Structural differences and specificities of mRNA translation initiation complexes between trypanosomes and their hosts

Anthony Bochler

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ
INSTITUT DE BIOLOGIE MOLÉCULAIRE ET CELLULAIRE
CNRS – UPR 9002

THÈSE DE DOCTORAT

Présentée par :

Bochler Anthony

Soutenue le : **2 Octobre 2020**

Pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Mention : **Sciences de la Vie et de la Santé**

Discipline : **Aspects Moléculaires et Cellulaires de la Biologie**

Structural differences and specificities of mRNA translation initiation complexes between trypanosomes and their hosts

Spécificités et différences structurales des complexes de l'initiation de la traduction de l'ARNm entre les trypanosomes et leurs hôtes

Thèse dirigée par :

Dr. Yaser Hashem	Chargé de recherche, Université de Bordeaux
Dr. Pascal Auffinger	Directeur de recherche, Université de Strasbourg

Rapporteurs :

Pr. Zoya Ignatova	Professeure, Université de Hambourg
Pr. Alexey Amunts	Professeur, Université de Stockholm

Autres membres du jury :

Dr. Eric Ennifar	Directeur de recherche, Université de Strasbourg
Dr. Marie Sissler	Directrice de recherche, Université de Bordeaux

Acknowledgments/Remerciements

Dans un premier temps, je tiens à remercier Yaser Hashem de m'avoir donné l'opportunité de poursuivre une thèse de doctorat au sein de son équipe. Je remercie également Pascal Auffinger pour sa co-direction. Cette thèse m'a permis de m'épanouir professionnellement dans des domaines que j'apprécie tels que la cryo-microscopie électronique, la modélisation moléculaire et la science en générale.

Plus généralement, je tiens à remercier le personnel de l'Institut de Biologie Moléculaire et Cellulaire de Strasbourg et de l'Institut Européen de Chimie et Biologie de Bordeaux pour leur accueil et leur sérieux dans l'accompagnement de mes démarches administratives.

Je remercie les membres de mon jury Alexey Amunts, Zoya Ignatova, Marie Sissler et Eric Ennifar pour avoir accepté d'assister à ma soutenance et de discuter mon manuscrit.

Je remercie également mes anciens collègues de Strasbourg, Angelita, Margarita, Jay et Vincent pour les nombreux projets sur lesquels nous avons travaillé ensemble, toujours dans la bonne humeur.

Je tiens à remercier mes collègues et surtout amis Heddy, Florent, Camila, Thomas et Elodie pour les bons moments passés à Bordeaux.

Une pensée également pour mes amis de longue date et ma famille pour le soutien qu'ils m'ont apporté tout au long de mon doctorat. Une attention particulière pour ma moitié, Béa qui m'a accompagné durant ces nombreuses années dans les bons et les mauvais moments.

CONTENTS

I. INTRODUCTION 6

1. The gene expression: from DNA to protein 6

2. Ribosomes and their role in translation 7

3. Structural biology of translation initiation 9

4. Initiation of translation: a highly regulated step 16

4.1. General mechanisms 16

4.2. Eukaryotic initiation factors and their coordinated action in the translation initiation 18

5. The Kinetoplastids: biology of a killer 21

5.1. Trypanosoma brucei and the African sleeping sickness 22

5.2. Trypanosoma cruzi and the Chagas disease 23

5.3. Leishmania genus and the leishmaniosis 25

6. Kinetoplastids ribosome specificities 27

6.1. The unique expansion segments 27

6.2. The large subunit puzzle 28

7. Contribution of structural biology in the structural and functional understanding of large RNP complexes 29

7.1. Different methods for a common goal 29

7.2. Single particle Cryo-electron microscopy and the overcome of its bottlenecks 34

8. Molecular modeling: from data to 3D structure 36

8.1. The interpretation of structural biology data 36

8.2. The different levels of protein architecture 37

8.3. The RNA architecture and folding 40

8.4. Model building: from sequence to 3D structure 42

8.5. Model refinement: improving the model 43

2. PhD PROJECTS 44

3. RESULTS 45

Article 1: The cryo-EM structure of a Novel 40S Kinetoplastid-Specific Ribosomal Protein 45

Article 2: Structural differences in translation initiation between pathogenic trypanosomatids and their mammalian hosts 67

Article 3: Interaction Networks of Ribosomal Expansion Segments in Kinetoplastids 106

Article 4: Structural insights into the mammalian late-stage initiation complexes 130

4. CONCLUSION AND PERSPECTIVES 174

REFERENCES 177

Abbreviations

3D = three dimensional

cryo-EM = cryo-electron microscopy

eIF = eukaryotic initiation factor

ES = expansion segment

H bond = hydrogen bond

IC = Initiation complex

IF = initiation factor

NMR = nuclear magnetic resonance

MDFF = Molecular Dynamic Flexible Fitting

mRNA = messenger RNA

PDB = Protein Data Bank

PIC = pre-initiation complex

RNP = ribonucleoprotein particle

RP = ribosomal protein

rRNA = ribosomal RNA

snoRNA = small nucleolus-restricted RNA

TC = ternary complex

tRNA = transfer RNA

List of figures

Figure 1: The general steps of gene expression

Figure 2: The major steps of protein biosynthesis

Figure 3: General views of the eukaryotic ribosome from *S. cerevisiae*.

Figure 4: Table of small subunit ribosomal protein new nomenclature

Figure 5: Table of large subunit ribosomal protein new nomenclature

Figure 6: Schemes of ribosome layers.

Figure 7: The eukaryote translation initiation process

Figure 8: Structure and positioning of initiation factors on the 40S

Figure 9: Scheme of the anatomy of *T. cruzi* trypomastigote form

Figure 10: The life cycle of *T. brucei*

Figure 11: The life cycle of *T. cruzi.*

Figure 12: The life cycle of *Leishmania genus*.

Figure 13: Statistics of PDB structure deposition.

Figure 14: Protein homology modeling principle.

Figure 15: The four level of protein structure.

Figure 16. RNA architecture

I. INTRODUCTION

1. The gene expression: from DNA to protein

The gene expression concept was proposed in 1958 by Francis Crick. It is the process by which the information encoded in genes are expressed and synthetized into functional gene products. Most of the time, these products are proteins but some genes code for functional RNAs (non-coding RNAs) like transfer RNAs (tRNAs) or ribosomal RNAs (rRNAs) (Crick, 1958).

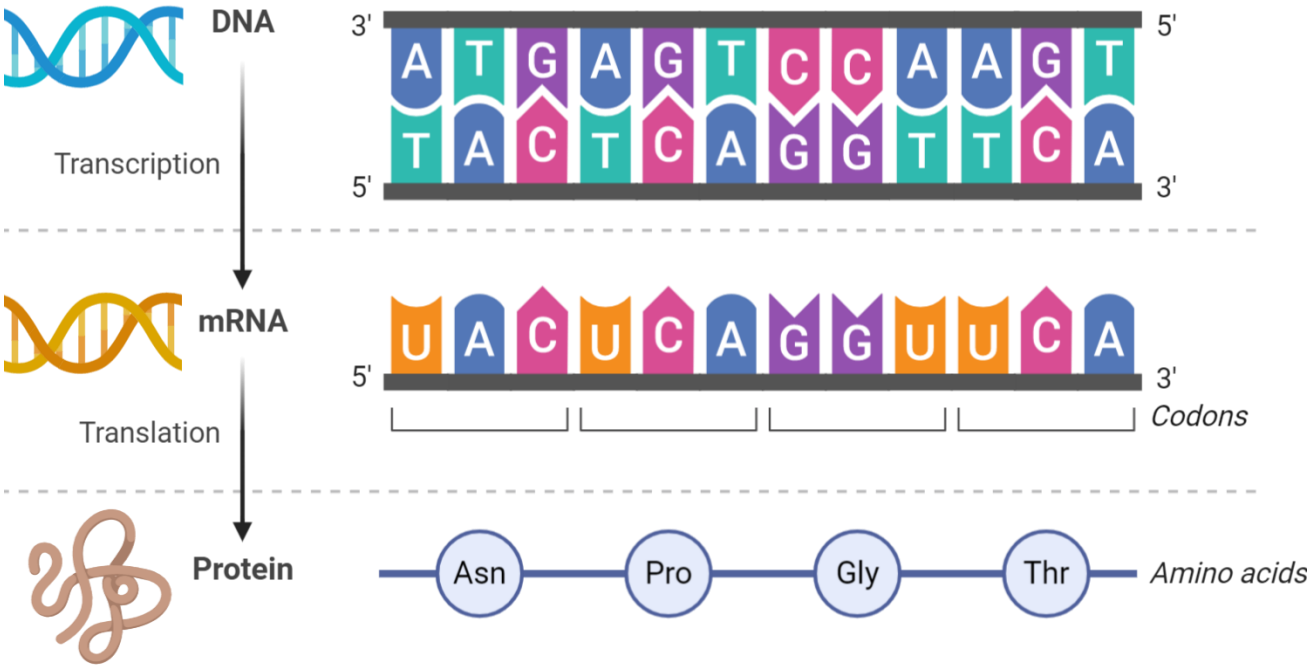


Figure 1: The general steps of gene expression. First the DNA is transcribed into mRNA. Second, the mRNA is translated into amino acids that are bound together into polypeptide chains.

The first step of gene expression is the DNA replication, which is the production of two replicas of DNA from one DNA molecule. The process starts from one double stranded DNA helix that will be separated into two single strands. Each strand will serve to produce one copy of the original DNA molecule through a molecular process carried out by a family of enzyme called DNA polymerase (Bell and Dutta, 2002).

The second step is the DNA transcription; this process is carried by the RNA polymerase enzymes. Those enzymes read DNA to produce a complementary RNA strand named primary transcript (Patikoglou and Burley, 1997). This step is followed by the maturation of the transcribed RNA. This maturation step differs depending on the nature of the RNA. The

non-coding RNAs, like rRNAs or tRNAs, are transcribed as precursors that need to undergo further processing to become mature RNAs before being exported out of the nucleus, in eukaryotes. For example, pre-rRNAs are cleaved and modified at specific sites by RNAs called small nucleolus-restricted RNA (snoRNAs), which associate with proteins to form snoRNPs. The snoRNA will basepair with the target RNA to position the modification at precise sites before the protein part performs the catalytic reaction of cleavage. The cleavage of the pre-RNAs will produce the different mature rRNAs. Later, they will form the cores of the ribosomal subunits in a process called ribosome biogenesis that occurs in a compartment of the cell nucleus called nucleolus (Sirri et al., 2008).

For coding RNAs, called messenger RNAs (mRNAs), maturation steps are required in eukaryotes. One of the pre-mRNAs maturation processes is the splicing, to get rid of non-coding regions (introns) and splice all coding regions (exons) together. The major actor of this process is the spliceosome, an RNA-protein catalytic complex (Gilbert, 1978; Matera and Wang, 2014). In prokaryotes, mRNA processing is quasi-inexistent, as transcripts are transcribed in an almost matured form, which do not require splicing (Apirion and Miczak, 1993). Another pre-mRNAs maturation step is the 5' capping enzymatic reactions that will modify the 5' guanine to confer it a methyl in its nitrogen 7. This modified cap will protect the mRNA from degradation by exonucleases and will aid the mRNA recognition by initiation factors to recruit it into preinitiation complex (Banerjee, 1980; Ramanathan et al., 2016a; Sonenberg and Gingras, 1998, p. 4). The last pre-mRNAs modification is the polyadenylation of the 3' end, which is necessary for mRNA export and translation re-initiation (Gilmartin, 2005; Zhao et al., 1999). Once mRNAs are fully processed, the protein biosynthesis can occur. This step is also known as translation and is an essential process for any living organism. It takes place in the ribosome, a large ribonucleoprotein complex.

2. Ribosomes and their role in translation

Ribosomes are ubiquitous molecular machines that read the genetic information contained in mRNAs and catalyze the assembly of amino acids to form polypeptide chains. Ribosomes are essential to any organism because of their key role in protein biosynthesis (Schmeing and Ramakrishnan, 2009). Due to their crucial role they are often the target of drugs, such as antibiotics (Steitz, 2008; Wilson, 2014; Yusupova and Yusupov, 2017). They are ubiquitous to both eukaryotes and prokaryotes. In eukaryotes, they are found in every compartments of the cell where translation takes place, cytosol but also mitochondria and chloroplast (Borst and Grivell, 1971; Harris et al., 1994; Palade, 1955).

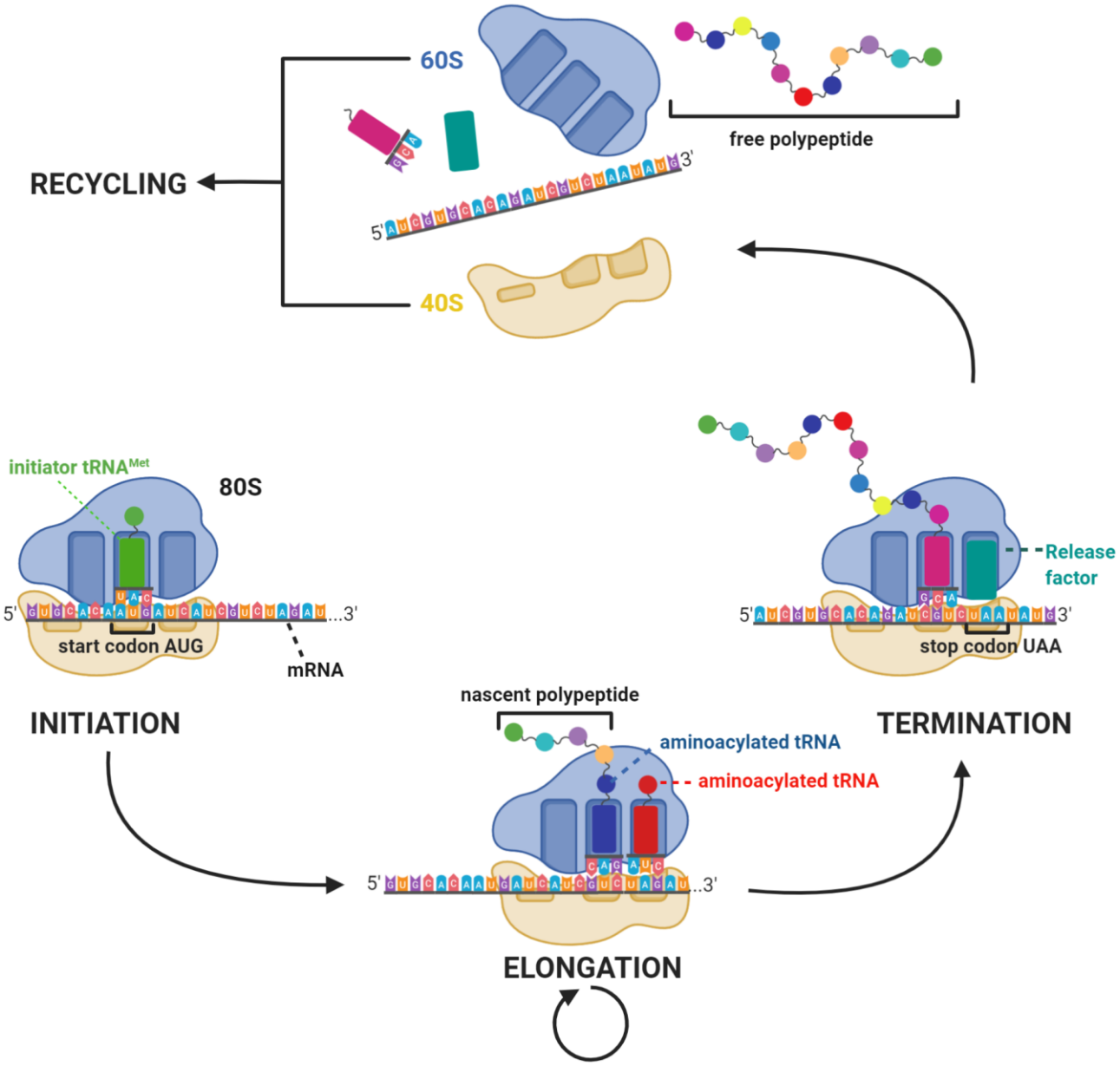


Figure 2: The major steps of protein biosynthesis. It starts with the initiation followed by elongation and finishes with the termination and the recycling of the ribosome subunits before starting a new translation process.

The translation mechanism can be roughly divided to three phases as seen in Figure 2: Initiation, Elongation and Termination. First, the initiation phase is a highly regulated step that involves several initiation factors and different conformational states. The goal is the assembly of a fully functional ribosome ready to commence translating the recruited mRNA. The elongation phase consists of a string of repetitive events mediated by several elongation factors, the purpose of which is to add one amino acid residue per mRNA base triplet following the AUG start codon into the nascent polypeptide chain (Dever and Green, 2012;

Valášek et al., 2017). Finally, the termination phase occurs when a stop codon is encountered. It will trigger the release of the nascent protein, aided by several release factors. The ribosomal subunits are then recycled for new round of translation (Dever and Green, 2012).

3. Structural biology of translation initiation

To fully understand the molecular mechanisms of these crucial steps of translation, especially the highly regulated step of initiation, structural characterization of each step is a prerequisite, and has become more than ever at reach, in association with biochemical molecular biology techniques. To do so, several biophysical techniques were developed over the past century in structural biology to witness, at an atomic level, the interactions of each factor with the small ribosomal subunit and the full ribosome. Historically, most of structural biology studies were conducted using X-ray crystallography, but a new technique was to come and redefine the frontiers of structural biology (Hauptman, 1990). Indeed, cryogenic electron microscopy (cryo-EM) is one of the most recent techniques of structural biology and has already proven its strength in structure determination over the past two decades (Frank, 2002). One of the major advantages of cryo-EM is the lack of need for crystallization that is often tedious especially for large molecular assemblies and hydrophobic molecules. Once a cryo-EM structure of a given complex is done, the resulting density map is used to build a molecular model using different specialized tools, similarly to X-ray crystallography. The ability to build molecular models is required to understand how the different components interact, therefore how factors work and what role they play in a complex. One of the key components of the translation process is the ribosome.

The first observation of a ribosome occurred back in the 50s by the American scientist George Emil Palade. He described the ribosome as “a particulate component of small dimensions (100 to 150Å)” (Palade, 1955). He observed it by electron microscopy with “a magnification of 5000 to 10000 and thereafter enlarged photographically”. His work on ribosome was awarded the Nobel Prize in medicine (1974), in association with Albert Claude and Christian de Duve.

The ribosomes is huge ribonucleoprotein complex. Although conserved in all living cells, its size can vary depending on species. For example, the prokaryotic ribosome of E. *coli* is a 21 nm particle. The eukaryotic counterpart is bigger and vary between 25 to 30 nm particle depending on the species. Universally, it is composed of both ribosomal RNAs (rRNAs) and

ribosomal proteins (r-proteins), organized in two major functional parts: the small ribosomal subunit (SSU), which that will bind the mRNA at the start codon before the large subunit association (LSU) (Figure 3), where the nascent protein is polymerized (Ramakrishnan, 2002).

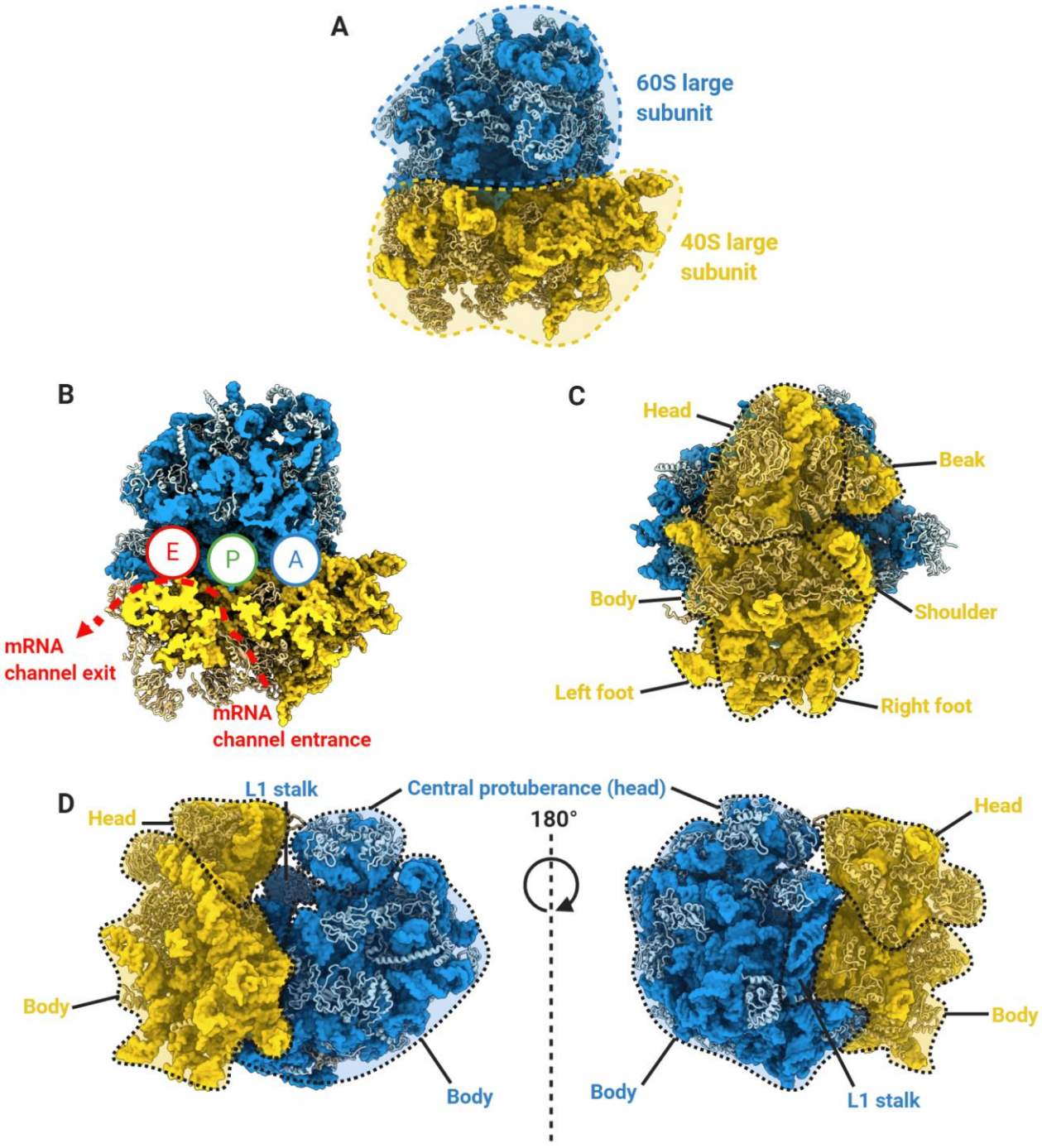


Figure 3: General views of the eukaryotic ribosome from *S. cerevisiae* (PDB: 4V7R, (Ben-Shem et al., 2010)). **A.** Overall view of the 80S eukaryotic ribosome **B.** View of the mRNA channel with the catalytic sites A, P and E highlighted. **C.** View from below the 40S small subunit with detailed anatomy. **D.** Views from the mRNA entry (left) and exit (right) channel.

Anthropomorphism is used to landmark the ribosome in order to navigate with ease on any ribosome structure. Subunits have a body and a head; the small subunit has two feet, a beak and a shoulder while the large subunit have the L1 stalk located near the exit channel of the mRNA (Figure 3 **D**) (Finkelstein et al., 2018).

The different steps of the translation of genetic information contained in the mRNA into a protein requires the coordination of several factors and ribosomal movements to read through the mRNA while assembling the amino acids brought by tRNAs. It is a complicated process combining high speed and accuracy (Green and and Noller, 1997) that occurs in the catalytic center of the ribosome, located in the interface between the two subunits.

When the ribosome is assembled, three catalytic sites are found at the inter-subunit interface (Figure 3 **B**). The first one is the A site (aminoacyl), where each aminoacylated tRNA will accommodate. Second is the P site (peptidyl), where the first initiation Methionine-tRNA is accommodated right before the elongation step. It is also the P-site that holds the tRNA with the nascent chain while the peptide bond is created between each translated amino acid. The third is the E site (exit), which holds the deacylated tRNA before its release from the ribosome to clear space for the next aminoacylated-tRNA incoming into the A site (Ramakrishnan, 2002).

The size and composition of the ribosome varies between prokaryotes and eukaryotes even though the overall structure and mechanism of translation is quite conserved. A new nomenclature has been instituted for the names and belonging (prokaryotes, archaea, eukaryotes) of each ribosomal protein for both cytoplasmic ribosome and mitochondrial ribosome. The following tables list these ribosomal protein (mitochondrial and chloroplastidial ribosomal proteins can be found in the full tables at the Nenad Ban lab webpage: https://bangroup.ethz.ch/research/nomenclature-of-ribosomal-proteins.html).

Cytoplasmic ribosome New name #	Occurrence*	Bacteria Old name	Yeast Old name	Human Old name
bS1	B m c	S1	–	–
uS2	BAE m c	S2	S0	SA
uS3	BAE m c	S3	S3	S3
uS4	BAE m c	S4	S9	S9
uS5	BAE m c	S5	S2	S2
bS6	B m c	S6	–	–
uS7	BAE m c	S7	S5	S5
uS8	BAE m c	S8	S22	S15A
uS9	BAE m c	S9	S16	S16
uS10	BAE m c	S10	S20	S20
uS11	BAE m c	S11	S14	S14
uS12	BAE m c	S12	S23	S23
uS13	BAE m c	S13	S18	S18
uS14	BAE m c	S14	S29	S29
uS15	BAE m c	S15	S13	S13
bS16	B m c	S16	–	–
uS17	BAE m c	S17	S11	S11
bS18	B m c	S18	–	–
uS19	BAE m c	S19	S15	S15
bS20	B c	S20	–	–
bS21	B m c	S21	–	–
bTHX	B c	THX	–	–
eS1	AE	–	S1	S3A
eS4	AE	–	S4	S4
eS6	AE	–	S6	S6
eS7	E	–	S7	S7
eS8	AE	–	S8	S8
eS10	E	–	S10	S10
eS12	E	–	S12	S12
eS17	AE	–	S17	S17
eS19	AE	–	S19	S19
eS21	E	–	S21	S21
eS24	AE	–	S24	S24
eS25	AE	–	S25	S25
eS26	E	–	S26	S26
eS27	AE	–	S27	S27
eS28	AE	–	S28	S28
eS30	AE	–	S30	S30
eS31	AE	–	S31	S27A
RACK1	E	–	Asc1	RACK1

Figure 4: Table of the small subunit ribosomal protein new nomenclature. From the Nenad Ban lab webpage.

Cytoplasmic ribosome New Name [#]	Occurrence*	Bacteria Old name	Yeast Old name	Human Old name
uL1	BAE m c	L1	L1	L10A
uL2	BAE m c	L2	L2	L8
uL3	BAE m c	L3	L3	L3
uL4	BAE m c	L4	L4	L4
uL5	BAE m c	L5	L11	L11
uL6	BAE m c	L6	L9	L9
bL9	B m c	L9	–	–
uL10	BAE m c	L10	P0	P0
uL11	BAE m c	L11	L12	L12
bL12	B m c	L7/L12	–	–
uL13	BAE m c	L13	L16	L13A
uL14	BAE m c	L14	L23	L23
uL15	BAE m c	L15	L28	L27A
uL16	BAE m c	L16	L10	L10
bL17	B m c	L17	–	–
uL18	BAE m c	L18	L5	L5
bL19	B m c	L19	–	–
bL20	B m c	L20	–	–
bL21	B m c	L21	–	–
uL22	BAE m c	L22	L17	L17
uL23	BAE m c	L23	L25	L23A
uL24	BAE m c	L24	L26	L26
bL25	B m	L25	–	–
bL27	B m c	L27	–	–
bL28	B m c	L28	–	–
uL29	BAE m c	L29	L35	L35
uL30	BAE m	L30	L7	L7
bL31	B m c	L31	–	–
bL32	B m c	L32	–	–
bL33	B m c	L33	–	–
bL34	B m c	L34	–	–
bL35	B m c	L35	–	–
bL36	B m c	L36	–	–
eL6	E	–	L6	L6
eL8	AE	–	L8	L7A
eL13	AE	–	L13	L13
eL14	AE	–	L14	L14
eL15	AE	–	L15	L15
eL18	AE	–	L18	L18
eL19	AE	–	L19	L19
eL20[†]	AE	–	L20	L18A
eL21	AE	–	L21	L21
eL22	E	–	L22	L22
eL24	AE	–	L24	L24
eL27	E	–	L27	L27
eL28	E	–	–	L28
eL29	E	–	L29	L29
eL30	AE	–	L30	L30
eL31	AE	–	L31	L31
eL32	AE	–	L32	L32
eL33	AE	–	L33	L35A
eL34	AE	–	L34	L34
eL36	E	–	L36	L36
eL37	AE	–	L37	L37
eL38	AE	–	L38	L38
eL39	AE	–	L39	L39
eL40	AE	–	L40	L40
eL41	AE	–	L41	L41
eL42	AE	–	L42	L36A
eL43	AE	–	L43	L37A
P1/P2	AE	–	P1/P2 (AB)	P1/P2 (αβ)

Figure 5: Table of the large subunit ribosomal protein new nomenclature. From the Nenad Ban lab webpage.

The bacterial ribosome is called 70S, the “S” standd for “Svedberg” which is a centrifugation sedimentation unit. For the *E.coli* ribosome, its large subunit is 50S and is made up of 28 ribosomal proteins and two ribosomal RNAs: the 23S is the largest rRNA of the large subunit 2854nucleotides and the 5S 118 nt The small subunit 30S comprises 19 ribosomal proteins and one rRNA 16S of 1539 nt (Kaledhonkar et al., 2019).

The eukaryotic cytoplasmic ribosome 80S is bigger than its prokaryotic counterpart; its diameter varies between 25-30 nm (250-300 Å). For the *S.cerevisiae ribosome,* the large subunit 60S is made up of 41ribosomal proteins and three rRNAs: 28S of approximately 3396 nt, 5,8S 158 nt and 5S 121 nt. The small subunit 40S has only one rRNA called 18S of 1800 nt and 26 ribosomal proteins (Ben-Shem et al., 2010).

The number of nucleotides per rRNA and the number of proteins vary depending on the species.

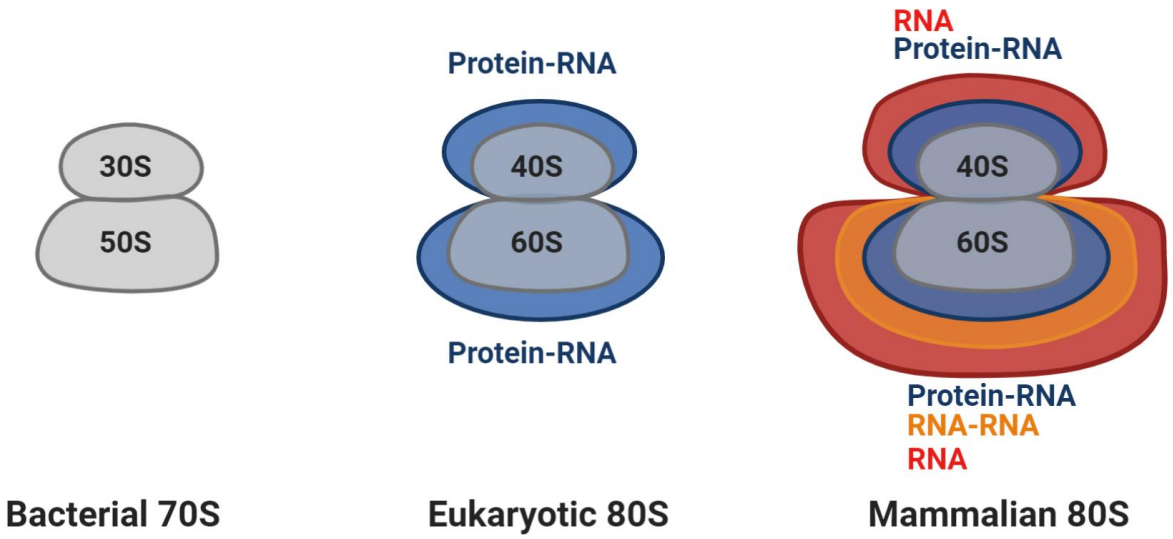


Figure 6: Schemes of ribosome layers. The universally conserved core (grey) corresponds to the 70S prokaryotic ribosome, the 80S has longer rRNA and more ribosomal proteins (blue layer). The orange layer corresponds to the rRNA expansion segments or additional r-proteins that are conserved among a group of species (example: mammals or kinetoplastids). Finally, the red layer corresponds to species-specific features like the expansion segments that are different in length and sometimes architecture between species of the same group (adapted from (Anger et al., 2013).

Thanks to the great number of structural studies carried out on a wide diversity of organisms, it is now clear that all ribosomes possess a conserved structural core (Figure 6) (Anger et al., 2013). Indeed, this universally conserved core among prokaryotes and eukaryotes, represent the catalytic site of the ribosome, where the activity is performed by the rRNA

(ribozyme). An additional layer of ribosomal proteins and extended rRNA form the eukaryotic ribosome 80S (Figure 6). Extra layers of RNA are found in eukaryotic species such as Kinetoplastids. Their 80S ribosome contains large RNA expansion called expansion segments. They are found in both subunits (for example ES6S and ES7S on the small subunit, ES27L and ES31L on the large subunit)(Anger et al., 2013; Hashem et al., 2013a).

The first successful crystallization of the ribosome was accomplished in the 1980s. Indeed, the large subunit of a bacterial ribosome from *Bacillus stearothermophilus* was crystallized (Yonath et al., 1980). Over the two following decades, progress was made in ribosome crystallization as well as in X-ray diffraction allowing the resolution of several additional structures. The large subunit 50S of an archeon *Haloarcula marismortui* at 2.4 Å (Ban, 2000) and the bacteria *Deinococcus radiodurans* at 3.3 Å (Schluenzen et al., 2000) were solved while the small subunit 30S of the bacteria *Thermus thermophilus* was also determined the same year (Wimberly et al., 2000). These pioneering studies were awarded the Nobel Prize in chemistry to Ada Yonath, Thomas Steitz and Venkatraman Ramakrishnan in 2009. Later, the reconstruction of a 5.5 Å resolution of a full 70S ribosome of *Thermus thermophilus* in 2001 (Yusupov, 2001) was accomplished. These multiple crystal structures granted a lot of information about the architecture of ribosome and the multiple rRNA-RPs interaction. The first EM 3D structure of a ribosomal particle was published in 1987 (Radermacher et al. 1987, Journal of Microscopy) using the random conical tilt series technique and only in 1992 the first cryo-EM 3D structure of the *E. coli* ribosome was derived (Penczek et al. 1992, Ultramicroscopy), thus marking the beginning of a new era for the ribosome structural biology.

To study the molecular mechanisms of the protein translation, high resolution is required to see details of molecular interactions of the different actors of the translation initiation. Several groups achieved that by crystallizing complexes of ribosomes with tRNA and mRNA. The two first structures of a ribosome with mRNA and tRNA were published in 2006 from the team of Noller at 3.7 Å (Korostelev et al., 2006) and Ramakrishnan at 2.8 Å (Selmer, 2006).

4. Initiation of translation: a highly regulated step

4.1. General mechanisms

The first step of protein biosynthesis is the initiation. Initiation is the rate-limiting step of translation. The goal of initiation is to join the two subunits to form a full ribosome in complex with an mRNA and the initiator $tRNA^{Met}$; the two forming the codon-anticodon bound in the P site of the ribosome. Therefore, the ribosome is ready to commence the elongation step. Initiation is a highly regulated and dynamic step. In prokaryotes, it involves three major initiation factors called IF1, IF2 and IF3. In eukaryotes, it involves a dozen of eukaryotic initiation factors (eIFs) that will orchestrate several conformational changes to form a functional ribosome (Aitken and Lorsch, 2012).

Initiation is common to all organisms, but its regulation differs between prokaryotes and eukaryotes. For example in prokaryotes, the small subunit will bind to the mRNA through a consensus sequence called the Shine-Dalgarno sequence (Dalgarno and Shine, 1973; reviewed in Rodnina, 2018). In eukaryotes, this consensus sequence doesn't exist but there are two main ways of translation: the cap-dependent translation, also called canonical translation, and the cap-independent translation (Merrick, 2004). Cap-independent translation concerns a minority of mRNAs that does not have the m7G cap modification. One of the cap-independent translation processes uses Internal Ribosome Entry Site (IRES) to be directly recruited to the ribosome. This way of translation is usually observed in cells in response to abnormal conditions (Shatsky et al., 2018).

In the case of cap-dependent translation, the mRNA is bound to the SSU through a process called the cap-binding recognition (Sonenberg and Gingras, 1998, p. 5). Unlike prokaryotes mRNAs that are ready to be translated, eukaryotic mRNAs need a maturation step that will confer them the so-called cap. This cap is a modification of the first nucleotide located in the 5' extremity of the mRNA. The modification is called m7G because it is a methylation (-CH3) in the nitrogen N7 of the guanosine. This modification is required for the mRNA recognition by the small subunit (Ramanathan et al., 2016b). After binding of the mRNA to the small subunit, the latter will scan through the mRNA sequence until it reaches the start codon (AUG), defined in eukaryotes by the Kozak consensus (Kozak, 1999).

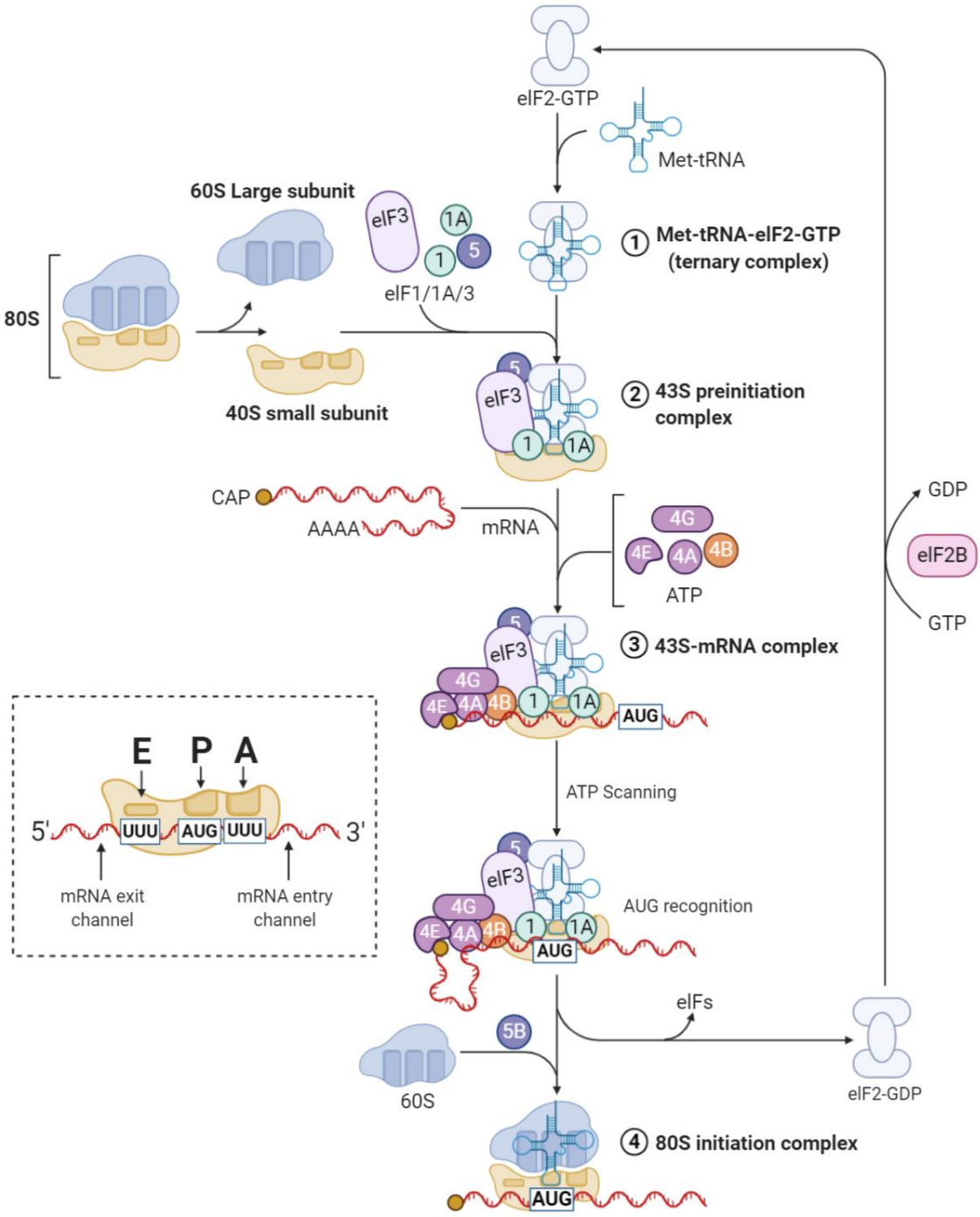


Figure 7: The eukaryote translation initiation process shows multiple steps and the coordinated action of a dozen initiation factors to achieve the assembly of a full elongation competent ribosome. Adapted from (Fraser and Doudna, 2007).

In eukaryotes, the first main step of initiation is the formation of the 43S preinitiation complex (PIC) (step 1 in Figure 7). It is subdivided in two steps, one is the formation of the ternary complex (TC) and the second is the binding of eIF1, eIF1a, eIF3 and eIF5 to the small subunit. The TC is formed by the association of the initiator $tRNA_i^{Met}$, eIF2 and a molecule of GTP that will then be recruited by small subunit to form the 43S PIC (Hashem et al.,

2013b; Jackson et al., 2010). The next step consists of the recognition of the mRNA and its recruitment to the 43S PIC (step 2 in Figure 7). The mRNA will be bound on its 5' cap by the initiation factor eIF4. This IF has several subunits: the core eIF4F (composed of eIF4A, eIF4E and eIF4G) and eIF4B (des Georges et al., 2015). Once the mRNA is bound to the 43S PIC, the scanning to the start codon occurs (step 3 in Figure 7). The 43S PIC will move downstream the mRNA from 5' to 3' until it reaches the start codon (AUG) that will bound to the anticodon triplet of the $tRNA_i^{Met}$. This codon-anticodon formation marks the transition from 43S PIC to 48S initiation complex (IC) (Aitken and Lorsch, 2012). Once 48S IC is formed, initiation factors must be released from the IC to clear space for the accommodation of the large subunit 60S to form the 80S functional ribosome, ready for the elongation step (step 4 in Figure 7).

4.2. Eukaryotic initiation factors and their coordinated action in the translation initiation

An important step that involves initiation factors is the proper assembly of the preinitiation complex on the correct mRNA, followed by accurate selection of the start codon AUG (Valášek et al., 2017).

To begin the translation cycle, mRNA must be recruited to the ribosome. The recruitment is performed by the group of eIF4F factors forming the cap-binding complex. The eIF4E subunit recognizes the mRNA modified 5'-cap, while eIF4G interacts with the helicase eIF4A and poly(A)-binding protein PABP1. Together with eIF3, these factors represent the major driving force in mRNA recruitment and accommodation in the mRNA channel of the small subunit 40S (Valášek et al., 2017). One of the co-operative roles of eIF1, 1A, 3 and 5 is to open the 40S mRNA channel before the docking of the mRNA (Valášek et al., 2017).

The mRNA scanning requires the unwinding of mRNA's secondary structure. It is the role of eIF4A coupled to the DHX29 helicase (for higher eukaryotes) to permit the ribosome to move smoothly along 5'UTR until a start codon is reached (des Georges et al., 2015; Rogers et al., 2002, p. 4). Start codon recognition is ensured by the CAU anticodon of the $tRNA_i^{Met}$ complementary to the common initiation codon AUG. This initiator tRNA is a part of the TC that is recruited to the ribosome aided by several eIFs such as eIF1, 1A, 3 and 5.

The AUG recognition will trigger hydrolysis of the GTP molecule bound to eIF2, aided by eIF1, 1A, 3 and 5 (des Georges et al., 2015; Valášek et al., 2017).

Translation initiation factor 3 (eIF3) is the most complex of all eIFs due to its composition. In mammals, it is composed of a total of 12 subunits including an octamer core of 8 subunits (a, c, e, f, h, k, l and m) and 4 peripheral subunits (b, d, g and i) (des Georges et al., 2015; Hinnebusch, 2006, p. 3; Valásek, 2012). Due to its complexity and its interaction with multiple eIFs, eIF3 has been studied intensively and it appears to play roles in nearly all steps of initiation. The current knowledge about eIF3 is that it keeps the 40S and 60S subunits from associating too early (Kolupaeva et al., 2005), it stimulates the recruitment of TC and mRNA to the pre-initiation complexes and it is involved in the scanning process (des Georges et al., 2015, Simonetti et al. 2016, Mol Cell).

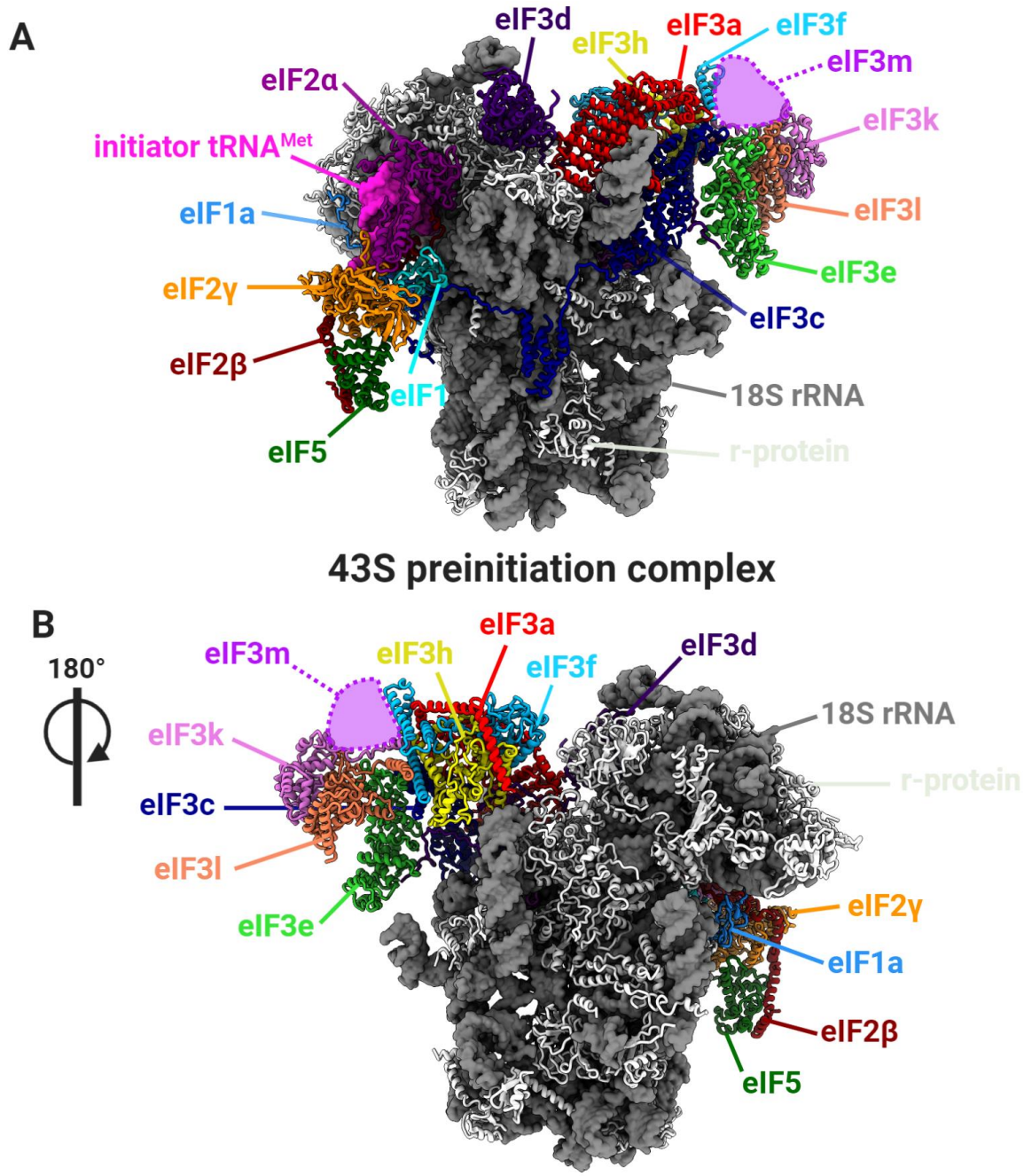


Figure 8: A. Structure and positioning of initiation factors on the 40S in the context of 43S pre-initiation complex of *Trypanosoma cruzi.* The eIF3m subunit is missing compared to mammalian preinitiation complex (Rezende et al., 2014). **B.** Same view from the other side of the complex (180°).

Translation initiation is a universal process, but some specificities exist in their regulation depending on the species, for example the lack of eIF3m subunit in *Trypanosoma cruzi* (Figure 8) common to others parasites of this group like *T. brucei* and *L. major* (Rezende et al., 2014). It is especially the case of a large group of eukaryotic parasites called Kinetoplastids.

5. The Kinetoplastids: biology of a killer

Kinetoplastida is a wide group of flagellated protists. It was initially classified in 1963 by the scientist Bronislaw Honigberg (Honigberg, 1963). The group is characterized by a specific cell component named the kinetoplast (See Figure 9). The kinetoplast is located at the base of the flagella, within the single large mitochondrion of the organism. It contains a unique mitochondrial DNA structure called kDNA organized in a giant network of interlocked rings (Vargas-Parada, 2010). The group is divided into several genus, the Bodonidae which are biflagellated bacteriophage free-living species, the Trypanosomatidae which are uniflagellated and contains several exclusive parasitic species such has *Trypanosoma brucei* and *Trypanosoma cruzi*. Additionally, the *Leishmania* are exclusively parasitic and closely related to Trypanosomatidae. Previously cited *Trypanosoma brucei* and *cruzi* are the causing agent of several human diseases called trypanosomiasis while *Leishmania* species are causing leishmaniosis.

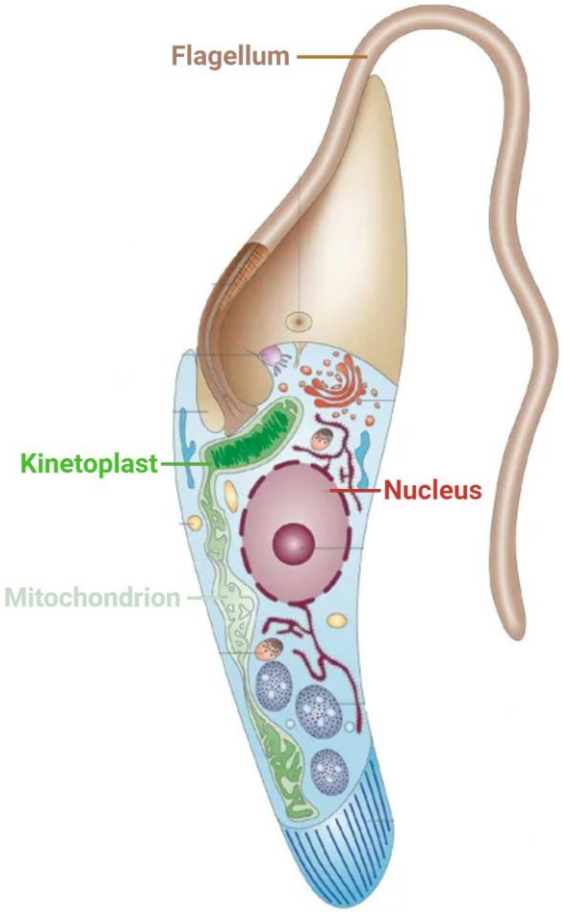


Figure 9: Scheme of the anatomy of *T. cruzi* trypomastigote form.This family of parasite is named after its specific organite called the Kinetoplast colored in green in the figure.(adapted from (Docampo et al., 2005)).

5.1. *Trypanosoma brucei and the African sleeping sickness*

Trypanosoma brucei regroup the African species of Kinetoplastids. Two of their species are causing the African trypanosomiasis commonly called the sleeping sickness, a disease vectored by tsetse flies. *Trypanosoma brucei gambianse* is found in west and central Africa and is responsible of the chronic infection representing the majority of sleeping sickness cases while *Trypanosoma brucei rhodesiense* causes the minor but deadliest form of the disease in east and southern Africa (Simarro et al., 2014). *T. brucei rhodesiense* can causes death within months for untreated infected patients while it can takes years for *T. brucei gambianse* to be lethal (Brun et al., 2010). The name sleeping sickness originates from the fact that Trypanosomal parasites produce a chemical compound called Tryptophol that induces sleep in humans (Cornford et al., 1979).

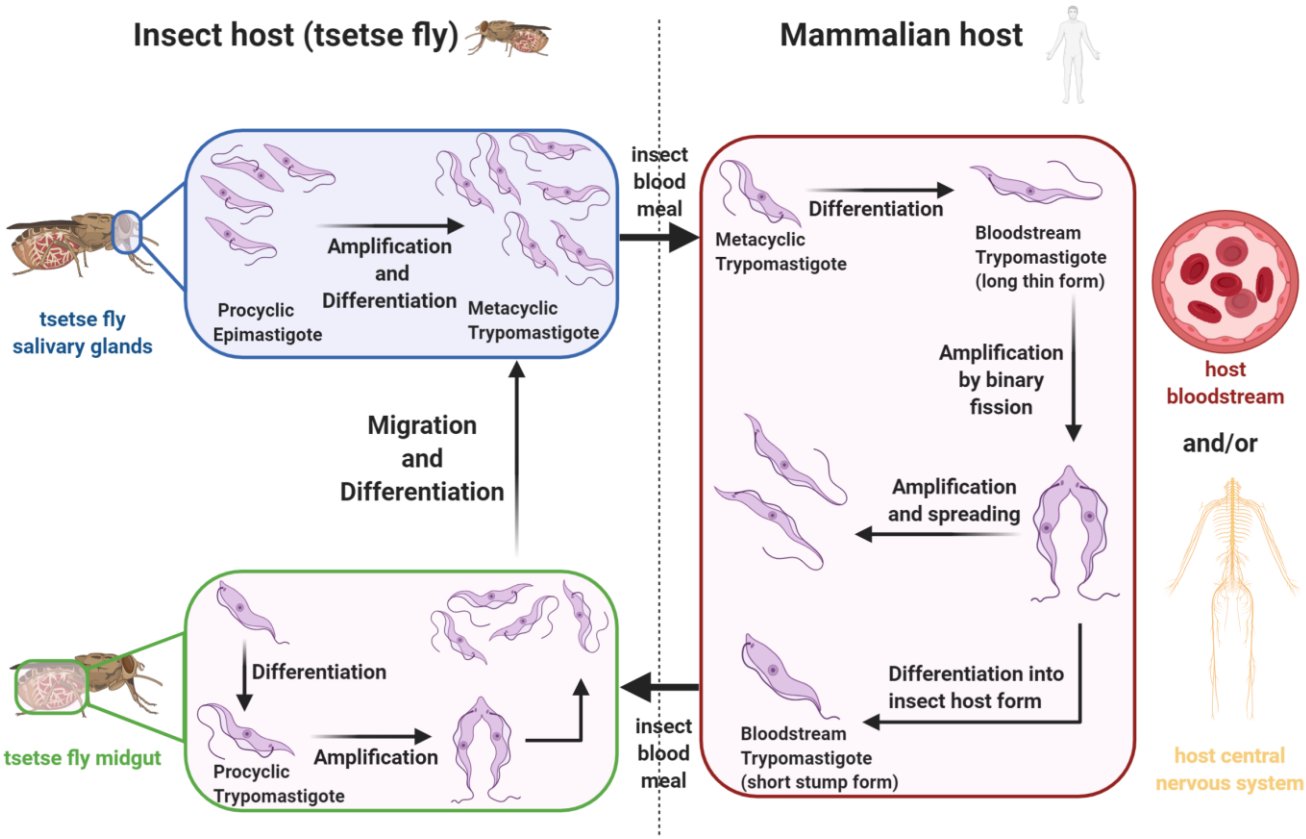


Figure 10: The Life cycle of *T. brucei.* The insect vector of this parasite is several species of tsetse flies.(adapted from (Field et al., 2017)).

As shown in Figure 10, *T. brucei* adopts different forms during the different stages of its life and depending on the infected host (Field et al., 2017). The vector of this parasite is the insect tsetse fly that feeds on blood from mammals, including from humans. The parasites are contained in the salivary glands of the insect in its metacyclic trypomastigotes form.

When the insect bites, it will inject metacyclic trypomastigotes into the mammalian host. The parasites are then transformed into extracellular bloodstream trypomastigotes that spread across different parts of the body. They multiply by binary fission in the blood, lymph or even spinal fluid. The bloodstream trypomastigotes long slender form can differentiate in short stump form that will be ingested in other tsetse flies while it feeds on infected mammalian hosts. Thus, spreading the parasites across multiple new hosts (Pollitt et al., 2011). The short stump form is now in the insect guts, where it differentiates into procyclic trypomastigotes. This form has the ability to multiply by binary fission and to differentiate into procyclic epimastigotes, which will migrate into the salivary glands of the tsetse fly. The last transformation of the cycle is from procyclic epimastigotes into metacyclic trypomastigotes ready to infect a new mammalian host through the tsetse fly bite.

The pick of the disease occurred in the late 1990s, it spread across 36 sub-Saharan countries, threatening millions of people (Malvy and Chappuis, 2011). In reaction to this alarming pick, the World Health Organization (WHO) launched important operations for the disease surveillance and control have been greatly strengthened. The hard work of the committed health workers resulted in the reversion of the epidemiological trend in the 2000s (Franco et al., 2014). Nowadays, the sleeping sickness is declining and is targeted for elimination in the 2020s by the WHO (Franco et al., 2017).

5.2. Trypanosoma cruzi and the Chagas disease

Trypanosoma cruzi is found in central and south America. It causes the American trypanosomiasis also called Chagas disease (Rassi et al., 2010). The parasite can infect humans and multiple other mammalian species. Several transmission types have been observed; the main one is through the main vector of the disease: the triatomine bug. This bug is a member of the Triatominae family, commonly found in Latin America. It is also known as kissing bugs or vampire bugs because it hides in roots or walls during the day and comes out at night to bite the face of people or animal to feed on its blood (Rassi et al., 2010). The triatomine becomes infected by *T. cruzi* by feeding on blood of an infected mammalian. After biting, the bug defecates near the fresh wound caused by the bite. It usually itches, causing the victim to scratch the wound, making the feces in contact with the wound. By this process, the parasites, hidden in the insect feces, are able to enter the mammalian host organism (“CDC - DPDx - American Trypanosomiasis,” 2019).

The human transmission usually occurs in poor rural areas like deforested and piassava palm culture areas where the insects are looking for new food sources because of the

deforestation causing the fauna to be greatly thinned (Teixeira et al., 2001). Other human transmission process have been recorded, such as blood transfusion, organ transplantation, mother to baby through placenta and even by contaminated food (Rassi et al., 2010).

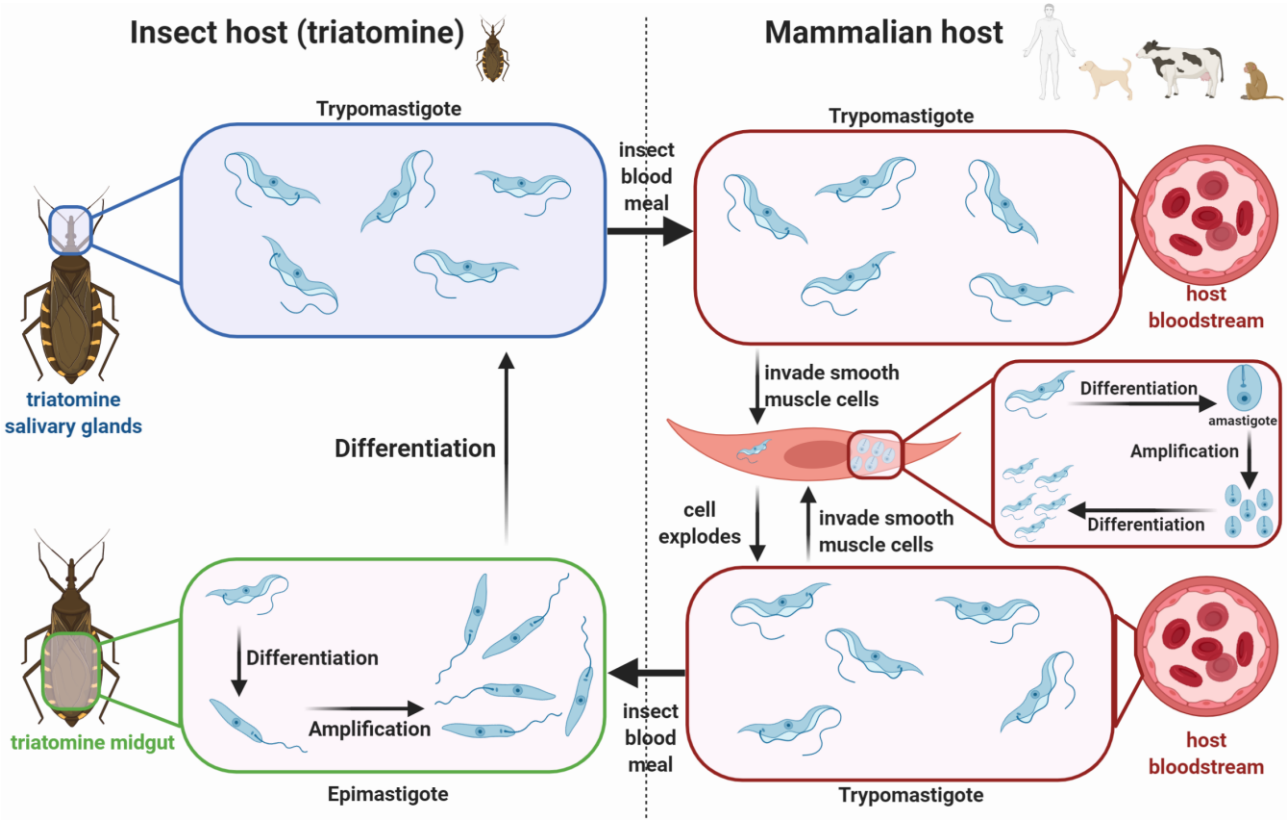


Figure 11: Life cycle of *T. cruzi*. This specie has a complex life cycle composed of several forms. First, epimastigote is the replicative form while in the insect host. Second, amastigote is the intracellular replicative form in mammal host. Finally, trypomastigote is the infective form found in both hosts. (adapted from (Field et al., 2017).

T. cruzi has also multiple forms during its different life stage, usually separated in two main stages: one in the insect vector and the other in the mammalian host (Figure 11). When the triatomine feasts on the blood of an infected mammalian host, it will ingest trypomastigotes. The parasites are now in the insect guts, where they differentiate into the replicative form called epimastigotes. After replication, the epimastigotes differentiate into the infective form called metacyclic trypomastigotes, ready to infect a new mammalian host through the insect feces (Field et al., 2017). Unlike its African counterpart, *T. cruzi* is an intracellular parasite; the trypomastigotes will invade several types of cells. Once in the cytoplasm, they differentiate into the intracellular replicative form called amastigotes. After the replication, the amastigotes redifferentiate into trypomastigotes provoking the rupture of the invaded cells, releasing parasites into the bloodstream where they will either invade new cells or

getting ingested by another triatomine through a bite.

Chagas disease was discovered and described by the Brazilian scientist Carlos Chagas back in 1909, it can affect human and more than 150 wild or domestic animal species. It is classified as a neglected tropical disease meaning that it infects low-income populations and are overshadowed by other public health issues ("CDC - Global Health - Neglected Tropical Diseases," 2020). Nowadays, Chagas disease affects 10 million people in endemic Latin American countries where its main vector, the triatomine, are proliferating. Several hundred thousand cases are also found in non-endemic countries like Mexico, United States and even in European countries like Spain. The estimated annual new cases are more than 41000 in endemic countries, and more than 14000 new-born getting infected directly by their mother (Rassi et al., 2010). The mortality of the disease is mainly by cardiomyopathy caused by the parasite, the number of death per year is approximately 10300 deaths in 2010 up from 9300 in 1990 (Lozano et al., 2012). Contrary to African trypanosomiasis, the American trypanosomiasis is still an up-to-date problem and is getting worse every year because it remains neglected.

There are several ways to fight the Chagas disease. The first one is to prevent the spreading of the insect vector using insecticides (permethrin or cypermethrin) (Zerba, 1999). The second method is an antiparasitic treatment to kill the parasites in the host. During the early stage of the infection, the drugs of choice are azole or nitro derivatives such as benznidazole or nifurtimox (Garcia et al., 2005) but their effectiveness decrease in chronic stage patients and resistance to these drugs has been reported (Buckner et al., 1998). The cure rate of the antiparasitic treatment is 90% for infant and 60-85% of adults treated in the first year of infection. Until this day, the need of an effective treatment for Chagas disease is still required.

5.3. Leishmania genus and the leishmaniosis

Leishmania is a genus of trypanosomes present in East Africa and Indian sub-continent. It causes a disease called leishmaniasis, affecting more than 90 mammalian species including canids, rodents and humans. The disease is vectored by the sandflies, an insect of the Phlebotomus genus (Chappuis et al., 2007) also known to spread Malaria. Leishmaniosis affect more than 12 million people in 98 different countries with more than 1 million new cases each year and are caused by 21 *Leishmania* species (World Health Organization, 2012). Leishmaniosis is also classified as a neglected tropical disease; it mainly affects people living in poverty. It appears in three different clinical forms caused by different

species of *Leishmania*. The most common is cutaneous, caused by several species like *L. donovani*, *L. major*, *L. mexicana* or *L. braziliensis*. The rarest is mucocutaneous and transmitted by *L. braziliensis* and *L. guyanensis*. The most severe and potentially lethal form is the visceral leishmaniosis also known as the black fever or *kala-azar* is transmitted by *L. donovani* and *L. infantum.* (Barrett and Croft, 2012).

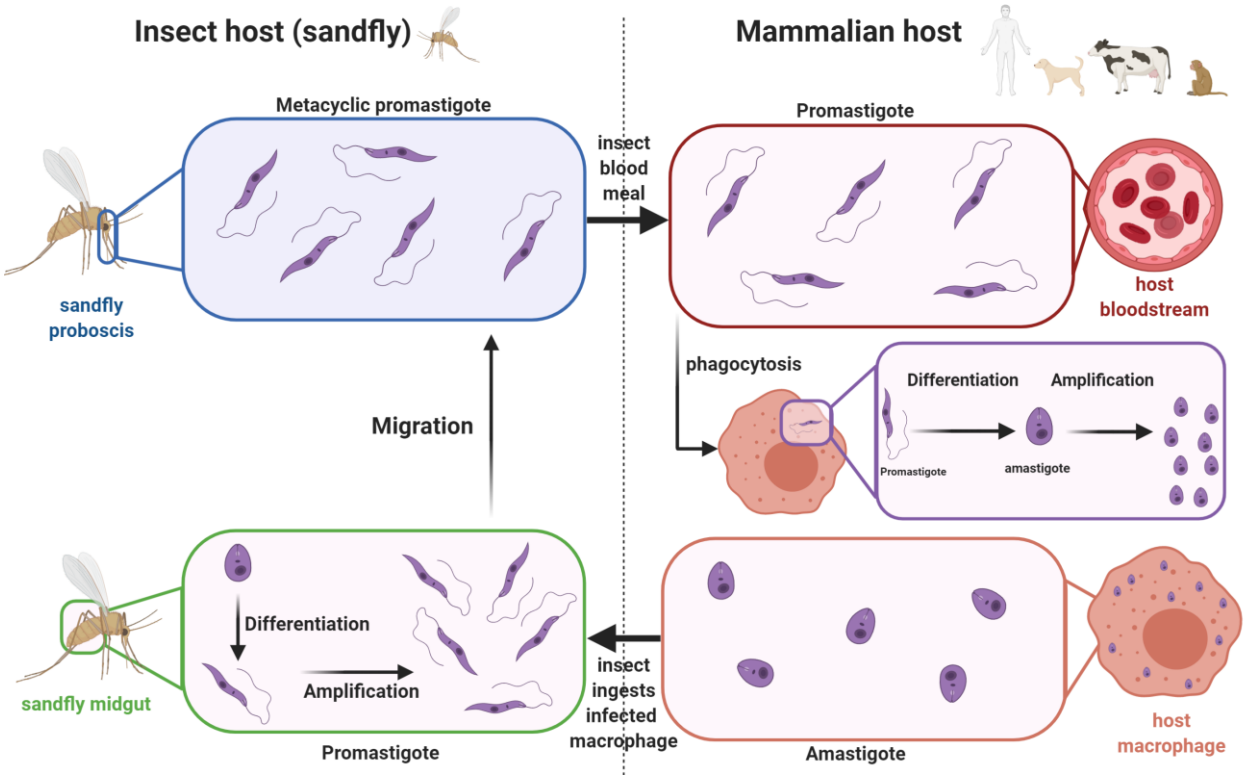


Figure 12: Life cycle of *Leishmania genus*. Leishmania has a slightly different spreading method compared to other kinetoplastids, indeed it uses the host immune system to spread across several hosts (adapted from (Field et al., 2017).

Just like other Kinetoplastids, *Leishmania* have several replicative and infective forms during its life cycle. It can be divided into two major cycles, one in the insect vector and one in the mammalian host (Figure 12). *Leishmania* are injected, in a metacyclic promastigotes form, into mammalian host during the blood meal of the sandflies. The immune system response of the host lead to the phagocytosis of the promastigotes by macrophages, where the parasites are transformed into the replicative intracellular form named amastigotes. The amastigotes will multiply and spread across multiple tissues causing the different clinical manifestations depending on the *Leishmania* species. The sandflies become infected by ingesting infected macrophages when they bite an infected mammalian. The ingested amastigotes are then in the insect midgut, where they differentiate into promastigotes and multiply before differentiating into metacyclic promastigotes and migrating into the insect

proboscis, where the cycle can restart with the next blood meal of the sandfly (Field et al., 2017).

6. Kinetoplastids ribosome specificities

The initiation complexes and ribosome have been studied a lot during the last decades and the general mechanisms of translation are known. The ribosome is used as a therapeutic target for drugs especially against bacteria (Hong et al., 2014; Wilson, 2014). Nonetheless, species-specific translation factors and mechanisms exist. These species-specific features can be the key targets for the development of new therapeutic strategies against neglected diseases such as Chagas disease or leishmaniosis. These two diseases are transmitted by kinetoplastids parasites, both are flagellated eukaryotic protozoa. They infect and threat millions of people, killing several thousand every year but they are still considered as “neglected tropical disease” because they mostly affect poor populations of the southern hemisphere. Nowadays, there is no vaccine for these diseases and the available anti-parasitic drugs show signs of resistance and lack efficiency (Baker et al., 2013). The need for new treatments is a necessity and the key might be in the translation process of those parasites.

6.1. The unique expansion segments

Despite the global conservation of eukaryotic ribosomes, kinetoplastid's one has multiple structural particularities split between the several complexes of the translation process (Clayton, 2016). When comparing mammalian 80S with a kinetoplastid 80S, the first thing that will catch the eye is the large structured rRNA expansions in the kinetoplastids ribosomes, especially on the SSU (Hashem et al., 2013a). Those long rRNA extensions are called expansion segments (ES). In kinetoplastid, the expansion segments can be up to ~600 nucleotides (ES6S)(Hashem et al., 2013a). The role of the RNA expansion segments is still unclear, nonetheless their positions all around the ribosome as an external layer and inter-subunit bridges might increase the ribosome structure stability (Hashem et al., 2013a). Also, the positions of ES6S (near the mRNA channel entrance and exit) and of ES7S (in close proximity to the mRNA channel entrance and eIF3) have been observed in a cryo-EM structure of 43S at ~5 Å in 2013 (Hashem et al., 2013b). RNA expansion segments are present in other eukaryotes like mammals, but they are usually smaller and less structured and compact than in their kinetoplastids counterpart. For now, kinetoplastids are the only

species presenting such structured and compact RNA expansion on their SSU. The question of their role was discussed for years but their high flexibility makes it hard to solve their structures at high resolution. Higher resolution structure could help in deciphering the role of these structural features.

6.2. The large subunit puzzle

As described previously (part 3), the eukaryotic ribosomal large subunit contains the 28S, 5S and 5,8S rRNAs. It has been demonstrated that in kinetoplastids, the 28S rRNA has the particularity of having multiple cleavage sites resulting on two big rRNAs fragments called LSU-α, LSU-β and four small rRNAs: srRNA1, srRNA2, srRNA3 and srRNA4 (Campbell et al., 1987). This particularity raises the question of how these eight rRNA parts are assembled inside the large subunit. The answer to this puzzle has been proposed in a publication showing the 2.5 Å structure of *Trypanosoma cruzi* ribosomal LSU (Liu et al., 2016). To propose a model for the assembly of the 60S, they compared the kinetoplastidian 60S with the yeast 60S. In their model, 5.8S rRNA together with LSU-α and LSU-β form the core that acts as a scaffold for the assembly of the other parts of the LSU. The core of the scaffold is stabilized by two interactions: the 5' end of LSU-α interacts with the whole 5.8S rRNA while the 3' end of LSU-α interacts with the 5' end of LSU-β. The 5.8S rRNA seems to play a central role in the 60S rRNAs assembly (Liu et al., 2016). The scaffold permits the assembly of srRNA2 and srRNA3. In kinetoplastids, some ribosomal protein presents extensions (Hashem et al., 2013a) like the globular domain of eL33 that provides a binding site for srRNA3 and LSU-α. The 5S rRNA can bind to uL5 and uL18. Finally, the two last pieces, srRNA1 and srRNA4 are assembled to form the 60S LSU (Liu et al., 2016).

This puzzle was proposed because of their exceptional (by their time in 2016) resolution and shows the importance of high-resolution structure to decipher the 3D arrangement but also propose a mechanism or role for factors based on their location on a molecular complex. This resolution was achieved by cryo-electron microscopy, a technique that evolved during the last decades to slowly grinding to the top tier method for the three-dimensional structure determination.

7. Contribution of structural biology in the structural and functional understanding of large RNP complexes

7.1. Different methods for a common goal

Structural biology is a branch of molecular biology that studies the three-dimensional structure of biological macromolecules. It provides a comprehensive understanding of how molecular architecture performs the biological reactions (Nitta et al., 2018). The structure determination uses three main techniques counting X-ray crystallography, nuclear magnetic resonance (NMR) and cryo-electron microscopy (cryo-EM). Each technique has its strengths and weaknesses compared to each other, but each of them allows to visualize how proteins and nucleic acids are organized in a three-dimensional space.

The first crystallography protein structure was solved in the 1950s when researchers used X-ray beams into protein crystals of myoglobin (Perutz and Weisz, 1947). This technique enabled to see how the myoglobin coiled into specific three-dimensional configurations (Kendrew et al., 1958). This discovery was awarded with the Nobel Prize in Chemistry to John Kendrew and Max Perutz in 1962.

NMR spectroscopy is a more recent technique developed in early 1980s by Richard Ernst, Kurt Wuthrich, Ad Bax, Marius Clore and Angela Gronenborn (Wüthrich, 2001). The first NMR protein structure solved was a proteinase inhibitor in 1985.
NMR uses the quantum mechanical properties of the nucleus of the atom to determine a 3D structure. NMR is a well-established method in structural biology. Its versatility permits the determination of macromolecules that cannot be crystallized or to solve structures of intrinsically disordered proteins that will never line up into a crystal lattice (reviewed in Marion, 2013). NMR is also useful to study the dynamics of flexible domains of proteins or in drug design to study the chemical shifts occurring when a protein encounters its binding partner.

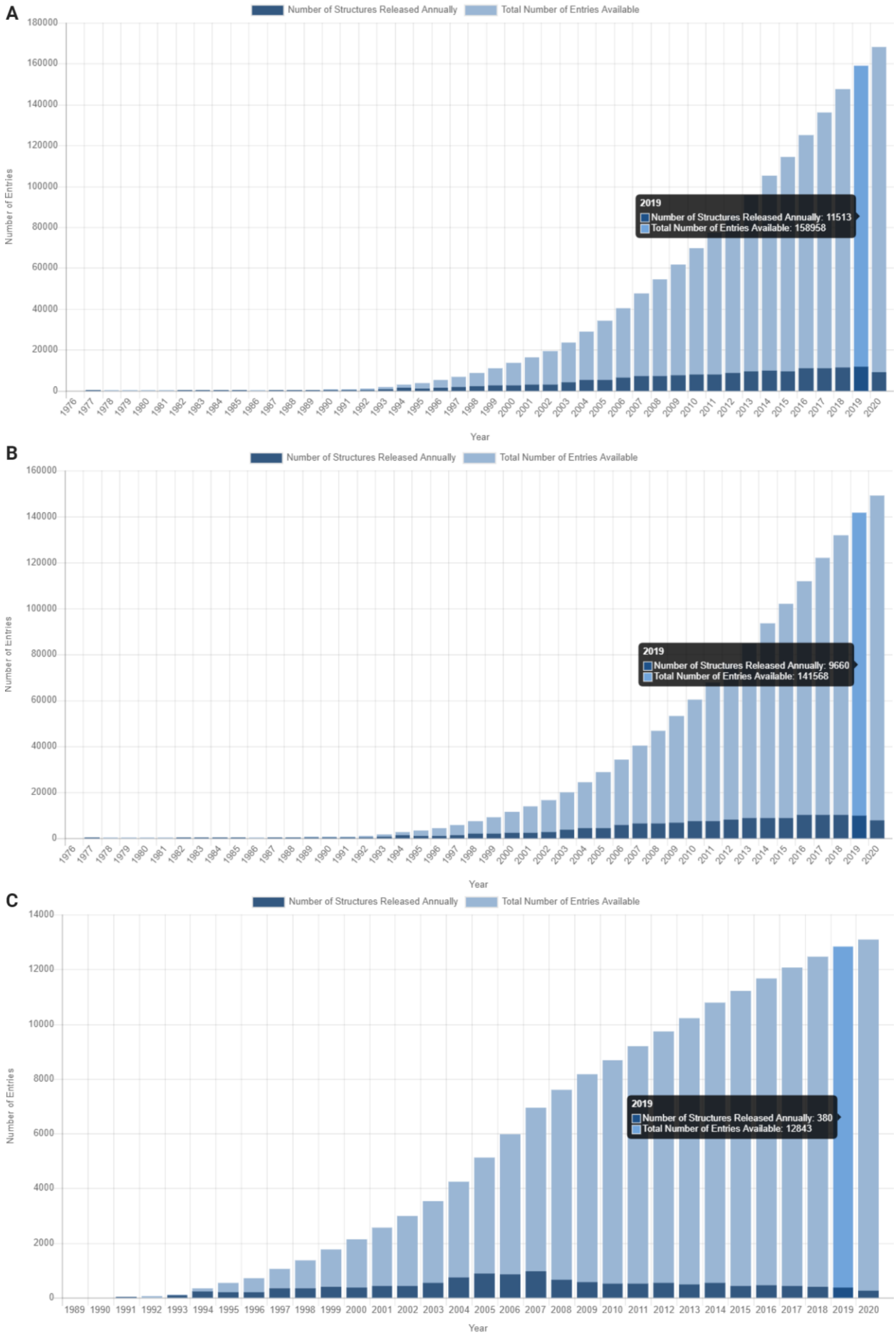
Among the biggest advances in structural biology, the development of electron microscopy was most influential. It is based on the electrons wave-behaviour that can be deflected in electric and magnetic fields using electromagnets as lenses to create focused electron beams. The first electron microscope was created by combining several of these lenses. It was a revolution in biology imaging by bringing a new level of resolving power in the 1940s

(reviewed in Masters, 2009). Unfortunately, one of the problems of using electron beams directly on a biological sample is the damages inflicted on the sample. This damage is in part the result of electrons interacting with exposed atoms of the sample, generating free radicals damaging the molecule and the X-rays generated by inelastic electrons. The electron doses have to be kept to a minimum to avoid sample damage but at the cost of a low-contrast images (reviewed in Masters, 2009). Another technical difficulty encountered by electron microscopy is the strong interaction of electrons with matter. To counter this, electron microscopes require high vacuums to prevent scattering of the beam by air molecules, making the study of living systems difficult (Masters, 2009). Samples must be dried or frozen. Early electron microscopy used techniques called positive and negative straining to improve the contrast of transmission electron microscopy (TEM). Those techniques consist of adding heavy metal atoms to the samples before being dried out as a thin layer on a transparent carbon film. The first usage of negative staining to recover 3D information from 2D images was done on purified particles virus in the late 1960s (Masters, 2009). The resulting micrographs contained an enormous number of 2D randomly oriented particles of the same 3D object (here the virus capsid). By attributing an angle for each imaged particle, the 2D information can be reassembled into a three-dimensional picture of the structure. The first 3D image reconstruction of virus particles was a major step forward in electron microscopy, but the resolution was limited to 70 Å because of the beam damage and artifacts introduced by the heavy atom staining techniques (Crowther et al., 1970). The heavy atoms artifacts were resolved with the introduction of cryo-electron microscopy (cryo-EM). This recent technique uses the flash cooling of the sample in liquid ethane, preserving the biological sample in an extremely thin layer of glassy ice under more physiological conditions, avoiding the exposure to heavy metals or drying (Dubochet and Stahlberg, 2001). This type of preservation lowers the contrast, but it can be counteracted by the defocusing techniques and by using a greater number of particles in reconstructions to improve the signal-to-noise ratios. It uses the same principle of data processing as staining techniques; it starts from a high number of 2D randomly oriented particles to get a 3D structure. The technique was first used with viruses, because of their symmetrical aspect the angle attribution is easier, but cryo-EM is now able to get atomic resolution for asymmetric molecules such as ribosomes.

A few years ago, a variant of cryo-EM emerged known as cryo-EM tomography. Compared to single particle cryo-EM using purified preparations of particles, tomography can image the molecular landscape within cells, enlarging the territorial claims of cryo-EM in structural biology. The principle of cryo-EM tomography is to make 3D reconstructions of the cellular

interior by combining a series of images as a sample containing frozen cells is tilted or rotated within the microscope (Lučić et al., 2013). The current limitation is the thickness of 1 µm for the sample but is enough for imaging entire cells of some species of bacteria. Animal and plant cells are larger and less transparent than bacteria but the current development of a focused ion-beam to isolate 1 µm slices of those cells is ongoing (Curry, 2015).

Structural biology is in constant expansion thanks to technical improvements in the three-dimensional structure determination techniques and the evolution of computing power during the last decades. The number of deposited structures in the Worldwide Protein Data Bank (wPDB) (http://www.wwpdb.org/) shows that we entered in the high-throughput era of structural biology.


A
Number of Structures Released Annually
Total Number of Entries Available
Number of Entries
Year
2019
Number of Structures Released Annually: 11513
Total Number of Entries Available: 158958
B
Number of Structures Released Annually
Total Number of Entries Available
Number of Entries
Year
2019
Number of Structures Released Annually: 9660
Total Number of Entries Available: 141568
C
Number of Structures Released Annually
Total Number of Entries Available
Number of Entries
Year
2019
Number of Structures Released Annually: 380
Total Number of Entries Available: 12843

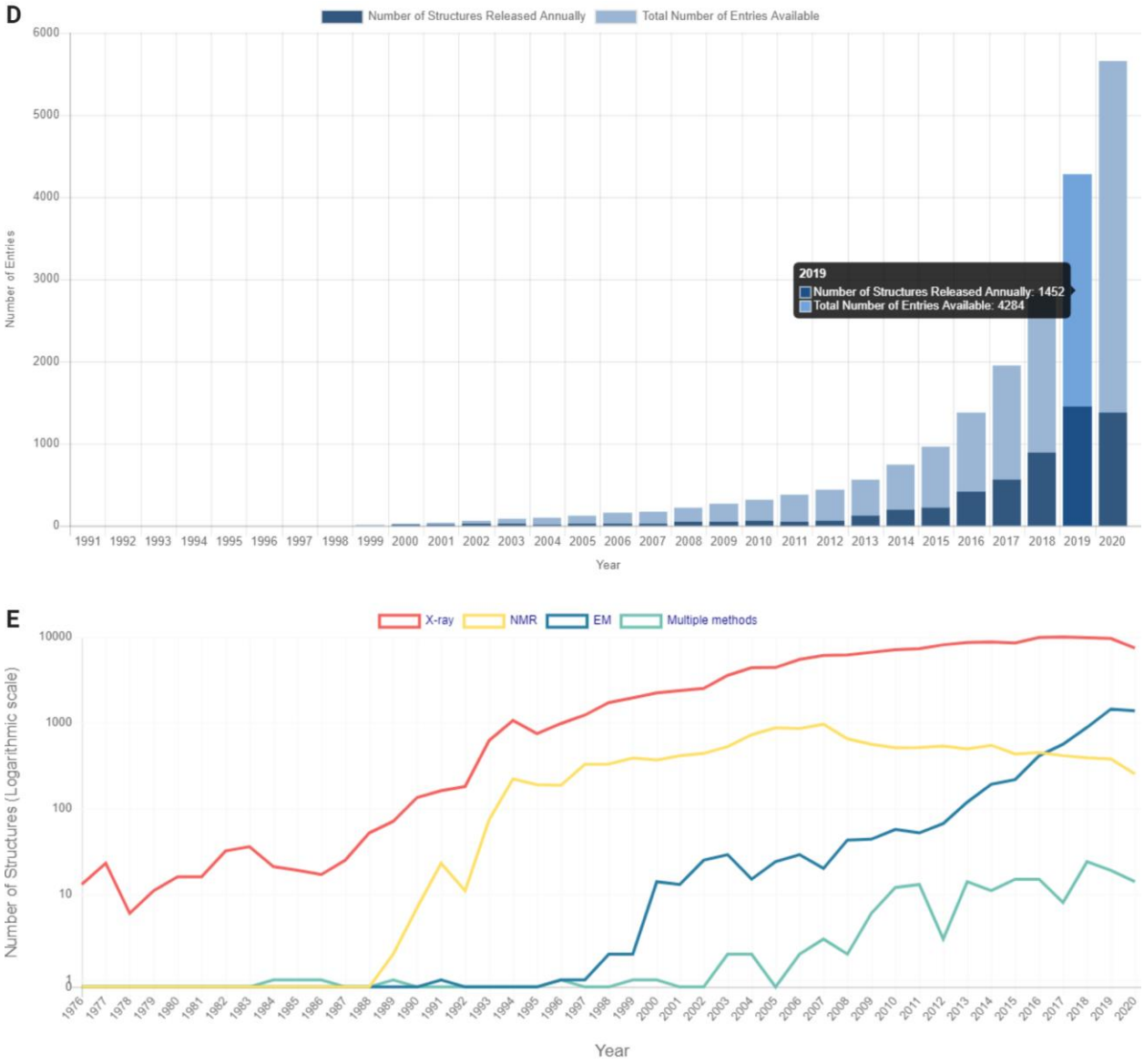


Figure 13: Statistics of PDB structure deposition with highlights in 2019. **A.** Overall statistics of deposited structure in the PDB. **B.** Number of X-ray structures deposited in the PDB. **C.** Number of NMR structures deposited in the PDB. **D.** Number of cryo-electron microscopy structures deposited in the PDB. **E.** Overall graphic of structure deposited per method with a logarithm scale.

In 2000 there were 2938 deposited structures versus more than 10000 per year since 2012 to reach 11513 in 2019 for a total of 158958 structures overall (Figure 13 panel **A**). But with such a high number of structures deposited each year emerges the issue of quality control. Among the high number of deposited structures in 2000, 2234 structures came from X-ray, 370 from NMR and only 11 from cryo-EM. In 2019, there were 9660 X-ray structures (Figure 13 panel **B**), 380 from NMR (Figure 13 panel **C**) and 1452 from cryo-EM (Figure 13 panel **D**). The X-ray technique is on a plateau since the last decade with a huge amount of structures of approximately 10000 structures per year while cryo-EM is in an exponential

phase, from 11 structures per year to 1452 in less than 20 years (Figure 13 panel **E**). When the last bottleneck of cryo-EM will be overcome, like the thickness of sample in tomography and the lengthy data acquisition and processing times, the number of structures will keep growing in the incoming years.

7.2. Single particle Cryo-electron microscopy and the overcome of its bottlenecks

In the 1970s, several scientists started to develop a new method to determine structure based on the electron microscopy. It was the beginning of the single particle cryo-electron microscopy (Dubochet et al., 1988; Frank, 1975).

At start, the results obtained were rather low resolution for example the structure of the ribosome at 30 to 50 Å resolution but the fact that sample doesn't need to be crystallized is a major advantage (Srivastava et al., 1995). Several decades of technical improvement made the atomic resolution possible in cryo-EM, these advances made the cryo-EM a major tool in structural biology to study molecules that are hard to crystallize (reviewed in (Cheng, 2018).

First problem of electron microscopy was to solve the 3D structure of a molecule by looking at 2D images of it. This problem was overcome by De Rosier and Klug that demonstrated the 3D structure reconstruction by combining 2D projection images of the same object in different orientation (De Rosier and Klug, 1968).

Second issue is the strong electron beam scattering due to its interacting with air molecules. EM samples need to be placed in a vacuum environment to avoid the scattering. The high vacuum is a problem for biological sample because it causes its dehydration, impacting the structural integrity of the sample. Keeping a biological sample hydrated under high vacuum was accomplished by Taylor and Glaeser by using frozen hydration on their catalase crystals, proving that a biological macromolecules can stay hydrated under vacuum by protecting the sample with ice (Taylor and Glaeser, 2008, 1974). This approach was hard to practice until the invention in the 1980s of the plunge freezing technique developed by Dubochet and its colleagues (Adrian et al., 1984; Dubochet et al., 1982).

The concept of this plunge technique is to apply purified protein samples in solution to an EM grid previously covered by a thin layer of carbon holey film. The solution is blotted on the grid with filter paper, removing most of the solution, leaving only a thin liquid film that will be drove into carbon film holes by surface tension. Next, the grid is plunged rapidly into

liquid ethane cooled by liquid nitrogen to froze it with an extreme speed (Dubochet et al., 1982). The swiftness of the freezing is crucial to form a thin layer of amorphous ice, if the ice has the time to crystallize, it will be opaque to electron and the sample will not be visible on the grid (reviewed in Cheng, 2018).

Now the grid contains a lot of the protein sample trapped into the ice in random orientations. and can be transferred into an electron microscope kept at liquid nitrogen temperature to acquire multiple images of the protein sample (reviewed in Cheng, 2018).

Third issue in cryo-EM is radiation damage caused by the high-energy electron beam. It limits the total electron dose used to image the biological samples, but this reduce the contrast or signal-to-noise ratios. A technique developed by Henderson and Unwin called electron crystallography countered this radiation damage problem by averaging images of many identical proteins packed as 2D crystals (reviewed in Cheng, 2018; Unwin and Henderson, 1975). Images were recorded with very low electron doses, so no features were visible, but their Fourier transforms show clear reflections. The combination of phases calculated from Fourier transformations of images and the amplitudes obtained from diffractions produced a high-resolution projection map of the specimen (reviewed in Cheng, 2018; Unwin and Henderson, 1975).

At the same time, Joachim Frank proposed to determine protein structures without crystallization by computationally combining images of many individual protein particles of the same type (reviewed in Cheng, 2018; Frank, 1975).

The combination of this idea with the plunge freezing sample preparation became the single-particle cryo-EM. The principle is to determine the structure by aligning and combining cryo-EM images of many molecules frozen under a thin layer of vitreous ice in random orientation. The signal-to-noise ratio can be increased by acquiring more images of the molecule in order to have the maximum different views to reconstruct the 3D structure of the molecule (Cheng et al., 2015; Fernandez-Leiro and Scheres, 2016).

One of the major breakthroughs in cryo-EM was the development of direct electron detection cameras (McMullan et al., 2016) avoiding the loss of information of the previous cameras that needed to convert electrons into photons. This enhancement coupled with the increasing computational power leading to new algorithms for data processing (Bai et al., 2013; Li et al., 2013) led the cryo-EM skyrocketing into the major structural biology method with X-ray crystallography.

Cryo-EM has been rewarded by the Nobel Prize in Chemistry in 2017 to Dubochet, Frank and Henderson for their work on the development of cryo-electron microscopy for high-resolution structure determination of biomolecules in solution.

8. Molecular modeling: from data to 3D structure

8.1. The interpretation of structural biology data

One of the keys to deciphering how a complex like the ribosome achieves its function in a biological process is the knowledge of how this ribosome interacts at an atomic level with its different actors (mRNA, tRNA and initiation factors). Then, structural study of these atomic details can be strengthened by biochemistry analysis to confirm the atomic observation in order to deepen the knowledge of a given biological process. Molecular modeling is fundamental and can be used to build atomic structure with the help of several specialized software and webservices.

Molecular modeling is used to build the three-dimensional structure of molecules into a density map resulting from cryo-EM data (Malhotra et al., 2019). The goal is to interpret the density by assigning correct molecules in its corresponding density. For large structure such as initiation complex or ribosome, prior knowledge is required to know how to correctly interpret such complexes composed of multiple proteins, initiation factors and RNAs. Already solved structure can be used as template if it is from a genetically close-related organism compared to the target model, for example the *T. brucei* ribosome can be used as a starting point to build *T. cruzi* ribosome. This technique is named “homology modeling”, and multiple tools exist such as *Swissmodel* (Waterhouse et al., 2018) or *Phyre2* (Kelley et al., 2015). The principle of these tools is to thread a target sequence in a template model in order to have the target model (Figure 14).

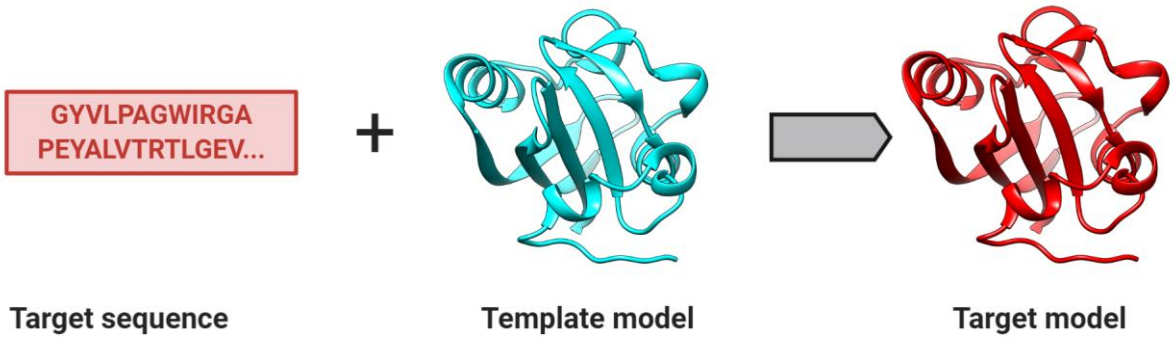


Figure 14: Protein homology modeling principle. The sequence is thread into a solved structure called “template” to obtain the structure of the target protein using homology modeling tools like *Swissmodel* or *Phyre2*. The same principle applies for RNA. (PDB : 2PYP (Genick et al., 1997))

The other modeling technique is *Ab initio*. It means that the model is built without any template, only based on prior knowledge, the target sequence and the density hints to build the model (Wu et al., 2007). If the resolution is good enough, side chains of amino acids will be visible, especially “bulky” side chains such as aromatic (tryptophan, phenylalanine and tyrosine) or basic (arginine, lysine and histidine). Also, the density will reveal secondary structures such as alpha helices or beta sheets to help the model building. It is usually the go-to method for freshly discovered factors, proteins or RNA.

8.2. The different levels of protein architecture

Proteins are found in every cells of every organisms. It is composed of one or several polypeptide chains formed by amino acids. The shape and structure of a protein is usually correlated to its function, a slight change in the sequence or structure can lead to different conformation with different function. It can be depicted as four levels of protein structure: primary, secondary, tertiary, and quaternary.

The primary structure corresponds to the sequence of amino acids in a polypeptide chain (panel **A** in Figure 15), as encoded by the corresponding gene. The order and composition of the sequence is decisive for the function of the protein. A nonsynonymous mutation in the DNA sequence leads to an amino acid substitution, which can alter the structure and/or the function of the protein, and possibly lead to a disease. The next level of protein structure corresponds to the amino acids sequence folding into secondary structures (panel **B** in Figure 15). They are formed by interaction of the backbone atoms to form different type of

secondary structure. Several secondary structure types exist, the most common are α helices and β-sheets (Ganapathiraju et al., 2004).For an α helix, hydrogen bonds (H bonds) are formed in the backbone between the oxygen of the carbonyl group (C=O) and the amino group (N-H) of an amino acid down the chain. The number of amino acids per helix turn is an average of 3.6 for a regular α helix. Some amino acids are favored to form α helix, for example methionine, leucine, lysine, alanine and glutamate have high helix-forming propensities compared to proline and glycine (Nick Pace and Martin Scholtz, 1998). Due to its backbone, proline tends to kink or break helixes and clash with the surrounding amino acids backbone. Inversely, glycine induces too much flexibility in the helix because of its lack of side chain (Richardson, 1981). This kind of secondary structure favor interaction because the side chains of each amino acid is pointing out of the helix. The β-sheet is the other common secondary structure motif found in proteins. It is formed by several β-strands interacting together through their backbone atoms to form hydrogen bonds. The backbone interaction is between the carbonyl group of one β-strand with the amino group of the neighboring β-strand, the β-strand minimum length is three amino acids. Favored amino acids to form β-sheets are aromatic residues and threonine, valine, isoleucine (Ganapathiraju et al., 2004). The third level of protein structure is the tertiary structure. It corresponds to the three-dimensional (3D) conformation of the protein (panel **C** in Figure 15). While some proteins are unfolded, others have conformation formed by the interaction of side chains of different parts of the protein. Usually, hydrophobic interactions play a key role in the three-dimensional shape of a protein because all hydrophobic side chains will cluster on the inside of the protein, leaving the hydrophilic amino acid side chains in the outside to interact with the water molecules of the surrounding solvent (Nič et al., 2009). The three-dimensional conformation of a protein can includes both α helix and β-sheet or only one type of secondary structure or none when it is a disordered protein. The last protein structure level is called quaternary structure. In some cases, a protein can only achieves its function while it is in a complex formed by multiple copies of itself or with other co-factors (panel **D** in Figure 15). It also includes biomolecular complexes of proteins with nucleic acids and other cofactors (Berg et al., 2002).

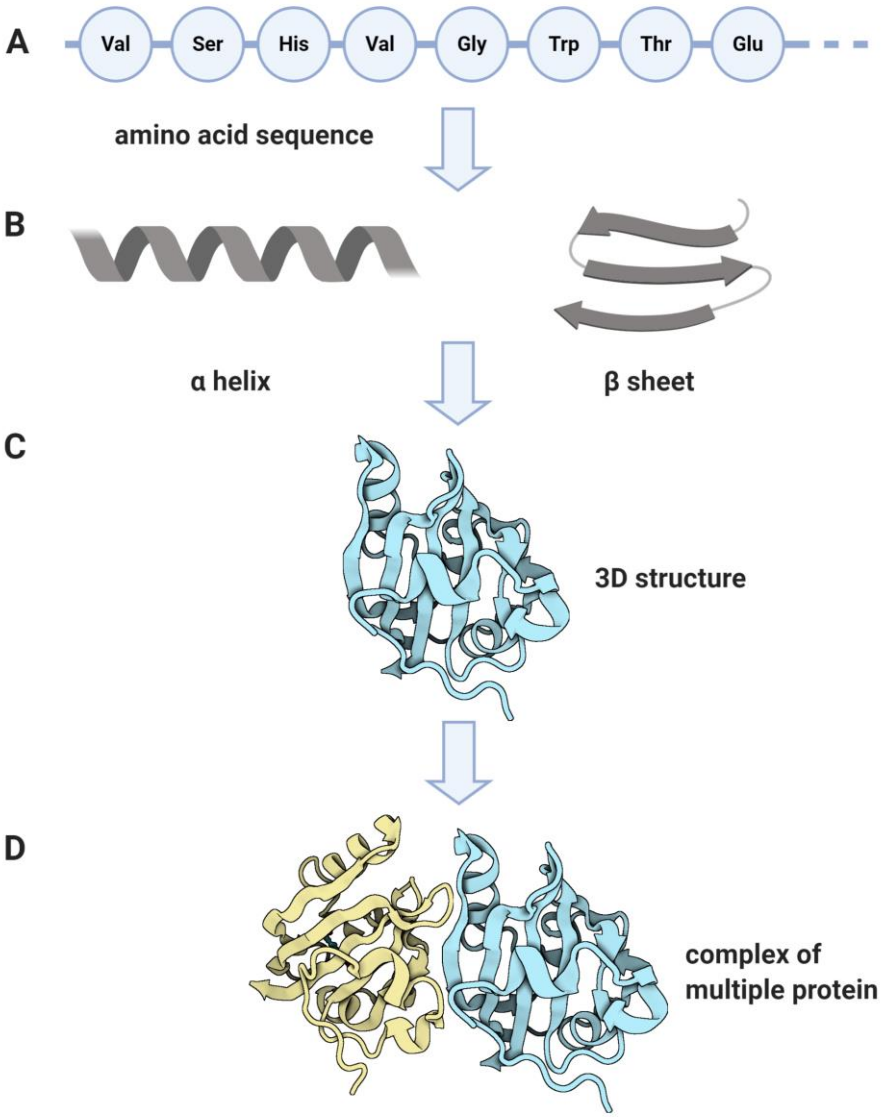


Figure 15: The four levels of protein structure. From sequence to 3D complexes, the function of a protein can depends on its sequence, but also on its conformation. (PDB : 2PYP (Genick et al., 1997))

The molecular modeling of proteins and RNAs require specialized software and webservices in order to build a 3D model.

8.3. The RNA architecture and folding

The ribonucleic acid (RNA) is essential to various biological roles. It is composed of nucleotides which are the association between a nitrogenous base, a ribose and a phosphate resulting in the nucleotides named uridine, cytidine, adenosine and guanosine (shortened to U, C, A and G) (Figure 16 **A**). These bases are divided into two categories, purines consist of a six-membered and a five-membered nitrogen-containing ring fused together, the adenine and guanine belong to the purines. Pyrimidines have only a six-membered nitrogen-containing ring, cytosine and uracil belong to the pyrimidines (Ezra et al., 1977). It can be found as a single-stranded molecule, but it often contains secondary structure and motif. RNA is able to assemble into three dimensional structure, either by base stacking or by base-pairing mediated by hydrogen bonding between complementary arrays of electrically polarized atoms (Leontis and Westhof, 2001). Each base has several ways of interacting with other base through three edges: Watson-Crick edge, Hoogsteen edge and sugar edge (Figure 16 **B**). Each base can also have two orientations called *cis* or *trans*. Resulting in 12 basic geometric types with at least two hydrogen bonds interaction between two bases (Leontis and Westhof, 2001). The results of such a huge number of base interactions are several possibilities of secondary and tertiary structure arrangement (Figure 16 **C**). To name a few, the canonical RNA helix, mainly formed by Watson-Crick and stacking interactions. The internal bulge which is one or more nucleotides forming a single strand in one strand of a helix, it is often needed to give flexibility to the structure. The stem loop, which is found at the edge of a double-stranded helix. The stem loop is the edge of a helix, it often interacts with other RNA or protein in context of biological complex. The three-way junction is a special arrangement between several co-axial RNA helices, it is for example found in ribozyme and plays a key role in the folding and activity. (Figure 16 **C**) (Lescoute and Westhof, 2006). In addition, more complex interactions can be found, such as multiloop junctions. The pseudoknot is a characteristic motif that is often found in viral IRES (Internal Ribosome Entry Site) (Jost and Everaers, 2010). These different secondary structures compose motifs and complex 3D structure such as the ribosomal RNA for example.

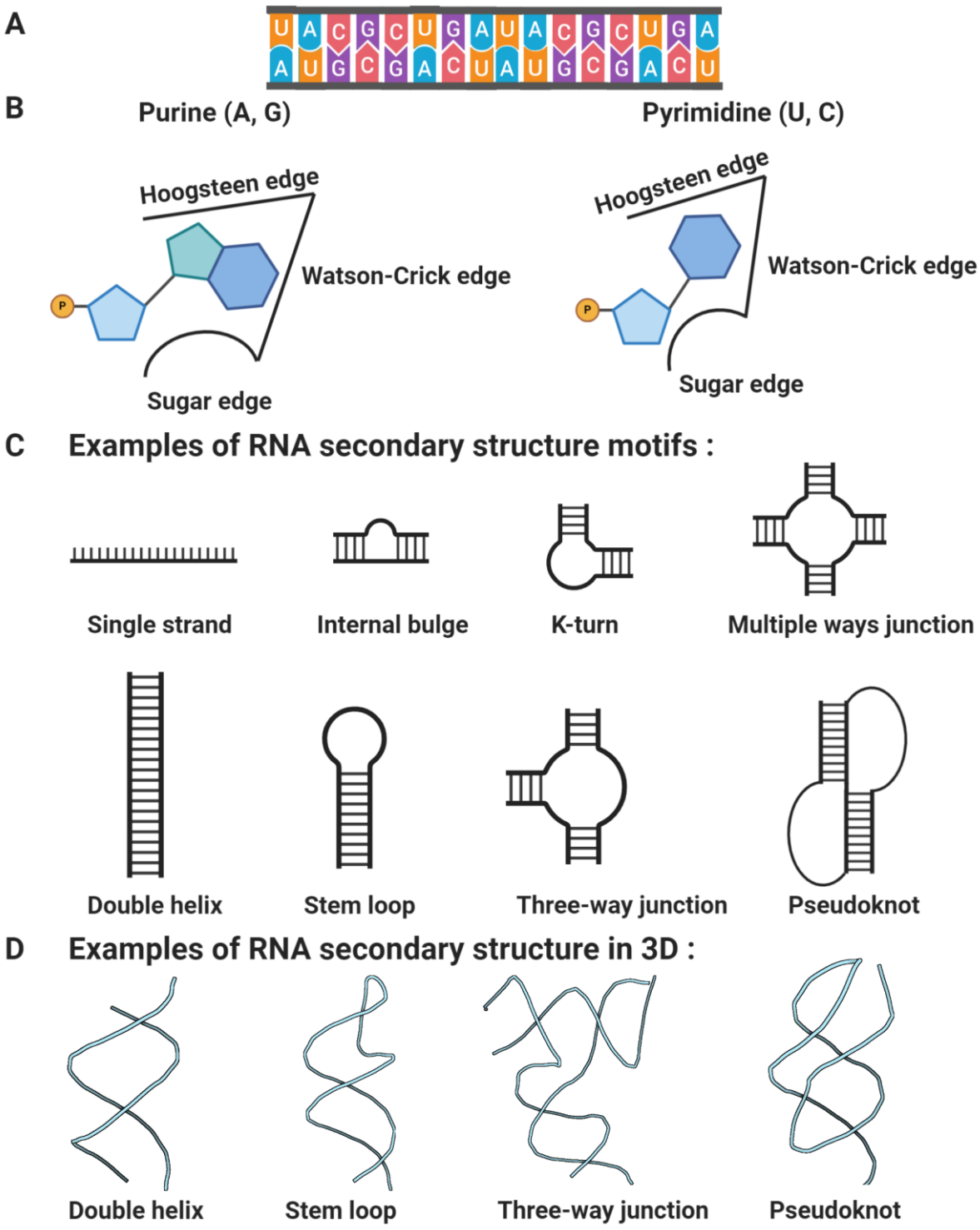


Figure 16. RNA architecture, from sequence to secondary structure to complex 3D interactions. **D.** From left to right, the double helix (PDB 413D, (Tanaka et al., 1999)), the stem loop (PDB : 6UGI (Chan et al., 2020)), the three-way junction from a ribozyme (PDB : 4KZ2 (Zhang et al., 2013)) and the pseudoknot (PDB : 2TPK (Holland et al., 1999)).

8.4. Model building: from sequence to 3D structure

The first step of modeling is to find the sequence of interest in one of the existing databases. The most substantial data banks are the Protein Data Bank (PDB) for proteins (Burley et al., 2019) and Nucleic Acid Database (NDB) (Berman et al., 1992) for nucleic acid. In order to find the target sequence, several sequence alignment tools called BLAST are available (Altschul et al., 1990). Some more specific database exists, for example, the TriTrypDB (Aslett et al., 2010) is a data bank specialized in Kinetoplastid sequences.

Once the sequence is retrieved, the next step depends on the molecule. If the target molecule has already been solved in a closely related organism, the model must be found on the PDB to be used as template for *homology* modeling. If the target molecule structure has not been solved yet, secondary structure information is required before going further in the *Ab initio* modeling. To do so, SYMPRED is a webservice containing several protein secondary structure prediction algorithms in order to produce a consensus secondary structure prediction (Simossis and Heringa, 2004). For RNA, several folding prediction algorithm exist such as RNAFold (Hofacker et al., 1994) or LocARNA (Will et al., 2012). These data are crucial to know where in the target sequence is likely to be found secondary structure like α helix or β-sheet for proteins. For RNA, these algorithms are used to predict where the RNA is more likely to be single or double stranded and if motifs like a triple junction or stem loop are located.

With all gathered information, the target molecule can be modelled in 3D in one of the specialized visualization and modeling software such as Coot (Emsley et al., 2010) and Chimera (Pettersen et al., 2004). Crossing the 2D prediction with the density map is essential to know where to start building. With a good resolution (~3-4 Å), secondary structures and even side chains (especially aromatics) should be clearly visible to build a correct 3D model of the target molecule.

Initiation complex and ribosomes contain numerous proteins, factors and RNAs and it takes some time to model everything in an organism that structure hasn’t been solved. When the initial model is complete, refinement steps are required in order to improve fitting of the model in the density map, to check for bad clashes and to correct geometries of all RNA and protein residues.

8.5. Model refinement: improving the model

A correct molecular model is crucial because local model errors can affect the afterward-biological interpretation. Having the perfect model requires a homogenous density map and resolution, which is not possible nowadays in cryo-EM due to several issues like beam damage or flexible parts of ribosomes. Fortunately, getting close to perfection is possible with the help of several specialized tools for model refinement.

One of the refinements methods is the Molecular Dynamic Flexible Fitting (MDFF). It is used to flexibly fit atomic structures into density map. The principle of MDFF is to add external forces proportional to the gradient of the density map into a molecular dynamic simulation of the atomic structures. These forces will attract the atomic structures in the density map to improve the general fitting while keeping the model geometrically correct (Trabuco et al., 2008). This method is efficient to improve the global fit of the structure in its density but the usage of forces in the atomic structures tend to distort the backbone atoms and rotamers are not always optimal when the density is not clear. To overcome this problem and correct geometry parameter, the next step consists of the usage of several tools. One of the most complete refinement and modeling software is PHENIX (Python-based Hierarchical Environment for Integrated Xtallography). This software regroups plenty of tools including refinement with the phenix.refine module for protein and phenix.erraser for RNA (Liebschner et al., 2019).

Phenix.refine is specialized in protein refinement, it will correct backbone geometry and find the best rotamers for each amino acid of your proteins while keeping a good fitting (Afonine et al., 2012). The complementary module to refine RNA is phenix.erraser and results in the same corrections as phenix.refine: backbone correction, base pairing and global fitting (Chou et al., 2016).

After these several corrections, the model needs to be validated. In order to do so, Molprobity is a webservice that will diagnose and output several geometrical parameters of the model with an associated lists of badly modelled atoms (Chen et al., 2010). Molprobity will output the global clashscore, the Ramachandran outliers, rotamers outliers, bad backbones, bad sugar puckers and more. Despite the several software ran, some mistakes and correction could remain. Molprobity will found them in order to correct them manually in Coot afterwards (Emsley et al., 2010) to obtain a correct molecular model.

2. PhD PROJECTS

Today, millions of people worldwide are struck by the Chagas disease and Leishmaniosis. The development of safer anti-parasitic drugs is a necessity because current treatments are not 100% efficient and even lethal for around 10% of patients. The African sleeping sickness is under control for now but the need to stay careful is an obligation. The core of the ribosome of Kinetoplastids and Human are conserved, making the use of ribosome targeting drugs toxic and dangerous for the patient. One way to find new therapeutic targets is to analyze and decipher structures at an atomic level in order to precisely understand the molecular mechanism and find potential new therapeutic targets. Our team is working on the analysis of translation initiation complexes of kinetoplastids and mammalian by cryo-electron microscopy and biochemistry. The kinetoplastids translation complexes such as ribosomes have already been solved and is different from their mammalian counterpart due mainly to their big RNA expansion. In addition, their translation initiation seems to have a lot of specificities that still need to be discovered and modeled in a 3D structure. Therefore, My PhD project focused on the modeling of kinetoplastids and mammalian translation initiation complexes in order to compare and highlight kinetoplastids structural specificities.

3. RESULTS

Article 1: The cryo-EM structure of a Novel 40S Kinetoplastid-Specific Ribosomal Protein

In Brief

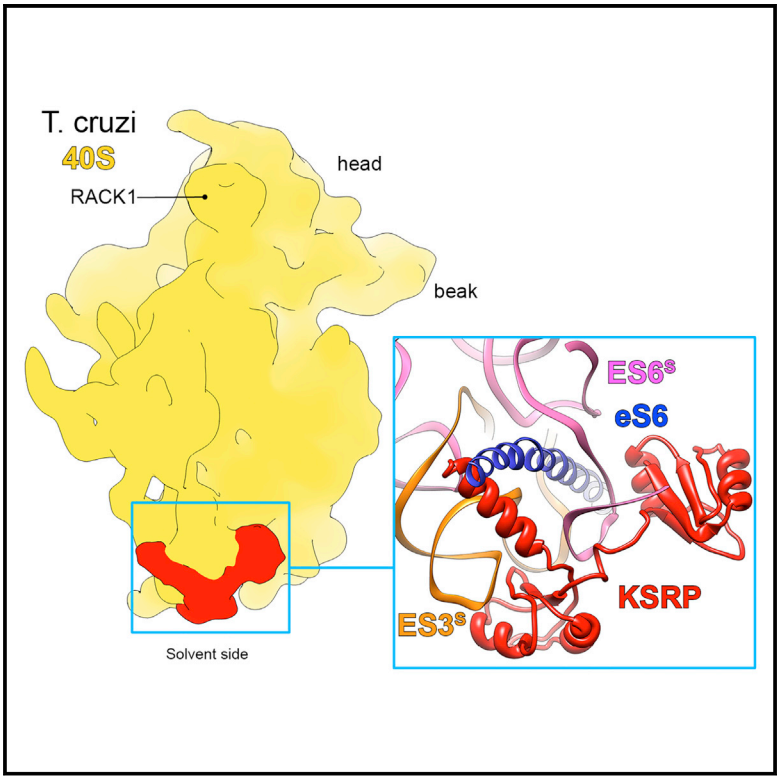
In this paper, we present the cryo-EM structure of a novel kinetoplastid-specific ribosomal protein (KSRP). Structure of upper eukaryotes ribosomal proteins and rRNAs has been solved for several years now compared to the study of Kinetoplastid ribosomes which is a recent field. Structure of Kinetoplastid ribosomes were performed mainly according to homology of their mammalian counterpart. Recent advances in cryo-EM permit to highlight an empty density on the foot of the small subunit 40S that appears to be a specific ribosomal protein of Kinetoplastid. Its structure and binding site are completely specific and conserved among all Kinetoplastid making it a new potential target for the development of safer and specific anti-parasitic treatment.

Article

Structure

The cryo-EM Structure of a Novel 40S Kinetoplastid-Specific Ribosomal Protein

Graphical Abstract



Authors

Jailson Brito Querido, Eder Mancera-Martínez, Quentin Vicens, Anthony Bochler, Johana Chicher, Angelita Simonetti, Yaser Hashem

Correspondence

y.hashem@ibmc-cnrs.unistra.fr

In Brief

Querido et al. present a cryo-EM structure of a novel kinetoplastid-specific ribosomal protein (KSRP) bound to the ribosome.

Highlights

- Cryo-EM structures of a novel kinetoplastid-specific ribosomal protein
- KSRP is an essential kinetoplastid protein that stabilizes specific rRNA elements
- KSRP interacts with the kinetoplastid-specific C-terminal region of protein eS6
- The entire interacting network of KSRP is kinetoplastid-specific

Brito Querido et al., 2017, Structure *25*, 1–10
December 5, 2017
https://doi.org/10.1016/j.str.2017.09.014

CellPress

Please cite this article in press as: Brito Querido et al., The cryo-EM Structure of a Novel 40S Kinetoplastid-Specific Ribosomal Protein, Structure (2017), https://doi.org/10.1016/j.str.2017.09.014

Structure
Article

CellPress

The cryo-EM Structure of a Novel 40S Kinetoplastid-Specific Ribosomal Protein

Jailson Brito Querido,[1,3] Eder Mancera-Martínez,[1,3] Quentin Vicens,[1] Anthony Bochler,[1] Johana Chicher,[2] Angelita Simonetti,[1] and Yaser Hashem[1,4,*]
[1]Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, 67000 Strasbourg, France
[2]Université de Strasbourg, CNRS, Plateforme Protéomique Strasbourg-Esplanade FRC 1589, 67000 Strasbourg, France
[3]These authors have contributed equally
[4]Lead Contact
*Correspondence: y.hashem@ibmc-cnrs.unistra.fr
https://doi.org/10.1016/j.str.2017.09.014

SUMMARY

Kinetoplastids are potentially lethal protozoan pathogens affecting more than 20 million people worldwide. There is a critical need for more specific targets for the development of safer anti-kinetoplastid therapeutic molecules that can replace the scarce and highly cytotoxic current drugs. The kinetoplastid ribosome represents a potential therapeutic target due to its relative structural divergence when compared with its human counterpart. However, several kinetoplastid-specific ribosomal features remain uncharacterized. Here, we present the near-atomic cryoelectron microscopy structure of a novel *bona fide* kinetoplastid-specific ribosomal (r-) protein (KSRP) bound to the ribosome. KSRP is an essential protein located at the solvent face of the 40S subunit, where it binds and stabilizes kinetoplastid-specific domains of rRNA, suggesting its role in ribosome integrity. KSRP also interacts with the r-protein eS6 at a region that is only conserved in kinetoplastids. The kinetoplastid-specific ribosomal environment of KSRP provides a promising target for the design of safer anti-kinetoplastidian drugs.

INTRODUCTION

Kinetoplastids are unicellular eukaryotic pathogens that infect crops, cattle, and humans. The best-described kinetoplastids are *Leishmania* and *Trypanosoma* species, which are pathogenic for humans and mammals. These organisms display complex life cycles with several developmental stages that are specific to the particular transit environment, including tissues or organs of the invertebrate vector and the host (Neto et al., 2016). Chagas disease and human African trypanosomiasis (HAT) are caused by *Trypanosoma cruzi* and *Trypanosoma brucei* sp., respectively, while different *Leishmania* species are the etiologic agents of different variants of leishmaniasis (such as *Leishmania donovani* causing visceral leishmaniasis in mammals; Chappuis et al., 2007). Kinetoplastid infections are classified as neglected diseases for the World Health Organization (WHO), as they affect millions of people living in poverty from Africa, South America, and Asia. According to a recent technical report of the WHO regarding the research priorities for Chagas disease, HAT, and leishmaniasis, there is an urgent need to substitute the few highly cytotoxic drugs that are currently available in the market, encouraging the discovery of new targets for the development of more efficient and secure drugs (Field et al., 2017).

Given its essential role in cell protein synthesis, the ribosome has been proven to be an excellent drug target to counteract bacterial infections (Moazed and Noller, 1987; Wilson, 2014). Several structural studies on the bacterial ribosome have largely facilitated the development of novel antibiotics with higher efficiency and specificity (Wilson, 2014). However, this ribosome-targeting-based strategy is strongly limited by the fact that the kinetoplastid and human ribosomes share the same eukaryotic nature. Hence, the main challenge for the specific targeting of the function or stability of the kinetoplastid ribosome is to uncover its kinetoplastid-specific structural features. Recently, intermediate to high-resolution structures of several kinetoplastidian ribosomes (Hashem et al., 2013; Liu et al., 2016; Shalev-Benami et al., 2016; Zhang et al., 2016) were determined by cryoelectron microscopy (cryo-EM) and pinpointed unique structural features when compared with other known eukaryotic ribosomes, such as large rRNA expansion segments (ES), as well as r-protein extensions. These cryo-EM studies led to the structural description of nearly all the rRNA chains, including most of the kinetoplastid-specific ESs and domains, as well as the assignment of all known eukaryotic ribosomal proteins (r-proteins).

In the 40S subunits of kinetoplastids, ES3^{S}, ES6^{S}, ES7^{S}, and ES9^{S} are substantially larger when compared with other known eukaryotic counterparts (Hashem et al., 2013). The secondary structures of these divergent rRNA ESs reveal that they could enclose various thermodynamically unstable domains, such as single-stranded RNA and large internal loops, possibly illustrating the need for more protein elements to stabilize these different rRNA domains (Hashem et al., 2013; Melnikov et al., 2015; Zhang et al., 2016). Consequently, due either to gene evolutionary shuffling or gene expansions, several kinetoplastid r-proteins have evolved larger N- or C- terminal tails (Ayub et al., 2009; Hashem et al., 2013; Zhang et al., 2016). This genetic divergence within the r-protein terminal regions has resulted in local structural rearrangements that guide and stabilize the conformation of some regions of these extended rRNA ESs

(Hashem et al., 2013). For example, in yeast and human ribosomes, the C-terminal α helix of r-protein eS6 spans only ~45 residues of the protein and remains solvent exposed, close to the 40S left foot formed by ES6^S (Ben-Shem et al., 2011; Khatter et al., 2015). In contrast, the kinetoplastidian eS6 displays a highly divergent C-terminal helix that spans the last 60 residues and is completely buried within additional kinetoplastid-specific domains of ES6^S and ES3^S (Hashem et al., 2013). Nevertheless, in spite of the recently reported high-resolution ribosome structures from several kinetoplastids (Hashem et al., 2013; Liu et al., 2016; Shalev-Benami et al., 2016; Zhang et al., 2016), no additional r-proteins accompanying these extra-large kinetoplastid-specific ESs were reported.

Given that kinetoplastid r-protein assignments were performed mainly according to homology to known r-proteins from other eukaryotes such as *Saccharomyces cerevisiae* (Hashem et al., 2013; Zhang et al., 2016) and *Tetrahymena thermophila* (Hashem et al., 2013), the discovery of novel kinetoplastid-specific r-proteins has been out of reach so far, as the assignment of any kinetoplastid-specific non-annotated protein to the ribosome complex is difficult to predict. Hence, despite the recent progress in the structural characterization of the kinetoplastidian ribosomal machineries, the full discovery of species-specific ribosomal interacting partners remains a major challenge in ribosome targeting for drug development. Using high-resolution cryo-EM and thorough biochemical analysis, we unveiled the first *bona fide* 40S kinetoplastid-specific ribosomal protein (KSRP) reported so far. In addition, we were able to unambiguously attribute several uninterpreted densities from the *Leishmania donovani* 80S cryo-EM structure solved previously (with an average resolution of ~3.0 Å) to KSRP, thus revealing the structure of the latter at nearly atomic resolution. These results reveal the structure of rRNA ES6^S, ES3^S, and r-protein eS6 C-terminal tail (C-ter) in direct interaction with KSRP, crucial details that were overlooked in previous reports (Hashem et al., 2013; Liu et al., 2016; Shalev-Benami et al., 2016; Zhang et al., 2016), and provide a new molecular framework for the development of safer anti-kinetoplastidian drugs.

RESULTS AND DISCUSSION

Characterization of 40S Ribosomal Complexes

We purified native 40S complexes from a lysate of *T. cruzi* epimastigotes by sucrose gradient fractionation (see STAR Methods) (Figures 1A and S1A) and imaged them by cryo-EM to an average resolution of 4.4 Å (Figures 1A–1C and S2; Table S1). Samples were also characterized by mass spectrometry (MS/MS) for determining the proteomic composition (Figures S1B and S3A). The cryo-EM map of native *T. cruzi* 40S subunits displays the presence of two unassigned densities located at the platform and the left foot, respectively (Figures 1B and 1C). The platform-bound density partially presents a helical shape and interacts with 18S rRNA residues belonging to h24 and h23 in proximity to the E site (Figure 1D). The second unassigned density is located at the bottom of the 40S left foot (Figure 1B), near ES6^S and ES3^S. Based on their shape and binding sites, almost exclusively composed of rRNA, we thought that these elements might correspond to RNA-binding proteins.

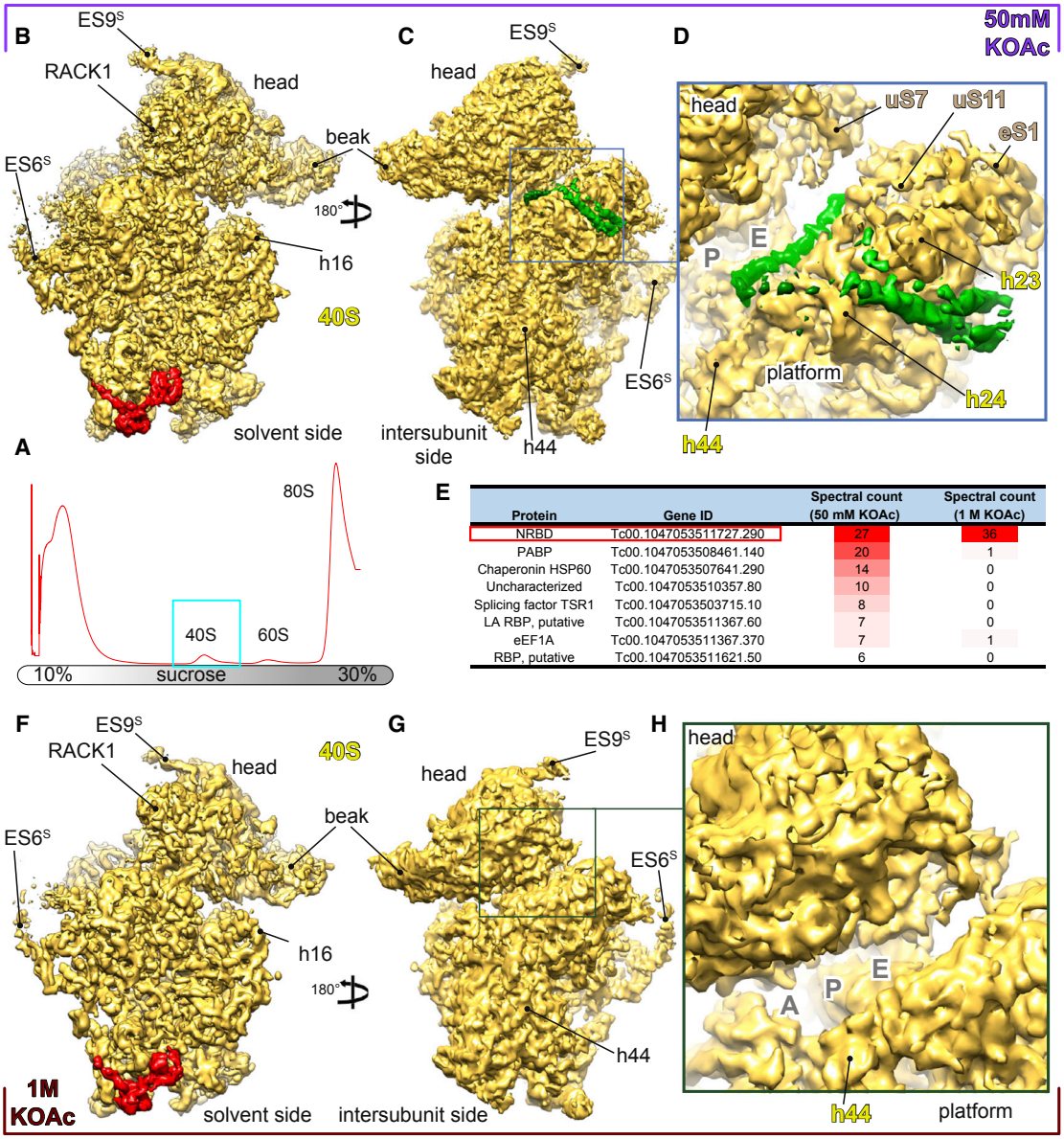
Classically, the characterization of most ribosome-interacting proteins has been achieved using biochemical approaches that include sucrose gradient fractionation, ribosomal salt washing, and peptide sequencing (Benne et al., 1976; Fleischer et al., 2006; Merrick et al., 1973). Hence, in an attempt to unravel the identity of the presumed proteins that would account for these densities, we purified native 40S ribosomal complexes in low-salt conditions (50 mM KOAc), similarly to the imaged complex (Figures 1B and 1C), then half of the sample was salt-washed with 1 M KOAc (see STAR Methods). Both low-salt and salt-washed 40S samples were subjected to size-exclusion chromatography (SEC, i.e. gel filtration), followed by MS/MS proteomic characterization. Finally, the salt-washed 40S sample was analyzed by cryo-EM.

MS/MS data from gel-filtrated native 40S subunits purified at low salt concentration (50 mM) revealed the presence of several non-ribosomal proteins, two of which were identified as putative RNA-binding proteins (NRBD/nuclear RNA-binding domain and PABP/Poly(A)-binding protein, Figures 1E and S3). Although our MS/MS analysis is semi-quantitative, based on the spectral counts, both RNA-binding proteins were present at nearly equimolar levels when compared with 40S r-proteins. On the other hand, in salt-washed 40S subunits, only the RNA-binding protein NRBD (accession number Q4DY30_TRYCC/Tc00.1047053511727.290) was present at equimolar levels when compared with the rest of the 40S r-proteins, indicating its tight association with the 40S ribosomal subunit (Figures 1E, red frame, and S3). Importantly, the cryo-EM map of the salt-washed 40S complex lacks the platform-bound density but retains the density located at the 40S left foot (Figures 1F–1H), potentially revealing the latter as corresponding to NRBD.

Ribosomal Location of NRBD, a Novel Kinetoplastid-Specific Ribosomal Protein

In *T. cruzi*, NRBD exists as two highly similar homologs, TcNRBD1 and TcNRBD2, which are synthesized from two different mRNAs arrayed in the same polycistronic transcription tandem (Oliveira et al., 2016). Notably, NRBD homologs are only encoded by kinetoplastid genomes, indicating that they are a kinetoplastid-specific protein family (Pitula et al., 2002). They contain three conserved motifs, a KPA-rich N-terminal tail, and two RNA recognition motifs (RRM) (Figure S4A). NRBD2 has an 8-amino-acid longer KPA-rich motif when compared with NRBD1. Although the TcNRBD proteins, as well as their *T. brucei* orthologs, TbP34/NRBD1 and TbP37/NRBD2 (Prohaska and Williams, 2009), have been previously linked to kinetoplastidian ribosome biogenesis and in particular maturation of the LSU, by binding to 5S rRNA and L5 (Prohaska and Williams, 2009; Umaer and Williams, 2015; Umaer et al., 2014), nevertheless several biochemical reports have shown that NRBD consistently co-purifies with cytoplasmic 40S, 80S, and polysomes (Alves et al., 2010; Ayub et al., 2009; Hellman et al., 2007; Klein et al., 2015; Oliveira et al., 2016; Umaer et al., 2014), suggesting that NRBD may be an integral kinetoplastid-specific r-protein.

Even though no unidentified densities possibly corresponding to a protein have been reported for the available kinetoplastid ribosome structures, a deep analysis of the high-resolution cryo-EM reconstruction of the 80S ribosome from *Leishmania*



Protein	Gene ID	Spectral count (50 mM KOAc)	Spectral count (1 M KOAc)
NRBD	Tc00.1047053511727.290	27	36
PABP	Tc00.1047053508461.140	20	1
Chaperonin HSP60	Tc00.1047053507641.290	14	0
Uncharacterized	Tc00.1047053510357.80	10	0
Splicing factor TSR1	Tc00.1047053503715.10	8	0
LA RBP, putative	Tc00.1047053511367.60	7	0
eEF1A	Tc00.1047053511367.370	7	1
RBP, putative	Tc00.1047053511621.50	6	0

Figure 1. Cryo-EM reconstructions of native and salt-washed 40S complexes from *T. cruzi*

(A) Absorbance profile from 10% to 30% sucrose gradient at 260 nm, taken from the top to the bottom, displaying peaks for 40S (light blue square), 60S, and 80S complexes.

(B and C) Cryo-EM reconstruction of native 40S complex (purified in 50 mM KOAc) from *T. cruzi* viewed from different orientations. Density segments correspond to 40S (yellow) and unassigned densities (red and green).

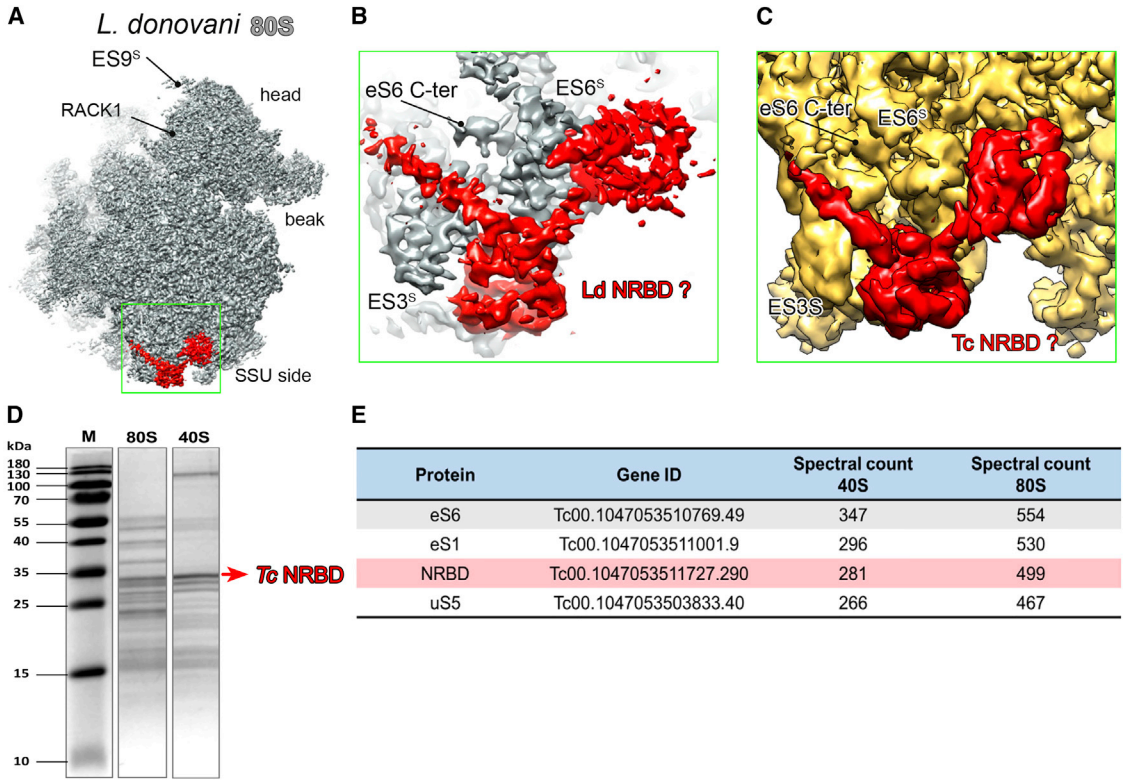
(D) Close-up view on the platform of the 40S, highlighting unassigned density.

(E) List of top-hit non-ribosomal proteins detected from native and salt-washed 40S by mass spectrometry. The intensity of the red background indicates the occupancy of the proteins in the sample compared to 40S r-proteins (red = equimolar to 40S r-proteins). NRBD was consistently detected with native and salt-washed (1 M KOAc) 40S complex.

(F and G) Cryo-EM reconstruction of salt-washed 40S complex from *T. cruzi* viewed from solvent and intersubunit sides shows the unassigned density (red) on the 40S left foot.

(H) Close-up view on the platform of salt-washed 40S, highlighting the absence of the additional densities at the platform. See also Figures S1–S3 and Table S1.

donovani (Zhang et al., 2016) revealed the presence of the same unassigned density on the left 40S subunit foot (Figures 2A–2C). Consistently, we located this density in the cryo-EM reconstruction of *T. brucei* (Hashem et al., 2013) and *T. cruzi* (Gao et al., 2005) 80S ribosome structures, previously solved at lower resolutions (Figures S5A and S5B). In the case of the *T. brucei* structure, residues 193–212 and 274–307 from ES3^S and 825–856 from ES6^S had been erroneously modeled into the density now attributed to NRBD. A closer look at this extra mass in the *L. donovani* structure revealed features that more likely



Protein	Gene ID	Spectral count 40S	Spectral count 80S
eS6	Tc00.1047053510769.49	347	554
eS1	Tc00.1047053511001.9	296	530
NRBD	Tc00.1047053511727.290	281	499
uS5	Tc00.1047053503833.40	266	467

Figure 2. Conservation of the Uncharacterized Density on the Left Foot of 40S Subunits among Kinetoplastids and Detection of NRBD in the 40S and 80S Ribosome

(A) Cryo-EM reconstruction of *L. donovani* 80S (Zhang et al., 2016), viewed from the SSU side. Density segments correspond to 80S (gray) and the unassigned density located on the foot (red).

(B) Close-up view on the *L. donovani* 40S left foot, highlighting an unassigned density (red) that could be attributed to a protein.

(C) Close-up view on the *T. cruzi* 40S left foot, highlighting the presence of the same density (red) found in *L. donovani* 80S reconstruction.

(D and E) SDS-PAGE gel of r-proteins from fractions of purified *T. cruzi* 40S subunits and 80S ribosomes. Mass spectrometry analysis of the ≈30 kDa gel band (marked with a red arrow) detected the NRBD protein at equimolar levels when compared with other ≈30 kDa 40S r-proteins. See also Figures S1, S3, and S5.

corresponded to α helices and β sheets rather than rRNA (Figures 2B and 2C). Overall, the segmented density displayed two globular domains that contact ES3^S and ES6^S, respectively, and a helical domain that interacts with the C-terminal tail of r-protein eS6 (Figures 2B and 2C). The organization with two RNA recognition motifs (RRM) matches that predicted for NRBD (Umaer and Williams, 2015). Furthermore, the theoretical molecular weight of this unassigned density was close to 30 kDa, which corresponds to the predicted molecular weight for NRBD based on SDS-PAGE analysis (Figures 2D, 2E, and S1A–S1B). Together, these observations suggested that this extra density may correspond to NRBD.

The MS/MS analysis consistently detected NRBD in the 40S subunits and 80S ribosomes from *Trypanosoma cruzi* and *Leishmania tarentolae* (Figures 2D, 2E, S1A–S1B, and S3). Furthermore, we confirmed the presence of *T. cruzi* NRBD (TcNRBD) in pure 80S ribosomes by an ELISA test and an immunoblot, in which a polyclonal antibody against TcNRBD (kindly provided by Dr. Goldenberg, Carlos Chagas Institute, Fiocruz, Curitiba, Brazil) (Oliveira et al., 2016) was used as the primary antibody (Figures S1C and S1D). To further confirm the location of NRBD on the kinetoplastid ribosomes, we obtained low-resolution immuno-labeled cryo-EM reconstructions of *T. cruzi* 80S ribosomes (~9 and ~11 Å after particle sorting) that were pre-incubated with the anti-TcNRBD antibody before imaging. These reconstructions show the presence of additional densities on the most exposed domain of this unassigned portion, which can be explained by the binding of the anti-TcNRBD antibody to NRBD at the left foot of 40S subunits (Figures S1E and S1F).

A homology-based atomic model of *L. donovani* NRBD (LdNRBD), derived from human PABP (see STAR Methods), fits well into this density (Figure 3). Thus, on the basis of our comprehensive structural and biochemical analysis, we attributed this density to NRBD, the first *bona fide* kinetoplastid-specific r-protein (NRBD/KSRP) unveiled so far. A comparative analysis of 80S ribosome cryo-EM maps from *T. cruzi* (this study), *T. brucei* (Hashem et al., 2013), and *L. donovani* (Zhang et al., 2016), as well as a sequence alignment of NRBD/KSRP from various species of kinetoplastids, demonstrate the general conservation of NRBD/KSRP among kinetoplastids (Umaer and Williams, 2015) (Figures S4A, 5A, and 5B). The kinetoplastidian specificity of NRBD/KSRP was further confirmed by examination of the human 80S ribosome (Zhang et al., 2016) at this location, showing the absence of such density but also the absence of the parts of ES3^S and ES6^S interacting directly with NRBD/KSRP (S 5C).

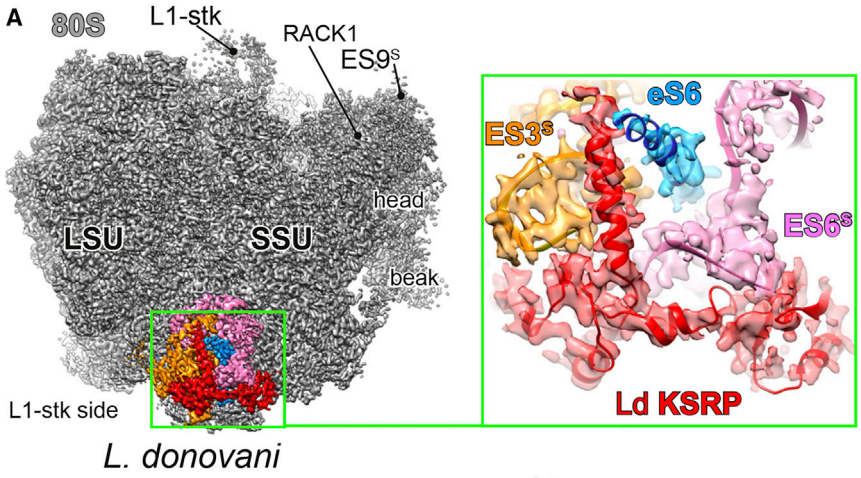


Figure 3. Fitting of the KSRP Homology Model into Its 80S Cryo-EM Segmented Density

(A) View of KSRP bound to the *L. donovani* 80S ribosome showing two globular domains and a helical C-ter that contact ES3^S (orange) and r-protein eS6 (blue). The close-up view (right panel) highlights the interactions of KSRP (red) with ES3^S (orange), ES6^S (pink), and r-protein eS6 (blue).

(B) Atomic model of KSRP (right panel), showing the C-terminal end colored in cyan, RRM1 in green, RRM2 in red, and the α-helical linker colored in yellow. The black dashed line on the left panel represents a most likely flexible 55-aa N-terminal tail (68-aa in TcKSRP), which remained unmodeled due to its lack of density.

Structure of KSRP in *Trypanosoma* and *Leishmania* Ribosomes

In order to compare KSRP's fine architecture between *L. donovani* and *T. cruzi*, we solved the cryo-EM structure of the *T. cruzi* 80S. Based on these two structures (*L. donovani* at ~3.0 Å and *T. cruzi* 80S at 4.0 Å), we refined homology models of *L. donovani* NRBD/KSRP (LdKSRP) and TcKSRP into their corresponding densities (Figures 4 and 5). Accordingly, we assigned and placed the core domains of NRBD/KSRP into the cryo-EM maps, including the two RRMs (1 and 2), a hinge central region, and the nuclear-export-signal (NES)-containing C-terminal α helix (Figures 3B and 4A–4D). A side view of NRBD/KSRP bound to the 40S subunit shows RRM1 and RRM2 contacting the rRNA belonging to ES6^S and ES3^S, and the larger C-terminal α helix protruding from the bottom toward the C-terminal tail of r-protein eS6 (Figures 3A, 3B, and 4A–4D). The lower resolution in the RRM1 vicinity suggests that this domain is more flexible. Although NRBD/KSRP is far away from the peptidyl-transferase center or the mRNA channel, it has been suggested that the r-proteins located at the ribosomal feet, such as eS6 and eL24, can be also implicated in mRNA translation regulation (Ben-Shem et al., 2011; Thiébeauld et al., 2009; Zhou et al., 2010).

The two RRM domains of NRBD/KSRP fold independently and adopt the canonical β1-α1-β2-β3-α2-β4 topology (Daubner et al., 2013) (Figure 3B). A structured and conserved short α-helical linker positions both RRMs to interact optimally with ES6^S and ES3^S rRNA. The structure of the linker between the RRMs may affect the recognition of their rRNA targets as shown for other RRM-containing proteins (Daubner et al., 2013; Lunde et al., 2007). Indeed, it has been reported that the potential of these proteins to accurately recognize RNA relies largely on the linker between the two RRM domains (Lunde et al., 2007). Importantly, we found that ribosome-associated TcNRBD/KSRP is phosphorylated at the highly conserved Ser149 (Ser136 in LdNRBD/KSRP), which is located within a solvent-exposed loop of the α-helical linker, close to a key contact point between RRM2 and ES6^S (Figures S4B–S4D). This post-translational modification could be involved in some of the conformational rearrangements at the interface between this protein and its rRNA target, suggesting a possible impact on the ribosome function. Globally, NRBD/KSRP most likely acts as an RNA chaperone that warrants some of the flexible segments of ES6^S and ES3^S rRNA, a specific and stable structure, i.e. NRBD/KSRP RRM1 and RRM2 interact with the latter ESs and piece them in the observed conformation (Hashem et al., 2013; Lunde et al., 2007; Zhang et al., 2016). This additional feature on the 40S left foot could have a specific function in mRNA translation that is yet to be elucidated by future studies. Consistently, a previous report showed that knockdown of TbNRBD/KSRP variants (TbP37/TbP34) leads to growth arrest and cell death, which was interpreted as probably due to observed defects on ribosome biogenesis and thus protein synthesis (Hellman et al., 2007).

KSRP Interactions with the Ribosome

Based on our atomic models of the 40S ES6^S, ES3^S, and NRBD/KSRP from *L. donovani* and *T. cruzi*, we have determined the residues involved in the interaction network at this kinetoplastid-specific ribosomal region of the 40S left foot (Figures 5A–5C and S4E). Overall, we observe an extended binding interface between NRBD/KSRP and the 40S. Indeed, at least 30 direct interactions were observed, including H bonds, hydrophobic interactions, and intricate stacking systems.

Generally, NRBD/KSRP resembles most of the canonical RRM-containing proteins, as it binds RNA via aromatic residues located on the β strands of its RRMs, which establish stacking interactions with bases or intercalate between sugar rings (Daubner et al., 2013). However, some peculiar exceptions to

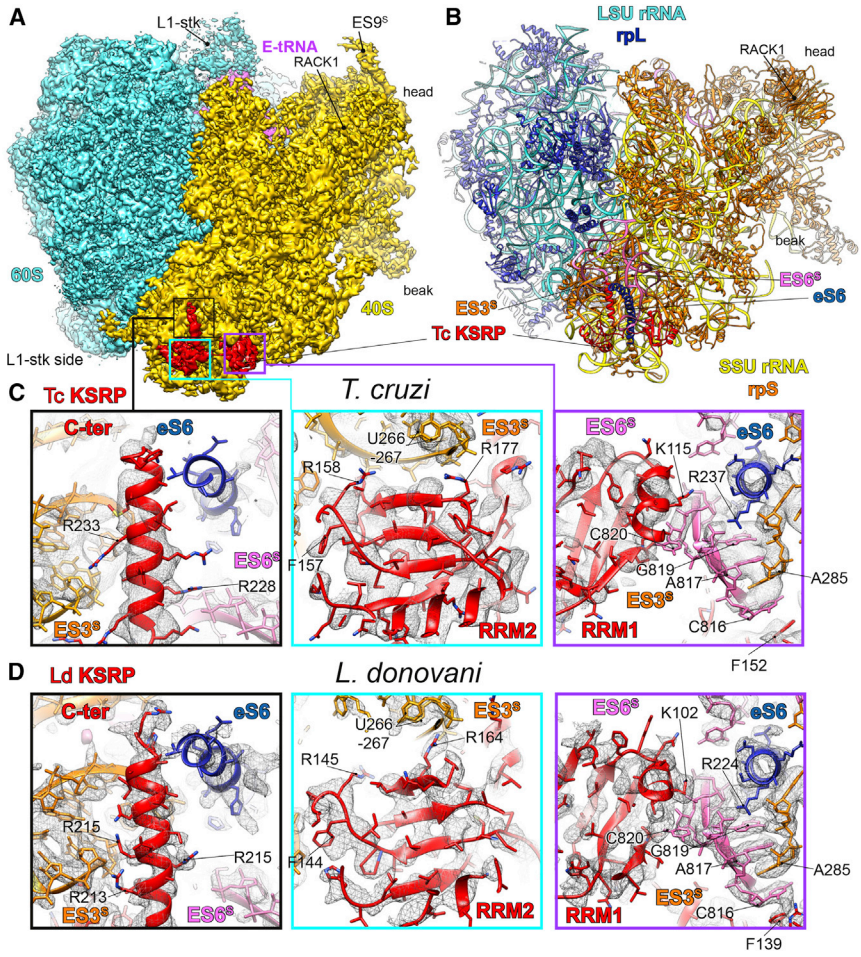


Figure 4. Architecture of Ribosome-Bound KSRP

(A) L1-stalk side view of the cryo-EM map of *T. cruzi* 80S, highlighting the 40S (yellow), 60S (cyan), E-tRNA (purple), and KSRP (red).

(B) Overview of the atomic model of *T. cruzi* 80S showing KSRP (red) bound to ES3^S (orange ribbon), ES6^S (pink ribbon), and r-protein eS6 (dark blue) at the left foot of 40S.

(C and D) Close-up views of the electron density (gray mesh) with molecular models for *T. cruzi* and *L. donovani* KSRP (red), ES3^S (orange), ES6^S (pink), and r-protein eS6 (dark blue), highlighting their interaction network within the ribosome and their consistency with the cryo-EM reconstructions. The blow-up squares highlight the main interactions in both Tc and Ld, respectively, of KSRP RRM2 with ES3^S (cyan frames), KSRP RRM1 with ES6^S (purple frames), and KSRP C-ter with eS6 C-ter, ES3^S, and ES6^S (black frames). The resolution of *L. donovani* 80S and to some extent *T. cruzi* cryo-EM maps (C and D) allowed accurate fitting of most side chains from the KSRP, ES3^S, ES6^S, and r-protein eS6. See also Figure S4.

this typical binding model have been reported (Daubner et al., 2013). For example, some RRM proteins can specifically recognize G triplets through conserved residues located within their inter-domain loops (Daubner et al., 2013). NRBD/KSRP interacts with rRNA via conserved residues located within both the β strands and the inter-domain loops of its RRMs (Figures 4C, 4D, 5A–5C, and S10). Accordingly, the highly conserved loop located between β2 and β3 of RRM2 contacts the junction between the stem loops 2A and 2 of ES3^S (Figures 5A–5C and S4E). These interactions occur between three key conserved amino acids (Y181, R182, and N184 in TcNRBD/KSRP) and a highly conserved CAC trinucleotide motif, similarly to RBPMS proteins, which were shown to be involved in RNA transport, localization, and stability (Teplova et al., 2016). Remarkably, the same key conserved tyrosine (Y181 in TcNRBD/KSRP and Y168 in LdNRBD/KSRP) located on the loop β2-β3 of RRM2 stacks with the A281 of ES^3 and positions three different regions of ES3^S into close proximity (Figure 5C, right panels). In addition, a highly conserved aromatic residue (F152 in TcNRBD/KSRP) on the β1 strand of RRM2 and surrounding hydrophobic tyrosines (Y185 and Y187 in TcNRBD/KSRP) on the β3 RRM2 strand form a docking site for the ES3^S/ES6^S base stack involving C816, A285, A817, and G819 (Figures 4C and 4D, purple frames, and Figure 5C, left panels). Another remarkable site of interactions involves three arginine side chains located on the β2 strand of RRM2, which bind to ES3^S (residues G263-U267) (Figures 4C and 4D, cyan frames, and Figure 5C, middle panels).

Although the interactions of the RRM1 of NRBD/KSRP are less extensive, but also less defined due to the lower local resolution of the map in that area, we observe that RRM1 recognizes the helix loop of domain ES6^S.B (named according to the previously reported nomenclature; Anger et al., 2013) via two conserved residues, K115 and R99 in TcNRBD/KSRP, located on the α2 helix and the loop β2-β3 of RRM1, respectively (Figures 5A, 5B, and S4E). As a result of this interaction, a specific conformation of ES6^S creates a protein/rRNA tunnel through which the C-terminal tail of r-protein eS6 extends (Figures 4C, 4D, purple frames, and S4C–S4E). The location, structure, and perhaps the function of NRBD/KSRP present similarities with the chloroplast-specific 30S r-proteins cS22 and cs23, which are also made of RRMs and which are located on the foot of the small subunit (Figure S5D; Bieri et al., 2016). Interestingly, cS22 was shown to possess RNA chaperone activity and to play a role in plant development (Xu et al., 2013). This suggests that ribosomes from completely different biological contexts have evolved similar rRNA/r-protein, pointing out a possible common function for cS22 and NRBD/KSRP.

As mentioned above, NRBD/KSRP was first identified as a protein that is important for nuclear export via binding to 5S rRNA and r-protein uL5. Previous data along with our work could suggest that 5S and 18S rRNAs compete for binding to KSRP through similar rRNA-binding surfaces to the two RRMs. It is difficult to assess this hypothesis at this stage; nevertheless the previously reported mutagenesis study (Kamina and Williams, 2017) may supply a valuable insight on NRBD/KSRP capacity to bind with its target rRNA. Thus, positively charged

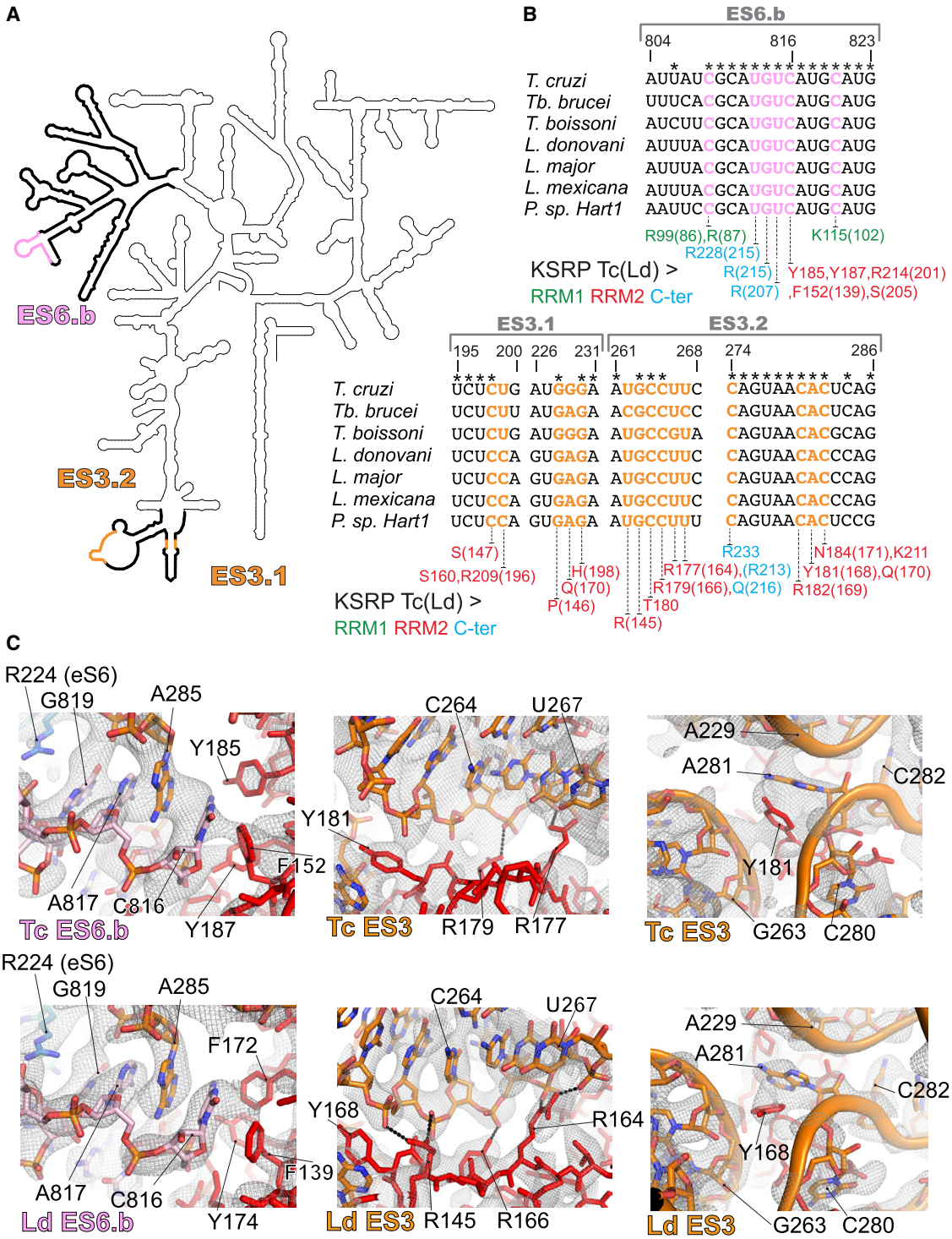


Figure 5. Detailed Ribosomal Interacting Network of KSRP

(A) Secondary structure of kinetoplastidian 18S rRNAs. The regions corresponding to ES3^S and ES6^S are highlighted in bold. The specific KSRP-binding regions are highlighted in orange (ES3^S) and pink (ES6^S).

(B) Alignment of several kinetoplastid 18S rRNA sequences showing the high conservation of KSRP interacting nucleotides within the ES6^S (upper panel) and ES3^S (bottom panel). The key KSRP interacting residues are shown on the bottom part of the alignment using the same color code as in Figures 4B and S4B.

(C) Close-up views highlighting key interactions between KSRP and ES6^S and ES3^S in the ribosome. See also Figure S4.

and aromatic residues within RRM1 and RRM2 that were identified to be important for binding to 5S/uL5 (Kamina and Williams, 2017) are in close proximity to the rRNA in our structure. The C-ter α helix of NRBD/KSRP also contacts kinetoplastid-specific domains of the $ES6^S$, $ES3^S$, and the C-ter α helix of r-protein eS6. The C-ter of NRBD/KSRP is rich in charged amino acids and is sandwiched between r-protein eS6 and $ES6^S$ on one side, and the minor groove of the ES3.2 helix formed by residues U262-U267 and U271-A275 on the other (Figures 4C, 4D, black frames, and S4E). R233 of TcNRBD/KSRP and R213 of LdNRBD/KSRP interact with the phosphate groups of C274 and U267, respectively, within the most N-ter region of the C-ter helix of NRBD/KSRP (Figures 5B and S4E). In addition to the interaction with $ES6^S$ and eS6, a conserved arginine residue, R228 in TcNRBD/KSRP and R215 in LdNRBD/KSRP, is within H-bonding distance to residues U813 and G814 from $ES6^S$, respectively (Figure S4E). According to our analysis, the first half of the helical C-ter of NRBD/KSRP is conserved among kinetoplastids and establishes similar interactions between charged residues on NRBD/KSRP and the bipartite rRNA target ($ES6^S$ on one side and $ES3^S$ on the other). In contrast, we noted some flexibility of the second half of the C-ter helix of NRBD/KSRP, which is supported by a lower sequence conservation in *T. cruzi* and *L. donovani* C-ters of NRBD/KSRP and their corresponding protein-binding partner (C-ter of eS6) (Figures S4C–S4E); thus this region is likely more flexible and may not contribute to binding specificity.

Conclusion

Overall, our results indicate that NRBD/KSRP is a kinetoplastid-specific r-protein that binds to a ribosomal region that is only conserved among these pathogens. Our cryo-EM structures therefore represent the first structures reported thus far of a kinetoplastid-specific r-protein bound to the ribosome. In addition, we show that NRBD/KSRP is phosphorylated at a key position between its two RRM domains. As NRBD/KSRP depletion has been previously shown to be lethal, this protein could play a pivotal role in ribosome integrity and consequently in mRNA translation regulation. The location of NRBD/KSRP may not only suggest its role as an rRNA chaperone but also as a scaffold for the binding of other regulatory proteins. In conclusion, our structural and biochemical study demonstrates the existence of a novel and unique ribosomal protein in kinetoplastids. The singularity and the importance of NRBD/KSRP in kinetoplastids highlight its potential as a new target for the development of safer and more specific anti-kinetoplastid therapeutic agents that specifically block mRNA translation in the pathogen, without affecting the host translational machinery.

STAR★METHODS

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

- KEY RECOURSES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHODS DETAILS
  - Kinetoplastids Cultures
  - Ribosome Isolation
  - Salt-Washing and Size-Exclusion Chromatography (SEC) of 40S Complexes
  - NanoLC-MS/MS Analysis
  - Characterization of Post-Translational Modifications of NRBD/KSRP
  - Anti-TcNRBD/KSRP ELISA
  - Immunoblot
  - Cryo-EM Data Acquisition
  - Immuno-Labeled Cryo-EM
  - Image Processing
  - Atomic Model
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and one table and can be found with this article online at https://doi.org/10.1016/j.str.2017.09.014.

AUTHOR CONTRIBUTIONS

J.B.Q., E.M.-M., and Y.H. designed, analyzed, and interpreted molecular and biochemical data experiments. J.B.Q., E.M.-M., and A.S. prepared samples. J.B.Q. performed the biochemical experiments and the immunochemical analysis. Y.H. prepared grids and carried out cryo-EM data processing. A.B., Q.V., and Y.H. performed the molecular modeling and the structural interpretation. E.M.-M., J.B.Q., and Y.H. wrote the manuscript. J.C. performed MS/MS analysis. All authors contributed to the discussion and the writing of the manuscript. Y.H. directed the research.

ACKNOWLEDGMENTS

We thank Christoph Diebolder, Ludovic Renault, and Julio Ortiz Espinoza (NeCEN, Leiden, Holland) as well as Jean-François Ménétret, Corinne Crucifix and Christine Ruhlmann (IGBMC, Strasbourg, France) for assistance with data acquisition. We also thank Dr. Samuel Goldenberg and Dr. Lysangela Alves (Instituto Carlos Chagas, Fiocruz-Paraná, Curitiba, Brazil) for providing the anti-TcNRBD polyclonal serum. We would like also to thank the High Performance Computing Center of the University of Strasbourg (funded by the Equipex Equip@Meso project) for IT support and the staff of the proteomic platform of Strasbourg-Esplanade for conducting the nanoLC-MS/MS analysis. The mass spectrometry instrumentation was granted from Université de Strasbourg (IdEx 2015 Equipement mi-lourd). This work was supported by the ANR grant @RAction program ANR CryoEM80S ANR-14-ACHN-0024 and the Labex: ANR-10-LABX-0036_NETRNA and benefits from funding from the state managed by the French National Research Agency as part of the Investments for the Future program (to Y.H.).

Received: June 8, 2017
Revised: July 20, 2017
Accepted: September 20, 2017
Published: October 26, 2017

REFERENCES

Abrishami, V., Zaldívar-Peraza, A., de la Rosa-Trevín, J.M., Vargas, J., Otón, J., Marabini, R., Shkolnisky, Y., Carazo, J.M., and Sorzano, C.O.S. (2013). A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs. Bioinformatics *29*, 2460–2468.

Abrishami, V., Vargas, J., Li, X., Cheng, Y., Marabini, R., Sorzano, C.Ó.S., and Carazo, J.M. (2015). Alignment of direct detection device micrographs using a robust optical flow approach. J. Struct. Biol. *189*, 163–176.

Adams, P.D., Afonine, P.V., Bunkóczi, 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.

Alves, L.R., Ávila, A.R., Correa, A., Holetz, F.B., Mansur, F.C.B., Manque, P.A., de Menezes, J.P.B., Buck, G.A., Krieger, M.A., and Goldenberg, S. (2010). Proteomic analysis reveals the dynamic association of proteins with translated mRNAs in *Trypanosoma cruzi*. Gene *452*, 72–78.

Anger, A.M., Armache, J.-P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., and Beckmann, R. (2013). Structures of the human and *Drosophila* 80S ribosome. Nature *497*, 80–85.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics *22*, 195–201.

Ayub, M.J., Atwood, J., Nuccio, A., Tarleton, R., and Levin, M.J. (2009). Proteomic analysis of the *Trypanosoma cruzi* ribosomal proteins. Biochem. Biophys. Res. Commun. *382*, 30–34.

Benne, R., Wong, C., Luedi, M., and Hershey, J.W. (1976). Purification and characterization of initiation factor IF-E2 from rabbit reticulocytes. J. Biol. Chem. *251*, 7675–7681.

Ben-Shem, A., Loubresse, N.G., de Melnikov, S., Jenner, L., Yusupova, G., and Yusupov, M. (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. Science *334*, 1524–1529.

Bieri, P., Leibundgut, M., Saurer, M., Boehringer, D., and Ban, N. (2016). The complete structure of the chloroplast 70S ribosome in complex with translation factor pY. EMBO J. *36*, 475–486.

Chappuis, F., Sundar, S., Hailu, A., Ghalib, H., Rijal, S., Peeling, R.W., Alvar, J., and Boelaert, M. (2007). Visceral leishmaniasis: what are the needs for diagnosis, treatment and control? Nat. Rev. Microbiol. *5*, S7–S16.

Chicher, J., Simonetti, A., Kuhn, L., Schaeffer, L., Hammann, P., Eriani, G., and Martin, F. (2015). Purification of mRNA-programmed translation initiation complexes suitable for mass spectrometry analysis. Proteomics *15*, 2417–2425.

Daubner, G.M., Cléry, A., and Allain, F.H. (2013). RRM–RNA recognition: NMR or crystallography…and new findings. Curr. Opin. Struct. Biol. *23*, 100–108.

Deo, R.C., Bonanno, J.B., Sonenberg, N., and Burley, S.K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell *98*, 835–845.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Field, M.C., Horn, D., Fairlamb, A.H., Ferguson, M.A.J., Gray, D.W., Read, K.D., De Rycker, M., Torrie, L.S., Wyatt, P.G., Wyllie, S., et al. (2017). Anti-trypanosomatid drug discovery: an ongoing challenge and a continuing need. Nat. Rev. Microbiol. *15*, 447.

Fleischer, T.C., Weaver, C.M., McAfee, K.J., Jennings, J.L., and Link, A.J. (2006). Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. Genes Dev. *20*, 1294–1307.

Gao, H., Ayub, M.J., Levin, M.J., and Frank, J. (2005). The structure of the 80S ribosome from *Trypanosoma cruzi* reveals unique rRNA components. Proc. Natl. Acad. Sci. USA *102*, 10206–10211.

Hashem, Y., Georges, A., des Fu, J., Buss, S.N., Jossinet, F., Jobe, A., Zhang, Q., Liao, H.Y., Grassucci, R.A., Bajaj, C., et al. (2013). High-resolution cryo-electron microscopy structure of the *Trypanosoma brucei* ribosome. Nature *494*, 385–389.

Hellman, K.M., Ciganda, M., Brown, S.V., Li, J., Ruyechan, W., and Williams, N. (2007). Two trypanosome-specific proteins are essential factors for 5S rRNA abundance and ribosomal assembly in *Trypanosoma brucei*. Eukaryot. Cell *6*, 1766–1772.

Kamina, A.D., and Williams, N. (2017). Non-canonical binding interactions of the RNA recognition motif (RRM) domains of P34 protein modulate binding within the 5S ribonucleoprotein particle (5S RNP). PLoS One *12*, e0177890.

Khatter, H., Myasnikov, A.G., Natchiar, S.K., and Klaholz, B.P. (2015). Structure of the human 80S ribosome. Nature *520*, 640–645.

Klein, C., Terrao, M., Gil, D.I., and Clayton, C. (2015). Polysomes of *Trypanosoma brucei*: association with initiation factors and RNA-binding proteins. PLoS One *10*, e0135973.

Kucukelbir, A., Sigworth, F.J., and Tagare, H.D. (2014). Quantifying the local resolution of cryo-EM density maps. Nat. Methods *11*, 63–65.

Liu, Z., Gutierrez-Vargas, C., Wei, J., Grassucci, R.A., Ramesh, M., Espina, N., Sun, M., Tutuncuoglu, B., Madison-Antenucci, S., Woolford, J.L., et al. (2016). Structure and assembly model for the *Trypanosoma cruzi* 60S ribosomal subunit. Proc. Natl. Acad. Sci. USA *113*, 12174–12179.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479–490.

Melnikov, S., Ben-Shem, A., Yusupova, G., and Yusupov, M. (2015). Insights into the origin of the nuclear localization signals in conserved ribosomal proteins. Nat. Commun. *6*, 7382.

Merrick, W.C., Lubsen, N.H., and Anderson, W.F. (1973). A ribosome dissociation factor from rabbit reticulocytes distinct from initiation factor M3. Proc. Natl. Acad. Sci. USA *70*, 2220–2223.

Moazed, D., and Noller, H.F. (1987). Interaction of antibiotics with functional sites in 16S ribosomal RNA. Nature *327*, 389–394.

Neto, O.P., de, M., Reis, C.R.S., Moura, D.M.N., Freire, E.R., and Carrington, M. (2016). Unique and conserved features of the protein synthesis apparatus in parasitic trypanosomatid (*Trypanosoma* and *Leishmania*) species. In Evolution of the Protein Synthesis Machinery and Its Regulation, G. Hernández and R. Jagus, eds. (Springer), pp. 435–475.

Oliveira, C., Carvalho, P.C., Alves, L.R., and Goldenberg, S. (2016). The role of the *Trypanosoma cruzi* TcNRBD1 protein in translation. PLoS One *11*, e0164650.

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.

Pintilie, G.D., Zhang, J., Goddard, T.D., Chiu, W., and Gossard, D.C. (2010). Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. J. Struct. Biol. *170*, 427–438.

Pitula, J., Ruyechan, W.T., and Williams, N. (2002). Two novel RNA binding proteins from *Trypanosoma brucei* are associated with 5S rRNA. Biochem. Biophys. Res. Commun. *290*, 569–576.

Prohaska, K., and Williams, N. (2009). Assembly of the *Trypanosoma brucei* 60S ribosomal subunit nuclear export complex requires trypanosome-specific proteins P34 and P37. Eukaryot. Cell *8*, 77–87.

Querido, J.F.B., Agirre, J., Marti, G.A., Guérin, D.M.A., and Silva, M.S. (2013). Inoculation of *Triatoma* virus (Dicistroviridae: Cripavirus) elicits a non-infective immune response in mice. Parasit. Vectors *6*, 66.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. *192*, 216–221.

de la Rosa-Trevín, J.M., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J.M., and Sorzano, C.O.S. (2013). Xmipp 3.0: an improved software suite for image processing in electron microscopy. J. Struct. Biol. *184*, 321–328.

de la Rosa-Trevín, J.M., Quintana, A., del Cano, L., Zaldívar, A., Foche, I., Gutiérrez, J., Gómez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V., et al. (2016). Scipion: a software framework toward integration, reproducibility and validation in 3D electron microscopy. J. Struct. Biol. *195*, 93–99.

Scheres, S.H.W. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530.

Shalev-Benami, M., Zhang, Y., Matzov, D., Halfon, Y., Zackay, A., Rozenberg, H., Zimmerman, E., Bashan, A., Jaffe, C.L., Yonath, A., et al. (2016). 2.8-Å cryo-EM structure of the large ribosomal subunit from the eukaryotic parasite *Leishmania*. Cell Rep. *16*, 288–294.

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.

Sorzano, C.O.S., Alcorlo, M., de la Rosa-Trevín, J.M., Melero, R., Foche, I., Zaldívar-Peraza, A., del Cano, L., Vargas, J., Abrishami, V., Otón, J., et al. (2015). Cryo-EM and the elucidation of new macromolecular structures: random Conical Tilt revisited. Sci. Rep. *5*, 14290.

Teplova, M., Farazi, T.A., Tuschl, T., and Patel, D.J. (2016). Structural basis underlying CAC RNA recognition by the RRM domain of dimeric RNA-binding protein RBPMS. Q. Rev. Biophys. *49*, e1.

Thiébeauld, O., Schepetilnikov, M., Park, H.-S., Geldreich, A., Kobayashi, K., Keller, M., Hohn, T., and Ryabova, L.A. (2009). A new plant protein interacts with eIF3 and 60S to enhance virus-activated translation re-initiation. EMBO J. *28*, 3171–3184.

Umaer, K., and Williams, N. (2015). Kinetoplastid specific RNA-protein interactions in *Trypanosoma cruzi* ribosome biogenesis. PLoS One *10*, e0131323.

Umaer, K., Ciganda, M., and Williams, N. (2014). Ribosome biogenesis in African trypanosomes requires conserved and trypanosome-specific factors. Eukaryot. Cell *13*, 727–737.

Wilson, D.N. (2014). Ribosome-targeting antibiotics and mechanisms of bacterial resistance. Nat. Rev. Microbiol. *12*, 35–48.

Xu, T., Lee, K., Gu, L., Kim, J.-I., and Kang, H. (2013). Functional characterization of a plastid-specific ribosomal protein PSRP2 in *Arabidopsis thaliana* under abiotic stress conditions. Plant Physiol. Biochem. *73*, 405–411.

Zhang, X., Lai, M., Chang, W., Yu, I., Ding, K., Mrazek, J., Ng, H.L., Yang, O.O., Maslov, D.A., and Zhou, Z.H. (2016). Structures and stabilization of kinetoplastid-specific split rRNAs revealed by comparing leishmanial and human ribosomes. Nat. Commun. *7*, 13223.

Zhou, F., Roy, B., and von Arnim, A.G. (2010). Translation reinitiation and development are compromised in similar ways by mutations in translation initiation factor eIF3h and the ribosomal protein RPL24. BMC Plant Biol. *10*, 193.

STAR★METHODS

KEY RECOURSES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-KSRP/NRBD	Gift of Dr. Samuel Goldenberg (Oliveira et al., 2016)	N/A
Goat anti-rabbit IgG (whole molecule) - peroxidase	Sigma	Cat#A0545
Amersham ECL Rabbit IgG, HRP-linked whole Ab (from donkey)	GE Healthcare	Cat#NA934V
Chemicals		
Protease inhibitor cocktail tablets	Roche	Cat#11873580001
RNasin® Ribonuclease Inhibitors	Promega	Cat#N251B
Cell lines		
Trypanosoma cruzi strain Y	This paper	N/A
Leishmania tarentolae strain T7-TR	Jena Bioscience	Cat#LT-110
Commercial Assays		
Amersham ECL Prime Western Blotting Detection Reagent	GE Healthcare	Cat#RPN2232
Deposited Data		
Structure of *T. cruzi* 80S	This paper, deposited at EMdatabank	EMDB- 3844
Structure of *T. cruzi* 40S	This paper, deposited at EMdatabank	EMDB- 3845
Structure of *T. cruzi* KSRP	This paper, deposited at PDB	PDB: 5OPT
Structure of *L. donovani* KSRP	This paper, deposited at PDB	PDB: 5OSG
Software and Algorithms		
SCIPION	(Abrishami et al., 2013)	N/A
Xmipp3	de la Rosa-Trevín et al., 2013)	N/A
CTFFIND4	(Rohou and Grigorieff, 2015)	N/A
RELION	(Scheres, 2012)	N/A
UCSF Chimera	(Pettersen et al., 2004)	N/A
SEGGER	(Pintilie et al., 2010)	N/A
Coot v. 0.8.2	(Emsley et al., 2010)	N/A
Phenix v. dev-2474	(Adams et al., 2010)	N/A
RESMAP	(Kucukelbir et al., 2014)	N/A
ChemiDoc Imaging Systems	Bio-Rad	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents should be direct to Yaser Hashem (y.hashem@ibmc-cnrs.unistra.fr).

METHODS DETAILS

Kinetoplastids Cultures

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

Ribosome Isolation

T. cruzi and *L. tarentolae* 40S subunits and 80S ribosomes were purified from cell suspension lysates. Briefly, parasites were harvested and lysed 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). The cleared supernatant was layered onto 10-30 % (w/v) sucrose gradients and centrifuged (35 000 rpm, 5h30min, 4°C) using an SW41 Ti rotor (Beckman-Coulter). The fractions containing 40S subunits and 80S ribosomes

were collected and pooled according the UV absorbance profile. Buffer was exchanged by pelleting ribosomal complexes and re-suspending them in sucrose-free buffer II (10 mM HEPES-KOH pH 7.4, 50 mM KOAc, 10 mM NH_4Cl, 5 mM $Mg(OAc)_2$, and 2 mM DTT).

Salt-Washing and Size-Exclusion Chromatography (SEC) of 40S Complexes

40S complexes were purified as described above and resuspended in buffer II containing either 50 mM or 1M KOAc. Both samples were further incubated for 1h at 4°C before SEC. SEC was performed at 0.05 or 1 M salt-conditions using a Superose 6/300 column (GE Healthcare) coupled to an 1260 Infinity Capillary LC system (Agilent Systems). The samples were analyzed by 15% SDS-PAGE gels and mass spectrometry.

NanoLC-MS/MS Analysis

The mass spectrometry (MS/MS) analysis of liquid *L. tarentolae* and *T. cruzi* 40S and 80S complexes were performed as described before(Chicher et al., 2015). Briefly, after precipitation, a reduction and alkylation, the proteins were digested overnight with trypsin (Promega). The peptides generated were injected on nanoLC-MS/MS (nano HPLC Easy nLC 1000 coupled to a Thermo-Scientific Qexactive Plus mass spectrometer). The data were searched against the *T. cruzi* and *L. major* proteome sets from the UniProt database. The identification of the peptides was performed with Mascot algorithm (version 2.5, Matrix Science, London, UK) through Proline Software (ProFi; http://proline.profiproteomics.fr/), validated with a minimum score of 25, and with a peptide and spectrum false discovery rate FDR<1%. A search on TriTrypDB (http://tritrypdb.org/tritrypdb/) complemented the pipeline.

Characterization of Post-Translational Modifications of NRBD/KSRP

T. cruzi 80S ribosomes were purified as described above except that all buffers were supplemented with phosphatase inhibitors (60 mM β-glycerol phosphate, 10 mM NaF and 2 mM $NaVO_3$). The ribosomal proteins were then separated by 15% SDS-PAGE gels and the ≈30 kDa gel band was subjected to in-gel trypsin digestion and analyzed by nanoLC-MS/MS. The phosphorylation site identified by Mascot was also validated manually on the spectra.

Anti-TcNRBD/KSRP ELISA

To investigate whether NRBD/KSRP was accessible for antibody binding in the context of the 80S ribosome, an enzyme-linked immunosorbent assay (ELISA) was performed as described previously(Querido et al., 2013). 400 ng/well of *T. cruzi* 80S, diluted in bicarbonate buffer (pH 8.5) was adsorbed overnight at 4°C onto a 96-well ELISA micro-plate. Bovine serum albumin (BSA) was used as a negative control. Rabbit polyclonal serum anti-TcNRBD1/KSRP, kindly offered by Dr. Samuel Goldenberg and Dr. Lysangela Alves (Instituto Carlos Chagas, Fiocruz-Paraná, Curitiba, Brazil), was used as a primary antibody(Oliveira et al., 2016). The samples were incubated with the anti-TcNRBD1/KSRP polyclonal sera (1:800 dilution) for 1h at room temperature, and washed for 5 times using PBS-0.05% Tween 20. Then, samples were incubated with goat anti-rabbit IgG (whole molecule)-peroxidase (Sigma; 1:5000 dilution) for 1h at room temperature. To detect the antibody-antigen complex, the samples were washed five times with PBS-0.05% Tween 20, and incubated with 3,3′,5,5′-Tetramethylbenzidine (TMB) Liquid Substrate (Sigma) for 30 min at room temperature. H_2SO_4 1 N was used to stop the reaction, followed by OD reading at 415 nm.

Immunoblot

80S ribosome was purified according to the protocol described above. 15 μg of ribosome and 15 μg of pure BSA (negative control) were resolved by 15% SDS-PAGE gel and transferred to Trans-Blot Turbo Mini PVDF membranes (Bio-Rad). Membrane was washed with PBS and blocked overnight with PBS containing 1 % skimmed milk, followed by an incubation with rabbit anti-*T. cruzi* NRBD polyclonal antibody (1:500) for 3 hours at room temperature. After the incubation, the membrane was washed, followed by an incubation with peroxidase-conjugated anti-rabbit IgG (1:10,000; GE Healthcare) for 1h at room temperature. The antibody-antigen complex signal was detected by chemiluminescence (ECL™ Western Blotting Detection Reagents; GE Healthcare), and the membrane was scanned using ChemiDoc Imaging Systems (Bio-Rad). Scanned image was analysed using Image Lab™ Software (Bio-Rad).

Cryo-EM Data Acquisition

The grids were prepared by applying 4 μL of 120 nM 40S or 80S to 400 mesh holey carbon Quantifoil 2/2 grids (Quantifoil Micro Tools), as described before(Simonetti et al., 2016). The grids were blotted for 1.5 sec at 4°C, 100% humidity, using waiting time 30 s, and blot force 5 (Vitrobot Mark IV). The data acquisitions were performed on a Titan Krios S-FEG instrument (FEI) operated at 300 kV acceleration voltage and at a nominal underfocus of $\Delta z = -0.6$ to -4.0 μm using the second-generation back-thinned direct electron detector CMOS (Falcon II) 4,096 x 4,096 camera and automated data collection with EPU software (FEI). The Falcon II camera was calibrated at nominal magnification of 75,000 x. The calibrated magnification on the 14 μm pixel camera is 134,615 x, resulting in 1.04 Å pixel size at the specimen level. The camera was set up to collect 22 frames and frames 2 to 13 where aligned. Total exposure was 1.5 s, with a dose of 60 /$Å^2$ (or 2.2 /$Å^2$ per frame).

Immuno-Labeled Cryo-EM

T. cruzi 80S-antibody complex was prepared in 50 μl total volume, containing 120 nM 80S and polyclonal serum anti-TcNRBD/KSRP (1:28 dilution). The mix was incubated at room temperature for 1h before grid preparation. 4 μL of 80S-antibody complex was applied

to 400 mesh holey carbon Quantifoil 2/2 grids (Quantifoil Micro Tools), blotted for 1.5 sec at 4°C, 100% humidity, using waiting time 30 s, and blot force 5 (Vitrobot apparatus Mark IV). Cryo-EM data were collected on a Polara Tecnai F30 cryo-transmission electron microscope (FEI instruments) operated at 300 KeV acceleration voltage and at a nominal underfocus of $\Delta z = -0.6$ to -4.0 μm, using a direct electron detector CMOS (Falcon I) 4,096 × 4,096 camera calibrated at a nominal magnification of 59,000 x, resulting in a pixel size of 1.815 Å.

Image Processing

SCIPION(Abrishami et al., 2015; de la Rosa-Trevín et al., 2016; Sorzano et al., 2015) package –integrates several data processing software– was used for image processing and 3D reconstruction. Optical Flow algorithm integrated in Xmipp3(de la Rosa-Trevín et al., 2013) was used for the movie alignment of ~3600 images from the low-salt 40S (native) sample, ~1140 images for the high-salt 40S sample (salt-washed) and 2350 images for the 80S sample (all from Tc). CTFFIND4(Rohou and Grigorieff, 2015) was used for the estimation of defocus for particles selection. Using SCIPION(Abrishami et al., 2013), approximately 413,500 particles were selected for the native 40S, 208,000 particles for the salt-washed 40S and 183,500 particles for 80S. RELION(Scheres, 2012) was used for particle sorting through 3D classification via SCIPION, refer to Extended Data Figure S2 for particle sorting details for all three complexes. Selected classes were refined using RELION's 3D autorefine, and the final refined classes were then post-processed using the procedure implemented in RELION applied to the final maps for appropriate masking, B factor sharpening, and resolution validation to avoid over-fitting(Scheres, 2012), indicating an average resolution of 4.4 Å for the native 40S complex, 6.2 Å for the salt-washed one and 4 Å for 80S ribosome (Extended Data Figure S2).

Regarding the 80S-antiTcNRBD complex, ~1400 images were collected on an FEI Polara microscope, yielding ~370,000 particles. The SCIPION package was also used to process the collected dataset. After a first particle-sorting step in RELION, a major class displaying some residual presence of an additional mass on the 80S left foot emerged (~61% of the total particles, see Extended Data Figure S2 for details). A large oval soft mask was applied around NRBD/KSRP region to further classify the obtained major class by focusing the particle sorting on the presence or absence of the antibody on NRBD/KSRP. We observed that only nearly 20.5% of the picked particles display densities of the antibody, represented by two classes. This modest apparent occupancy for the anti-body is probably due to the fact that the used anti-TcNRBD antibody is polyclonal and only a fraction of NRBD/KSRP is exposed to the solvent. We used the particles from these classes to perform a last 3D auto refinement in RELION. The final reconstructions of anti-NRBD/80S assemblies include ~63,000 and ~11,000 particles for Ab_C1 and Ab_C2 classes with resolutions of 9 and 11 Å, respectively.

Segmentation and Display of Density Maps

The cryo-EM maps were segmented by UCSF Chimera(Pettersen et al., 2004) using the SEGGER module(Pintilie et al., 2010). All the segments counting fewer than 1,000 voxels were discarded.

Atomic Model

Atomic models of the large ribosomal subunit or the entire 80S ribosome from *Trypanosoma cruzi* and *Leishmania donovani* were used to derive the atomic models of the *T. cruzi* ribosomal complexes. Near-complete atomic models of NRBD/KSRP from *T. cruzi* and *L. donovani* were derived by homology (32% identity) to the crystal structures of human PABP (Deo et al., 1999) using SWISS-MODEL (Arnold et al., 2006). After the construction of the atomic models, NRBD/KSRP was placed into its corresponding cryo-EM density map segment by rigid- body fitting using Chimera. The coordinates from the *Leishmania donovani* 80S ribosome (PDB ID 5T2A) (Zhang et al., 2016) were used as the starting model for correcting, building and fitting expansion segment $ES6^S$ (residues 700–770; 800–824; 977–1095) in the deposited 2.9 Å map, using Chimera v. 1.11.1(Pettersen et al., 2004) and Coot v. 0.8.2(Emsley et al., 2010). NRBD/KSRP was modelled within a carved-out box (generated using 'phenix.map_box') containing coordinates and density for NRBD/KSRP and surrounding residues (< 15 Å) from $ES3^S$ and $ES6^S$. Fitting of RRM1 and RRM2 was manually adjusted in Chimera (using 'torsions') and Coot (using 'real space refine zone', 'regularize zone' with appropriate secondary structure restraints). This model was refined in Phenix with simulated annealing (starting temperature = 450 K) and local grid search for local rotamer fitting. The resulting *L. donovani* 40S structure including NRBD/KSRP was used as a template for building the structure of the *T. cruzi* 40S in the 4.0 Å map of the *T. cruzi* 80S. The *T. cruzi* model was refined in real space using Phenix v. dev-2474 (Adams et al., 2010), with simulated annealing (starting temperature = 450 K), global minimization (5-10 macro-cycles), as well as RNA and protein secondary structure restraints (search_method = from_ca) (Table S1).

Local Resolution Measurement of the cryo-EM Map

RESMAP(Kucukelbir et al., 2014) was used to estimate the local resolution of the cryo-EM reconstruction. The resolution was represented as a variable color scale using UCSF Chimera (Pettersen et al., 2004).

DATA AND SOFTWARE AVAILABILITY

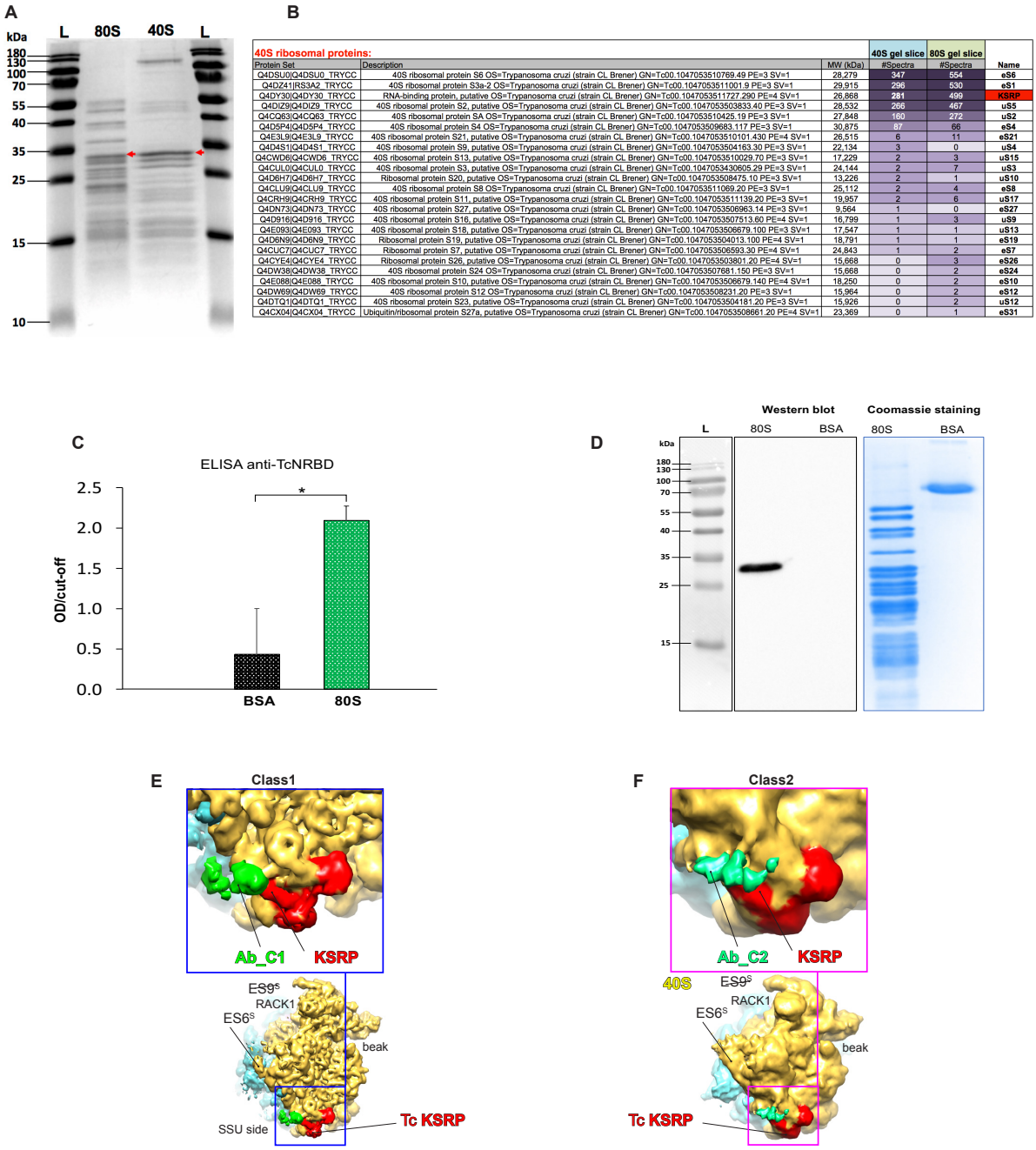
The accession numbers of the atomic models of *T. cruzi* and *L. donovani* KSRP reported in this paper are PDB: 5OPT and 5OSG respectively. The accession numbers of the cryo-EM maps of *T. cruzi* 80S and 40S reported in this paper are EMDB: 3844 and 3845, respectively.

Structure, Volume *25*

Supplemental Information

The cryo-EM Structure of a Novel 40S Kinetoplastid-Specific Ribosomal Protein

Jailson Brito Querido, Eder Mancera-Martínez, Quentin Vicens, Anthony Bochler, Johana Chicher, Angelita Simonetti, and Yaser Hashem



40S ribosomal proteins:

Protein Set	Description	MW (kDa)	40S gel slice #Spectra	80S gel slice #Spectra	Name
Q4DSU0|Q4DSU0_TRYCC	40S ribosomal protein S6 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053510769.49 PE=3 SV=1	28,279	347	554	eS6
Q4DZ41|RS3A2_TRYCC	40S ribosomal protein S3a-2 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511001.9 PE=3 SV=1	29,915	296	530	eS1
Q4DY30|Q4DY30_TRYCC	RNA-binding protein, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511727.290 PE=4 SV=1	26,868	281	499	KSRP
Q4DIZ9|Q4DIZ9_TRYCC	40S ribosomal protein S2, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053503833.40 PE=3 SV=1	28,532	266	467	uS5
Q4CQE9|Q4CQE9_TRYCC	40S ribosomal protein SA OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053510425.19 PE=3 SV=1	27,848	160	272	uS2
Q4D5P4|Q4D5P4_TRYCC	40S ribosomal protein S4 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053509683.117 PE=3 SV=1	30,875	87	66	eS4
Q4E3L9|Q4E3L9_TRYCC	40S ribosomal protein S21, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053510101.430 PE=4 SV=1	26,515	6	11	eS21
Q4D4S1|Q4D4S1_TRYCC	40S ribosomal protein S9, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053504163.30 PE=3 SV=1	22,134	3	0	uS4
Q4CWD6|Q4CWD6_TRYCC	40S ribosomal protein S13, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053510029.70 PE=3 SV=1	17,229	2	3	uS15
Q4CUL0|Q4CUL0_TRYCC	40S ribosomal protein S3, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053430605.29 PE=3 SV=1	24,144	2	7	uS3
Q4D6H7|Q4D6H7_TRYCC	Ribosomal protein S20, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508475.10 PE=3 SV=1	13,226	2	1	uS10
Q4CLU9|Q4CLU9_TRYCC	40S ribosomal protein S8 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511069.20 PE=3 SV=1	25,112	2	4	eS8
Q4CRH9|Q4CRH9_TRYCC	40S ribosomal protein S11, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511139.20 PE=3 SV=1	19,957	2	6	uS17
Q4DN73|Q4DN73_TRYCC	40S ribosomal protein S27, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506963.14 PE=3 SV=1	9,564	1	0	eS27
Q4D916|Q4D916_TRYCC	40S ribosomal protein S16, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053507513.60 PE=4 SV=1	16,799	1	3	uS9
Q4E093|Q4E093_TRYCC	40S ribosomal protein S18, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506679.100 PE=3 SV=1	17,547	1	1	uS13
Q4D6N9|Q4D6N9_TRYCC	Ribosomal protein S19, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053504013.100 PE=4 SV=1	18,791	1	1	eS19
Q4CUC7|Q4CUC7_TRYCC	Ribosomal protein S7, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506593.30 PE=4 SV=1	24,843	1	2	eS7
Q4CYE4|Q4CYE4_TRYCC	Ribosomal protein S26, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053503801.20 PE=4 SV=1	15,668	0	3	eS26
Q4DW38|Q4DW38_TRYCC	40S ribosomal protein S24 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053507681.150 PE=3 SV=1	15,668	0	2	eS24
Q4E088|Q4E088_TRYCC	40S ribosomal protein S10, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506679.140 PE=4 SV=1	18,250	0	2	eS10
Q4DW69|Q4DW69_TRYCC	40S ribosomal protein S12 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508231.20 PE=3 SV=1	15,964	0	2	eS12
Q4DTQ1|Q4DTQ1_TRYCC	40S ribosomal protein S23, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053504181.20 PE=3 SV=1	15,926	0	2	uS12
Q4CX04|Q4CX04_TRYCC	Ubiquitin/ribosomal protein S27a, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508661.20 PE=4 SV=1	23,369	0	1	eS31

Fig. S1. Identification of NRBD/KSRP. Related to Figure 1. (A), Five µg of either 80S or 40S were loaded in 15% SDS-PAGE gels. ~30kDa bands from 40S or 80S samples were individually excised from the gel and characterized by nanoLC-MS/MS **(B)**. **(C)**, The binding of anti-TcNRBD/KSRP antibody to the NRBD/KSRP in context of 80S complex was assayed by an indirect ELISA (enzyme-linked immunosorbent assay), using *T. cruzi* 80S as antigen source. BSA (bovine serum albumin) was used as negative control (NC). The cut-off value was defined using mean OD + 2SD (standard deviation) of the NC. The reported values represent the mean (+/- SD) of three experimental replicates. A Student's t-test indicates that there is a statistically significant difference between the mean signal for the NC and 80S ribosomes (P value = 0.03, P< 0.05). Error bars represent the standard error of the mean **(D)**, Western blot analysis of *T. cruzi* 80S ribosome, using antibody anti-NRBD/KSRP (rabbit IgG; 1:500). **(E–F)**, Immuno-labeled cryo-EM reconstructions of two different classes of *T. cruzi* 80S-antibody complex viewed from the solvent side, after particle sorting (see Methods). These two classes are different in the binding site of the antibody on KSRP. Density segments correspond to 40S (yellow), 60S (cyan), KSRP (red) and antibody (green). Both antibody classes (Ab_C1 and Ab_C2) interact with the most exposed domain of NRBD/KSRP, located on the solvent face of the small subunit, at slightly different positions.

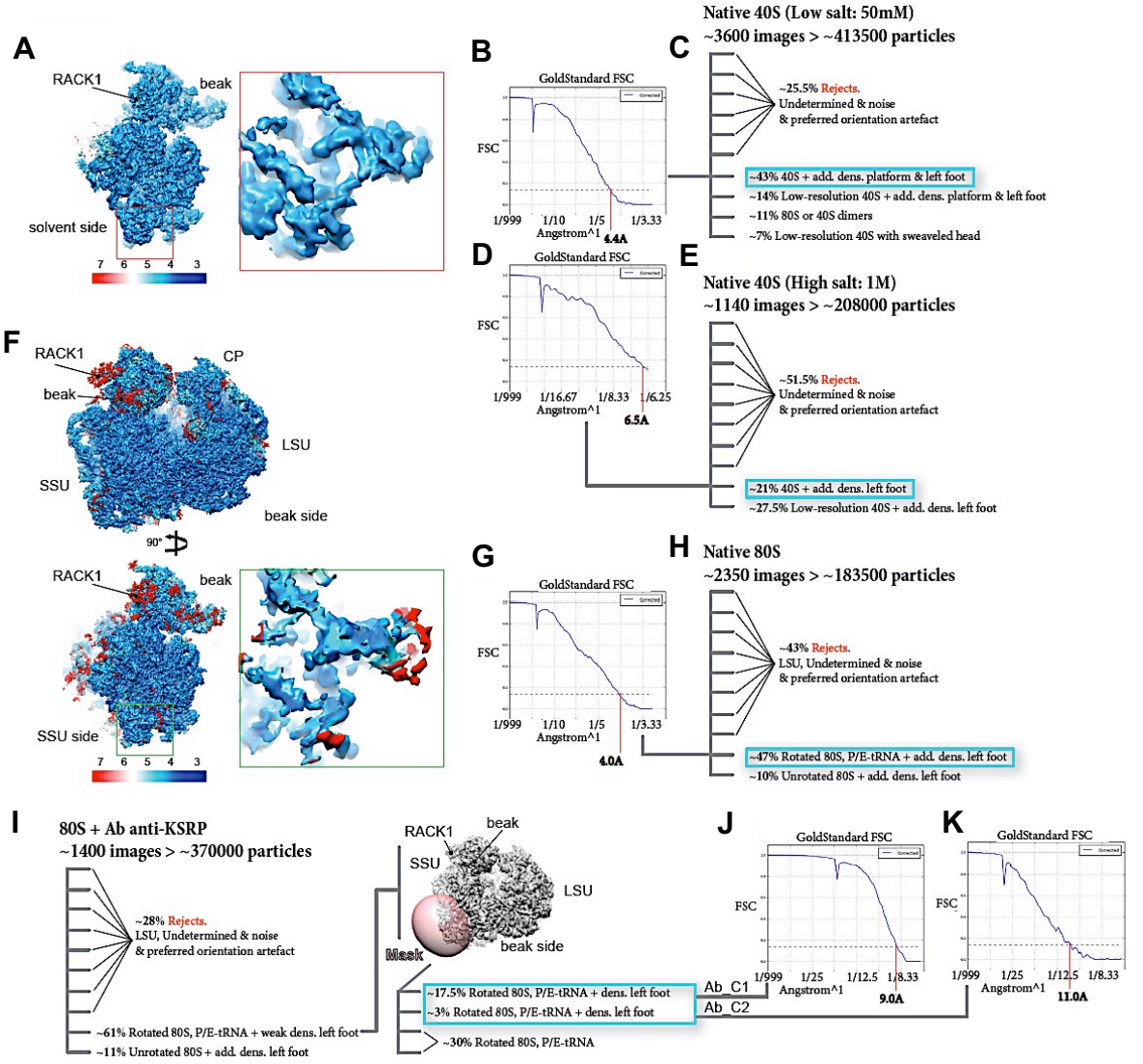


Fig. S2. Particle sorting of the cryo-EM images and the local and average resolutions of reported ribosomal complexes, Related to Figure 1. (A), Local resolution of the cryo EM map from *T. cruzi* native 40S subunits purified at low-salt viewed from solvent side. The map is colored according to the local resolution as estimated by RESMAP. Close-up view highlights the local resolution of the additional density on 40S left foot (~4.5Å). **(B),** Gold-standard FSC curve for the three-dimensional reconstruction of the latter class marking the corresponding resolution in angstrom (4.4 Å). **(C),** Overview of the particle sorting of the native low-salt 40S sample. The final reconstruction was performed from ~178,000 particles (see Methods), and using the eukaryotic small ribosomal subunit as a reference. **(D),** Gold-standard FSC curve of the of salt-washed 40S sample (6.5Å). **(E)**, Overview of the particle sorting of the native salt-washed 40S sample. The final reconstruction was performed from ~44,000 particles (see Methods). **(F),** Local resolution of the cryo-EM map from *T. cruzi* 80S, viewed from beak and SSU sides. Close-up view highlights the local resolution of the additional density on the 40S left foot (4Å for most of this density and up to ~7Å). **(G),** Gold-standard FSC curve of the of native 80S sample (4Å). **(H)**, Overview of the particle sorting for the native 80S sample. The final reconstruction was performed from ~86,000 particles (see Methods). **(I-K),** Overview of the particle sorting of the *T. cruzi* 80S-antibody complex and resolution estimation. 61% of the particles present a residual density attributable to a segment of the antibody. These particles were further sorted after the application of an oval mask at the antibody expected binding region, yielding two main classes. The first one (Ab_C1) includes ~65,000 particles and the second one (Ab_C2) includes ~11,000 particles (see Methods). In all panels, the populations of all classes are indicated by their particle percentages over the total number of particles. The cyan square indicates the retained major classes for further refinement in RELION.

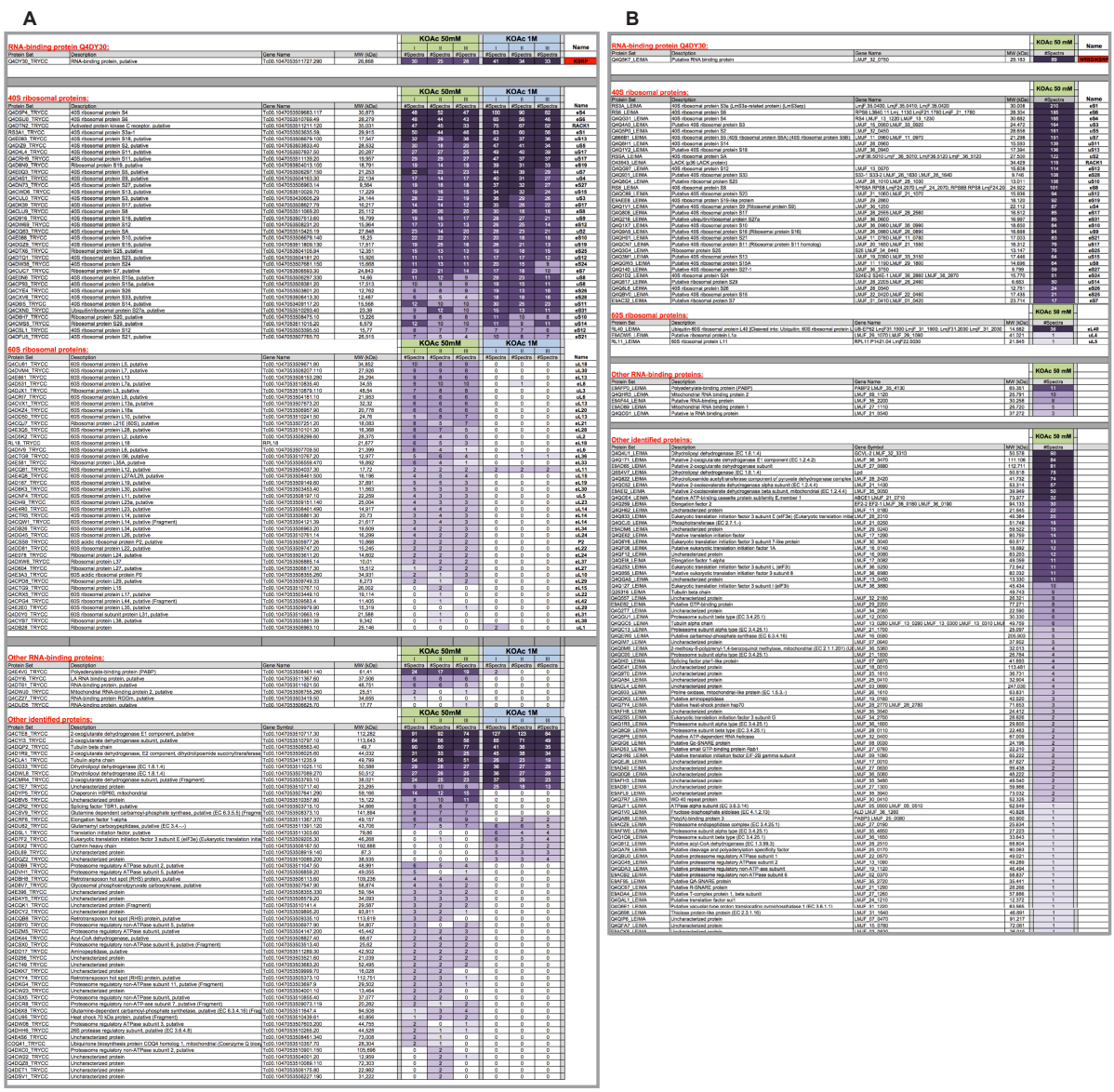
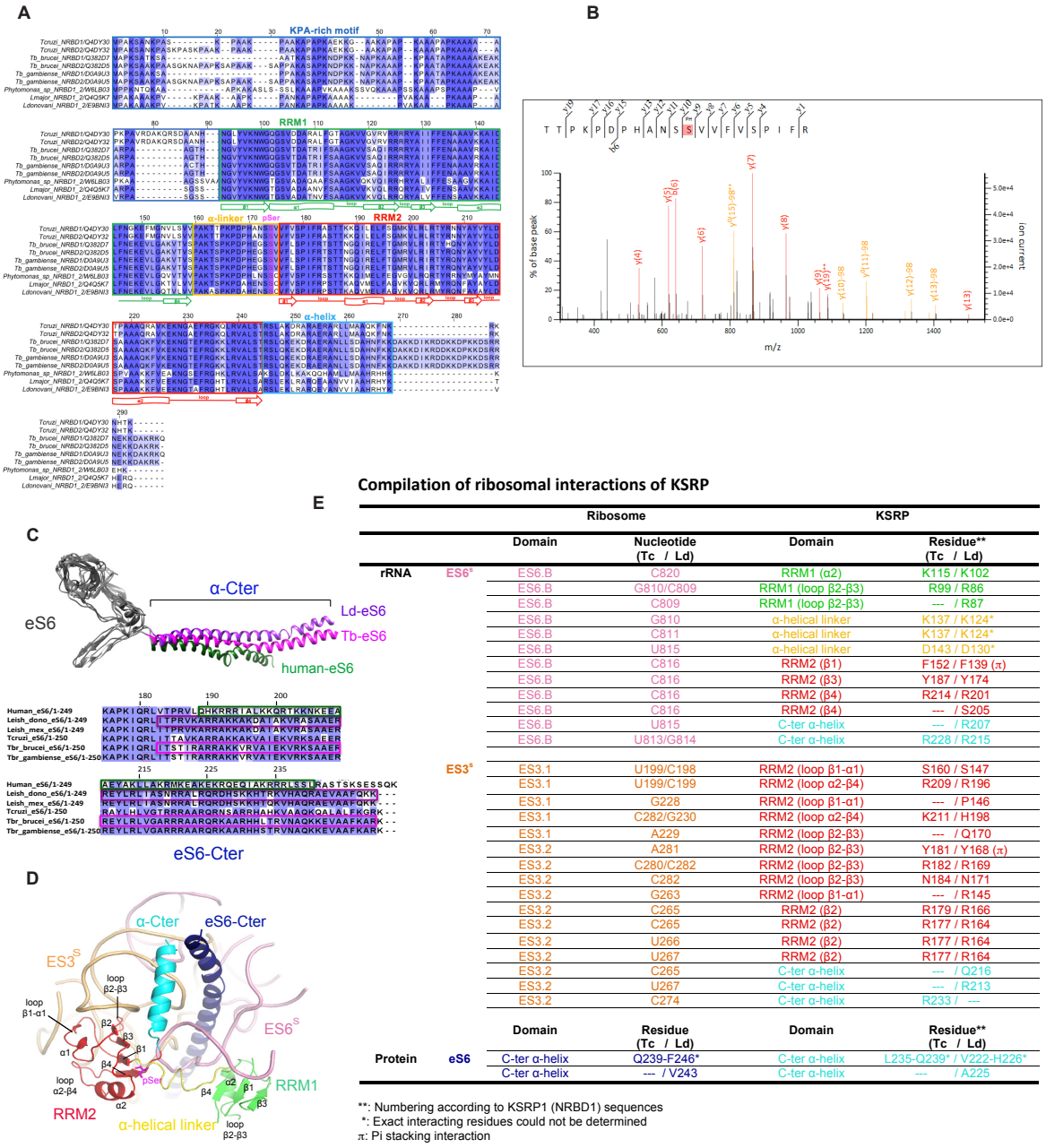


Fig. S3. Mass spectrometry characterization of native (50 mM KOAc) and salt-washed (1M KOAc) 40S complexes from *T. cruzi* and comparison with *L. tarentolae* (50 mM KOAc). Related to Figure 1. (A), The 40S complexes obtained after sucrose gradient and gel filtration were analyzed by MS/MS. The figure contains all the proteins identified from each complex, and TcNRBD/KSRP was highlighted on the top of the table. The spectral counts were used for relative quantification. Darker colors indicate the most abundant protein present in the sample. The analysis was performed in triplicate of each sample. **(B)**, Proteomic analysis of *L. tarentolae* 40S. The data were searched against the *T. cruzi* and *L. major* proteome sets from the UniProt database.



E

Compilation of ribosomal interactions of KSRP

		Ribosome		KSRP	
		Domain	Nucleotide (Tc / Ld)	Domain	Residue** (Tc / Ld)
rRNA	ES6S	ES6.B	C820	RRM1 (α2)	K115 / K102
		ES6.B	G810/C809	RRM1 (loop β2-β3)	R99 / R86
		ES6.B	C809	RRM1 (loop β2-β3)	--- / R87
		ES6.B	G810	α-helical linker	K137 / K124*
		ES6.B	C811	α-helical linker	K137 / K124*
		ES6.B	U815	α-helical linker	D143 / D130*
		ES6.B	C816	RRM2 (β1)	F152 / F139 (π)
		ES6.B	C816	RRM2 (β3)	Y187 / Y174
		ES6.B	C816	RRM2 (β4)	R214 / R201
		ES6.B	C816	RRM2 (β4)	--- / S205
		ES6.B	U815	C-ter α-helix	--- / R207
		ES6.B	U813/G814	C-ter α-helix	R228 / R215
	ES3S	ES3.1	U199/C198	RRM2 (loop β1-α1)	S160 / S147
		ES3.1	U199/C199	RRM2 (loop α2-β4)	R209 / R196
		ES3.1	G228	RRM2 (loop β1-α1)	--- / P146
		ES3.1	C282/G230	RRM2 (loop α2-β4)	K211 / H198
		ES3.1	A229	RRM2 (loop β2-β3)	--- / Q170
		ES3.2	A281	RRM2 (loop β2-β3)	Y181 / Y168 (π)
		ES3.2	C280/C282	RRM2 (loop β2-β3)	R182 / R169
		ES3.2	C282	RRM2 (loop β2-β3)	N184 / N171
		ES3.2	G263	RRM2 (loop β1-α1)	--- / R145
		ES3.2	C265	RRM2 (β2)	R179 / R166
		ES3.2	C265	RRM2 (β2)	R177 / R164
		ES3.2	U266	RRM2 (β2)	R177 / R164
		ES3.2	U267	RRM2 (β2)	R177 / R164
		ES3.2	C265	C-ter α-helix	--- / Q216
		ES3.2	U267	C-ter α-helix	--- / R213
		ES3.2	C274	C-ter α-helix	R233 / ---
		Domain	**Residue (Tc / Ld)**	**Domain**	**Residue** (Tc / Ld)**
Protein	eS6	C-ter α-helix	Q239-F246*	C-ter α-helix	L235-Q239* / V222-H226*
		C-ter α-helix	--- / V243	C-ter α-helix	--- / A225

**: Numbering according to KSRP1 (NRBD1) sequences
*: Exact interacting residues could not be determined
π: Pi stacking interaction

Fig. S4. Multiple sequence alignment, ribosomal interactions and MS/MS spectra of KSRP. Related to Figure 5 and Figure 6. (A), Multiple sequence alignment colored by BLOSUM 62 score of *T. cruzi* NRBD1 and NRBD2 orthologues highlighting conserved RRM1 and RRM2 regions when compared to its *T. brucei brucei*, *T. brucei gambiense*, *L. major*, *L. donovani*, and *Phonas sp. isolate hart1* counterparts. The most highly conserved regions are represented by darkest color. **(B)**, MS/MS spectra of the phosphorylated peptide (TTPKPDPHANSSpVVFVSPIFR), which is phosphorylated at the highly conserved serine S14 (y9 fragmented peptide on the spectrum). This phosphorylation site was previously predicted by the NetPhos 3.1 software (score = 0.61 for PKC kinase and 0.542 for Cdc2 kinase) (Blom et al., 1999). **(C)**, Multiple sequence alignment of kinetoplastidian and human variants of r-protein eS6 C-terminal end (colored by BLOSUM 62 score). This alignment and the 3D structures of eS6 proteins highlight the fact that in spite of having comparable lengths, the kinetoplastidian variants of the r-protein eS6 have evolved more structured and differently arranged C-terminal tails when compared to human eS6. The residue numbering is presented according to kinetoplastid eS6 sequences. **(D),** Overview of KSRP in the context of the ribosome as seen from the solvent side. The highly conserved phosphorylated serine 14 residue is represented as magenta sticks (pSer). **(E)**, Detailed interactions of KSRP and the ribosome. The interacting domains/residues are colored based on the color code used in the 3D atomic model shown in **D**.

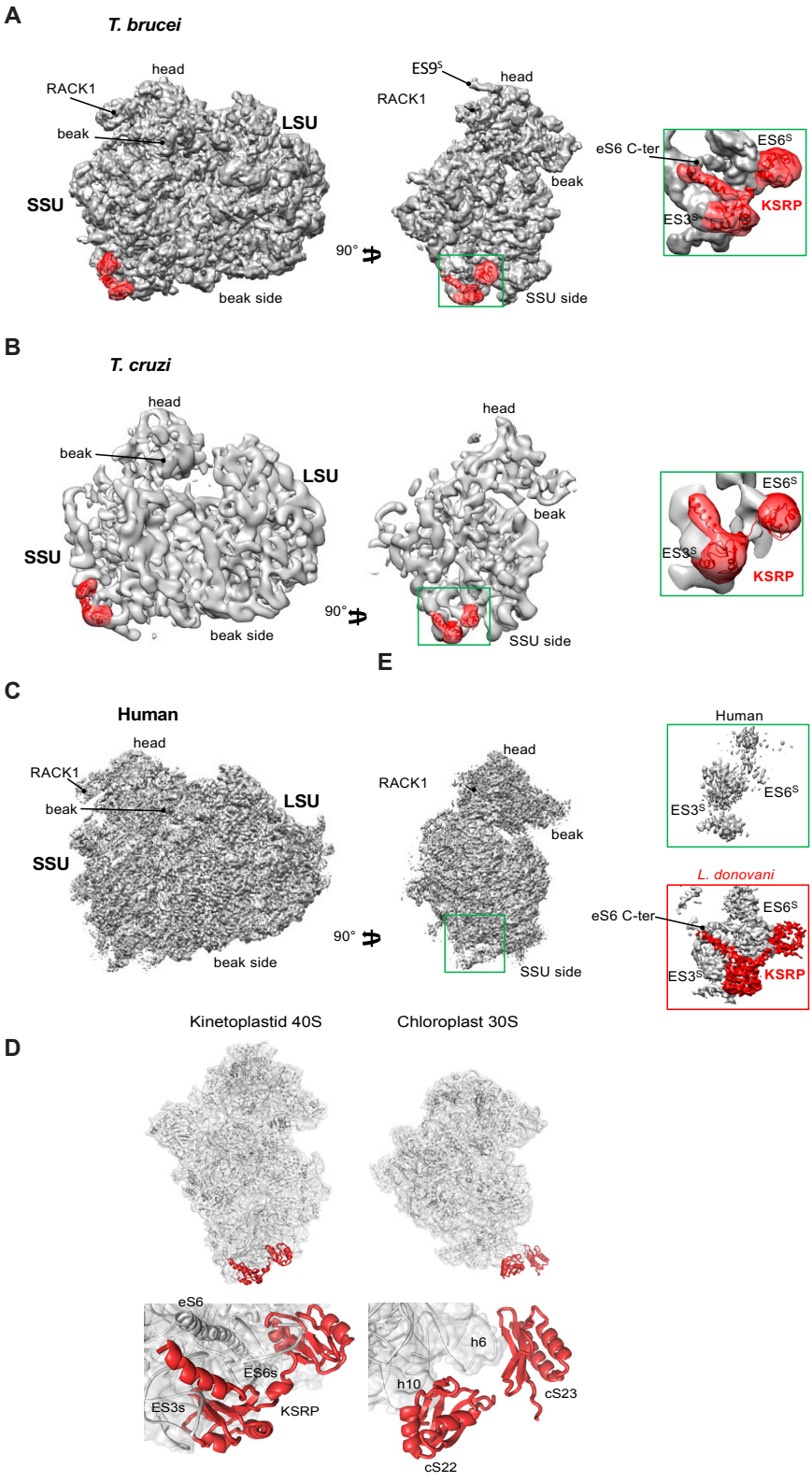


Fig. S5. Conservation of KSRP in the ribosomes of *T. brucei* and *T. cruzi*, and comparison of its kinetoplastid-specific ribosomal site with the human and spinach chloroplast ribosomes, Related to Figure 2 and 5. (A and **B),** Atomic model of LdKSRP fitted in *T. brucei* and *T. cruzi* 80S cryo-EM reconstruction map (Hashem et al., 2013; Kao et al., 2005). Here we highlighted the overlooked KSRP related density in red. **(C),** Human 80S cryo-EM reconstruction map, highlighting the absence of KSRP related density. Close-up view on the 40S left foot highlights the structural differences when compared to *L. donovani* ribosome (Zhang et al., 2016). **(D),** Cartoon model of the 40S from *T. cruzi* and the 30S from spinach chloroplast (PDB ID 5MMJ). KSRP and chloroplast-specific proteins cS22/cS23 are highlighted in red.

Supplementary Table 1. Refinement and validation statistics for the native 40S structure from *T. cruzi* built in the 80S map of *T. cruzi*.

Refinement	
Number of particles (total)	~183 500
Number of particles (used for 3D reconstruction)	~86 250
Resolution (Å; at FSC[a] = 0.143)	4.0
CC[a]	0.84
Model composition	
Non-hydrogen atoms	82265
Residues	6926
RMS[a] deviations	
Bonds (Å)	0.017
Angles (°)	2.057
Chirality (°)	0.130
Planarity (°)	0.020
Validation[b]	
Clashscore[c]	1.35 (97th percentile)
Proteins	
MolProbity score	1.42 (97th percentile)
Favored rotamers	3,830 (97.55%)
Ramachandran favored	4330 (89.59%)
Ramachandran allowed	412 (8.53%)
Ramachandran outliers	91 (1.88%)
RNA	
Correct sugar puckers	1983 (98.12%)
Correct backbone conformations	1582 (78.25%)

[a] FSC, Fourier shell correlation; CC, correlation coefficient (compiled by Chimera using a map simulated from the 40S model); RMS, root-mean square

[b] Compiled using MolProbity (80)

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

Article 2: Structural differences in translation initiation between pathogenic trypanosomatids and their mammalian hosts

In Brief

This study presents the structure of the 43S preinitiation complex of *Trypanosoma cruzi*, which provides molecular insight to translation initiation and reveals features common to other eukaryotes whilst also uncovering distinct features of Kinetoplastids.
Some of these specific features are the docking of eIF3 on expansion segment ES7S, the structure of eIF3 from *Trypanosoma cruzi* that is lacking the subunit m compared to their mammalian counterpart and the Kinetoplastid-specific helicase k-DDX60. This helicase is intriguing because one of its domains is completely inserted in the mRNA entry channel and k-DDX60 seems to induce stability to the whole complex. This stability permitted to see interactions between eIFs and 40S that has only been characterized biochemically.

Unique features of mRNA translation initiation in trypanosomatids

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

Affiliations:

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

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

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

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

[#]These authors contributed equally to this work.

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

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

Abstract:

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

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

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

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

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

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

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

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

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

Results and discussion:

Composition of the 43S PIC in trypanosomatids

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The specific features and binding site of eIF3 in trypanosomatids

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

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

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

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

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

The trypanosomatid-specific k-DDX60

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

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

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

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

Conclusion

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

References and Notes:

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

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

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

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

21. M. Miyashita, H. Oshiumi, M. Matsumoto, T. Seya, DDX60, a DEXD/H box helicase, is a novel antiviral factor promoting RIG-I-like receptor-mediated signaling. *Mol. Cell. Biol.* **31**, 3802–19 (2011).
22. H. Oshiumi, M. Miyashita, M. Okamoto, Y. Morioka, M. Okabe, M. Matsumoto, T. Seya, DDX60 Is Involved in RIG-I-Dependent and Independent Antiviral Responses, and Its Function Is Attenuated by Virus-Induced EGFR Activation. *Cell Rep.* **11**, 1193–207 (2015).
23. S. Alsford, D. J. Turner, S. O. Obado, A. Sanchez-Flores, L. Glover, M. Berriman, C. Hertz-Fowler, D. Horn, High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res.* **21**, 915–24 (2011).
24. J. L. Llácer, T. Hussain, L. Marler, C. E. Aitken, A. Thakur, J. R. Lorsch, A. G. Hinnebusch, V. Ramakrishnan, Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Complex. *Mol. Cell*. **59**, 399–412 (2015).
25. A. S. Y. Lee, P. J. Kranzusch, J. H. D. Cate, eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature* (2015), doi:10.1038/nature14267.
26. A. S. Lee, P. J. Kranzusch, J. A. Doudna, J. H. D. Cate, eIF3d is an mRNA cap-binding protein that is required for specialized translation initiation. *Nature*. **536**, 96–9 (2016).
27. Z. Wei, Y. Xue, H. Xu, W. Gong, Crystal structure of the C-terminal domain of S.cerevisiae eIF5. *J. Mol. Biol.* **359**, 1–9 (2006).
28. R. E. Luna, H. Arthanari, H. Hiraishi, J. Nanda, P. Martin-Marcos, M. A. Markus, B. Akabayov, A. G. Milbradt, L. E. Luna, H.-C. Seo, S. G. Hyberts, A. Fahmy, M. Reibarkh, D. Miles, P. R. Hagner, E. M. O'Day, T. Yi, A. Marintchev, A. G. Hinnebusch, J. R. Lorsch, K. Asano, G. Wagner, The C-terminal domain of eukaryotic initiation factor 5 promotes start codon recognition by its dynamic interplay with eIF1 and eIF2β. *Cell Rep.* **1**, 689–702 (2012).
29. C. R. Singh, Y. Yamamoto, K. Asano, Physical association of eukaryotic initiation factor (eIF) 5 carboxyl-terminal domain with the lysine-rich eIF2beta segment strongly enhances its binding to eIF3. *J. Biol. Chem.* **279**, 49644–55 (2004).
30. C. R. Singh, R. Watanabe, W. Chowdhury, H. Hiraishi, M. J. Murai, Y. Yamamoto, D. Miles, Y. Ikeda, M. Asano, K. Asano, Sequential eukaryotic translation initiation factor 5 (eIF5) binding to the charged disordered segments of eIF4G and eIF2β stabilizes the 48S preinitiation complex and promotes its shift to the initiation mode. *Mol. Cell. Biol.* **32**, 3978–89 (2012).
31. P. V Alone, T. E. Dever, Direct binding of translation initiation factor eIF2gamma-G domain to its GTPase-activating and GDP-GTP exchange factors eIF5 and eIF2B epsilon. *J. Biol. Chem.* **281**, 12636–44 (2006).
32. L. R. Kenner, A. A. Anand, H. C. Nguyen, A. G. Myasnikov, C. J. Klose, L. A. McGeever, J. C. Tsai, L. E. Miller-Vedam, P. Walter, A. Frost, eIF2B-catalyzed nucleotide

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

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

(2016).

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

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

ACKNOWLEDGMENTS

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

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

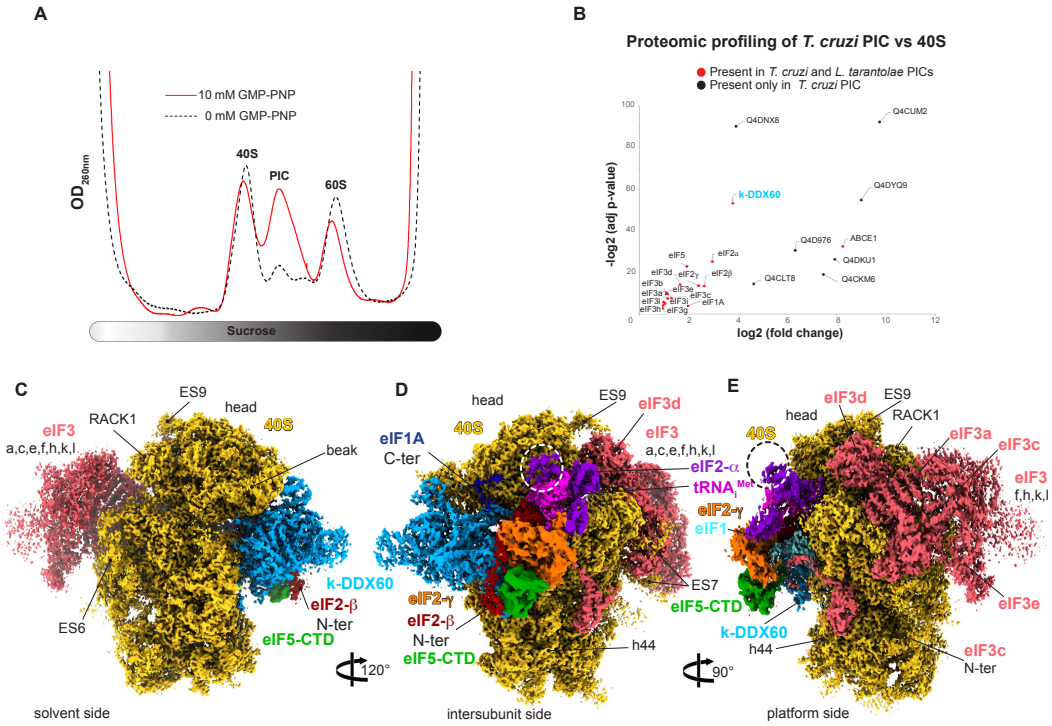


Fig. 1. Composition and cryo-EM structure of the T. cruzi 43S PIC. (**A**) The effect of the GMP-PNP treatment on the 43S PIC stabilization in the *T.cruzi* lysate assessed by UV absorbance profile analyses (**B**) Proteomic profiling of the endogenous pre-initiation complex in comparison with native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

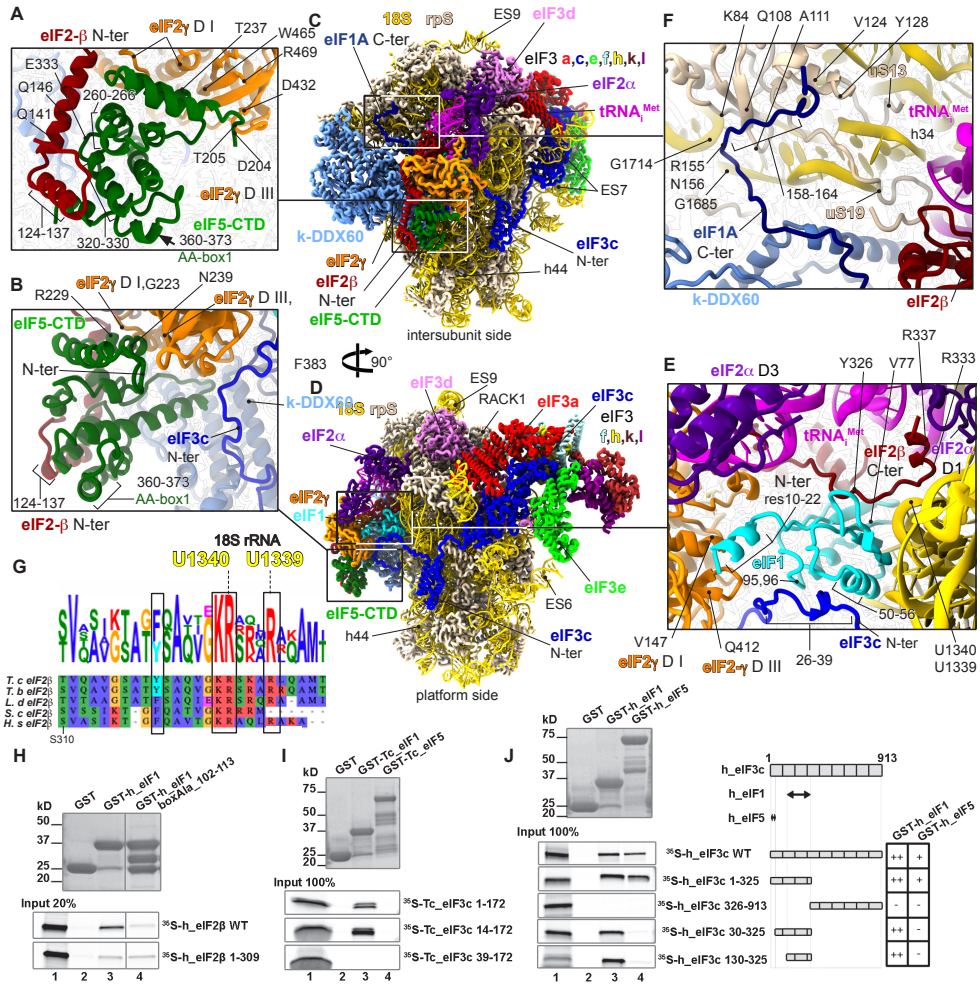


Fig. 2. Atomic model of the 43S PIC showing the interaction network of various eIFs. (**A**) Close-up view of an atomic model of the eIF5-CTD (in green), the eIF2β-NTT (in cherry red) and eIF2γ (in orange) shown from the intersubunit side. (**B**) Close-up view of the eIF5-CTD (in green) and its interaction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

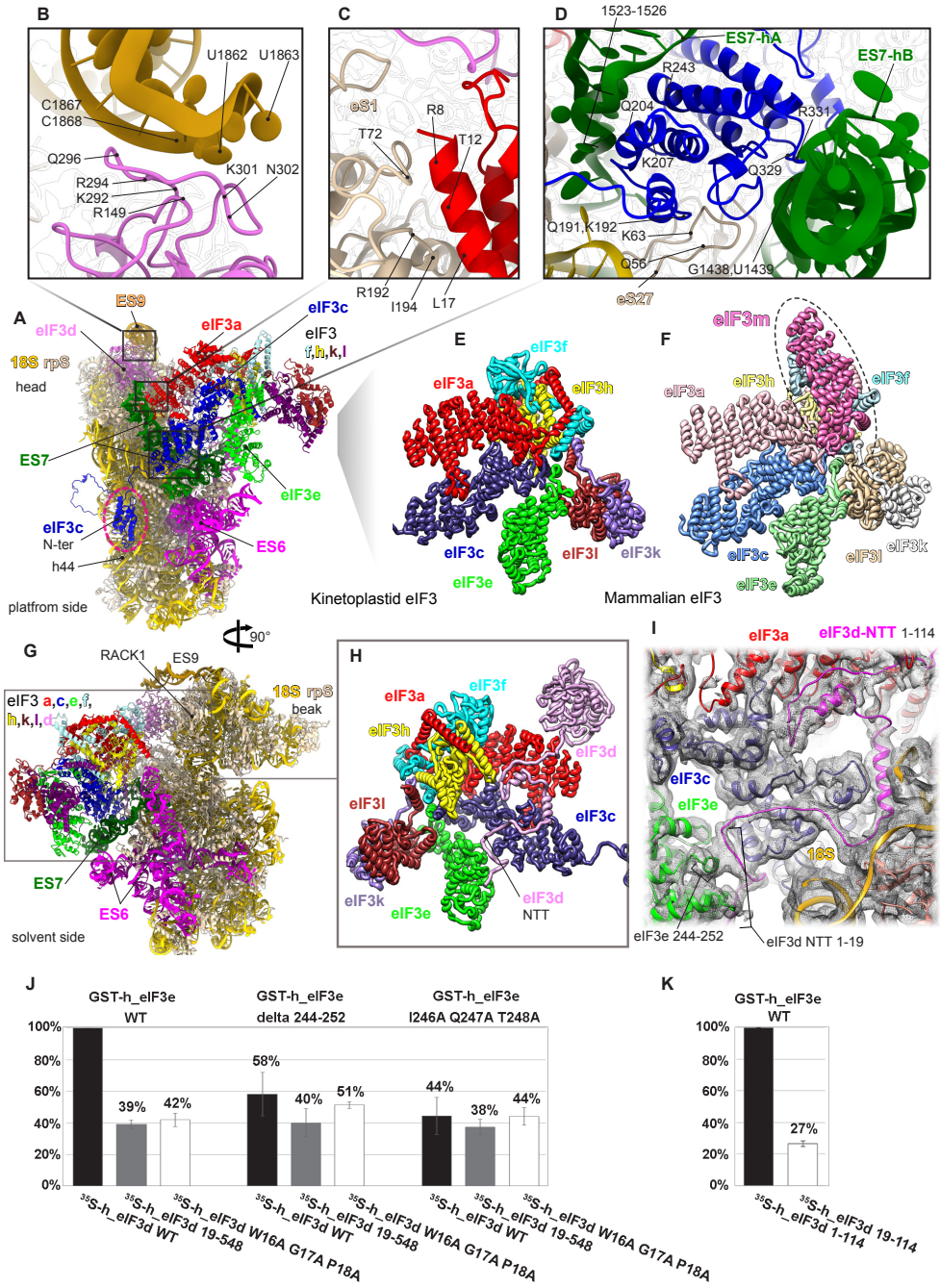


Fig. 3. Kinetoplastidian eIF3 and its unique binding site. (A) The overall view of the atomic model of the 43S PIC from the platform side. The conserved helical domain of the eIF3c-NTD is encircled with dashed line, eIF3 subunits are colored variably and 18S RNA in yellow. (**B**) Close-up view of the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

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

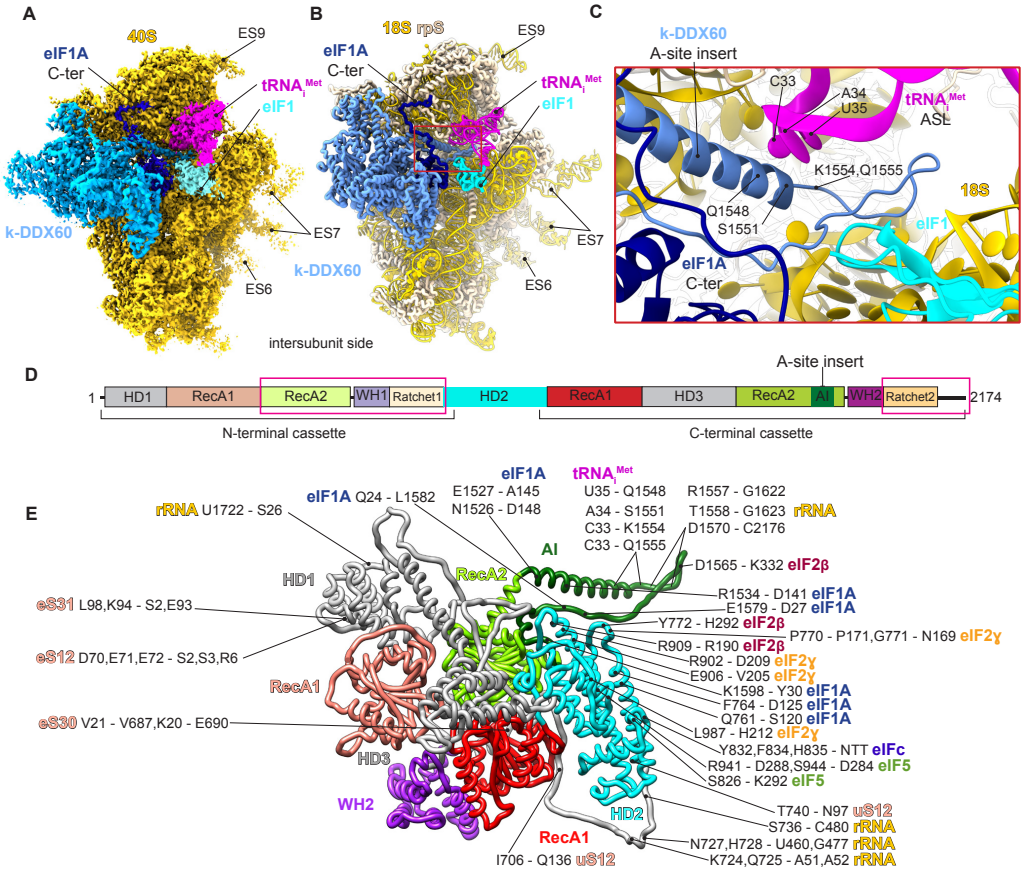


Fig. 4. k-DDX60 structure and interactions within the 43S PIC. (**A**) The cryo-EM structure of the *T. cruzi* 43S PIC highlighting k-DDX60 (colored in dark turquoise). eIF 2, 3 and 5 densities were removed for clarity (**B**) Cartoon representation of a partial atomic model of the *T. cruzi* 43S PIC.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

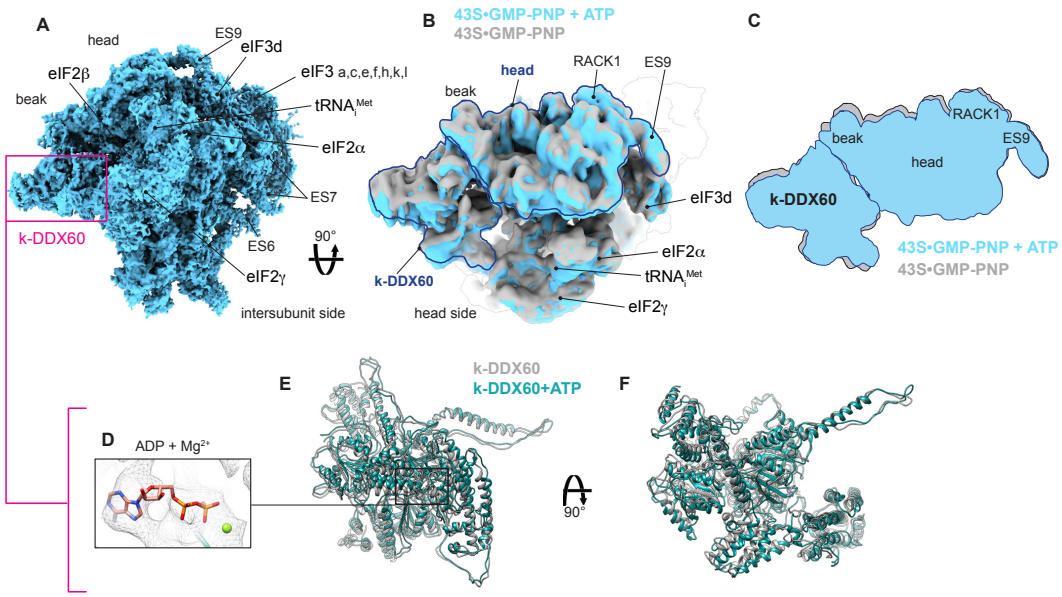


Fig. 5. Global conformational rearrangement of the 43S PIC driven by ATP binding to k-DDX60. (**A**) Cryo-EM reconstruction of the *T. cruzi* 43S PIC in the presence of ATP. (**B**) Superposition of the cryo-EM reconstructions of the 43S•GMP-PNP (in grey) and the 43S•GMP-PNP supplemented with ATP (in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

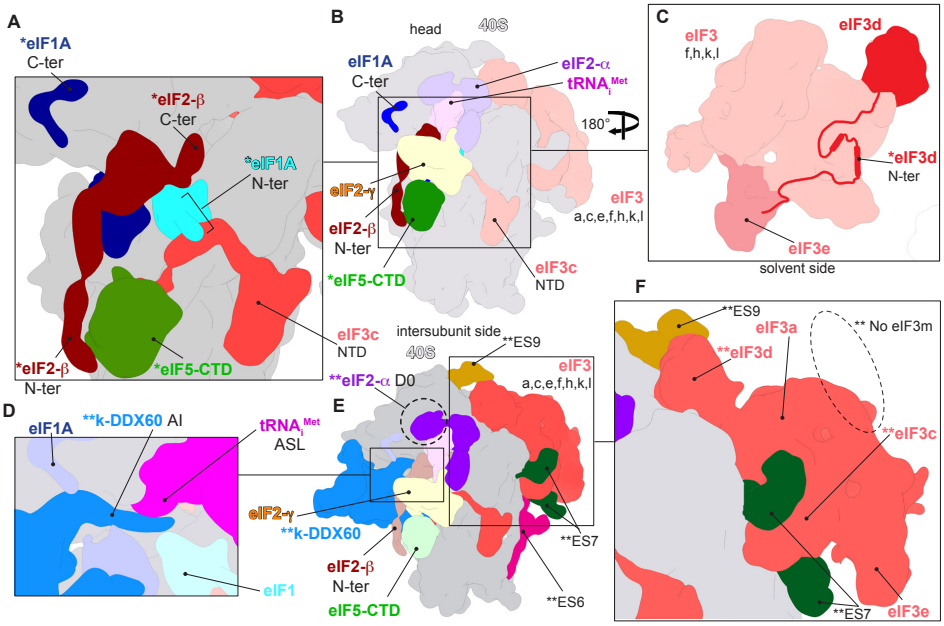


Fig. 6. Novel eukaryotic-conserved and trypanosomatid-specific features of the 43S PIC revealed in our work. (**A**) Schematic model representing a close up view on the N-terminal tails of eIF 1, 1A, 2β, eIF5-CTD and eIF3c-NTD, all conserved among eukaryotes and revealed in the current work. The ternary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

Supplementary Materials

Unique features of mRNA translation initiation in trypanosomatids

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

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

Materials and Methods

Kinetoplastids Cultures

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

48S Initiation Complex Purification

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

Cryo-EM Grid preparation

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

Cryo-EM Image acquisition

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

Image processing

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

Mass spectrometry analysis and data post-processing

Protein extracts were precipitated overnight with 5 volumes of cold 0.1 M ammonium acetate in 100% methanol. Proteins were then digested with sequencing-grade trypsin (Promega, Fitchburg, MA, USA) as described previously (Brito Querido et al., 2017). Each sample was further analyzed by nanoLC-MS/MS on a QExactive+ mass spectrometer coupled to an EASY-nanoLC-1000 (Thermo-Fisher Scientific, USA). Peptides and proteins were identified with Mascot algorithm (version 2.5.1, Matrix Science, London, UK) and data were further imported into Proline v1.4 software (http://proline.profiproteomics.fr/). Proteins were validated on Mascot pretty rank equal to 1, and 1% FDR on both peptide spectrum matches (PSM score) and protein sets (Protein Set score). The total number of MS/MS fragmentation spectra was used to relatively quantify each protein (Spectral Count relative quantification). Proline was further used to align the Spectral Count values across all samples. The whole MS dataset was then normalized. The mass spectrometric data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXDxxxx.

Volcano plot

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

Model building and refinement

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

GST pulldown assay

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

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

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

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

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

40S ribosomal proteins:

			BASIC Spectral Count (# spectra)		
				BEFORE Gel Filtration	AFTER Gel Filtration
accession		description	40S	43S	43S
Q4D5P4|Q4D5P4_TRYCC		40S ribosomal protein S4 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053509683.117	131	131	93
Q4DTN2|Q4DTN2_TRYCC		Activated protein kinase C receptor, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.10	100	96	48
Q4E0Q3|Q4E0Q3_TRYCC		40S ribosomal protein S5, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506	65	51	40
Q4DZ41|RS3A2_TRYCC		40S ribosomal protein S3a-2 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511001.9	98	84	46
Q4E093|Q4E093_TRYCC		40S ribosomal protein S18, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	75	73	64
Q4DSU0|Q4DSU0_TRYCC		40S ribosomal protein S6 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053510769.49 P	89	72	58
Q4CLU9|Q4CLU9_TRYCC		40S ribosomal protein S8 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511069.20 P	66	60	46
Q4D4L4|Q4D4L4_TRYCC		40S ribosomal protein S11, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	58	47	39
Q4D6I5|Q4D6I5_TRYCC		40S ribosomal protein S14, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705345	60	61	37
Q4CUL0|Q4CUL0_TRYCC		40S ribosomal protein S3, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705343	81	72	37
Q4D6N9|Q4D6N9_TRYCC		Ribosomal protein S19, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350401	39	36	36
Q4DY30|Q4DY30_TRYCC	KSRP	RNA-binding protein, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511727.2	79	72	29
Q4D4S1|Q4D4S1_TRYCC		40S ribosomal protein S9, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053504	38	38	28
Q4CUC8|Q4CUC8_TRYCC		Ribosomal protein S7, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506593	84	80	25
Q4CQU0|Q4CQU0_TRYCC		40S ribosomal protein SA OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053503719.20 P	65	58	22
Q4D916|Q4D916_TRYCC		40S ribosomal protein S16, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	48	52	19
Q4E0N6|Q4E0N6_TRYCC		40S ribosomal protein S15a, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705035	37	34	15
Q4DIZ9|Q4DIZ9_TRYCC		40S ribosomal protein S2, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705360	80	74	27
Q4CXN0|Q4CXN0_TRYCC		Ubiquitin/ribosomal protein S27a, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047	52	40	14
Q4DTX6|Q4DTX6_TRYCC		Ribosomal protein S25, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350410	46	44	8
Q4DK39|Q4DK39_TRYCC		40S ribosomal protein S17, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	58	57	16
Q4E088|Q4E088_TRYCC		40S ribosomal protein S10, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	52	54	22
Q4DW69|Q4DW69_TRYCC		40S ribosomal protein S12 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508231.20	34	39	13
Q4CXV6|Q4CXV6_TRYCC		40S ribosomal protein S33, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	34	31	17
Q4DTQ1|Q4DTQ1_TRYCC		40S ribosomal protein S23, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	33	28	28
Q4D6H7|Q4D6H7_TRYCC		Ribosomal protein S20, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350847	34	28	16
Q4CWD6|Q4CWD6_TRYCC		40S ribosomal protein S13, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705351	32	30	18
Q4DN73|Q4DN73_TRYCC		40S ribosomal protein S27, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	21	17	25
Q4DW38|Q4DW38_TRYCC		40S ribosomal protein S24 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053507681.150	30	26	15
Q4CMS5|Q4CMS5_TRYCC		Ribosomal protein S29, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705351101	17	15	13
Q4DGZ5|Q4DGZ5_TRYCC		40S ribosomal protein S15, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705351	23	20	11
Q4CYE4|Q4CYE4_TRYCC		Ribosomal protein S26, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350380	21	18	12
Q4E3L9|Q4E3L9_TRYCC		40S ribosomal protein S21, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705351	24	18	7
Q4DA48|Q4DA48_TRYCC		40S ribosomal protein S30, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	2	5	

Initiation factors:

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

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

	Name		Spectral Count IC
tr|E9ACL4|E	DDX60	Uncharacterized protein OS=Leishmania major GN=LMJF_03_0690 PE=4 SV=1	111
tr|Q4QCE4|(	ABCE1	Putative ATP-binding cassette protein subfamily E,member 1 OS=Leishmania major GN=ABCI	101
40S ribosomal proteins:			
	Name		Spectral Count
accession		description	IC
tr|Q868B1|C		40S ribosomal protein S5 OS=Leishmania major GN=LMJF_11_0960 PE=4 SV=1	188
tr|Q4Q216|C		Putative ubiquitin/ribosomal protein S27a OS=Leishmania major GN=LMJF_36_0600 PE=4 SV	265
tr|Q4Q1Y2|(		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|F		40S ribosomal protein S8 OS=Leishmania major GN=RPS8A PE=3 SV=1	108
sp|Q9NE83|		40S ribosomal protein S6 OS=Leishmania major GN=RPS6 PE=3 SV=1	175
tr|Q4Q817|C		Putative ribosomal protein S29 OS=Leishmania major GN=LMJF_28_2205 PE=4 SV=1	62
tr|Q4Q1V1|(		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|Q4Q0Q0|		40S ribosomal protein SA OS=Leishmania major GN=LmjF36.5010 PE=3 SV=1	145
tr|E9AEE8|E		40S ribosomal protein S19-like protein OS=Leishmania major GN=LMJF_29_2860 PE=4 SV=1	129
tr|Q4Q931|(		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|Q4QG97|(		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|O43943|O	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|Q4QBV0|(		Putative 40S ribosomal protein S15 OS=Leishmania major GN=LMJF_22_0420 PE=3 SV=1	31
tr|E9AC32|E		Putative ribosomal protein S7 OS=Leishmania major GN=LMJF_01_0410 PE=4 SV=1	27
Initiation factors:			
	Name		C Spectral Count (# spe
accession		description	IC
tr|Q4QEJ8|(	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	125
tr|Q4Q833|(	eIF3e	Eukaryotic translation initiation factor 3 subunit E OS=Leishmania major GN=LMJF_28_2310 P	84
tr|Q4Q253|(	eIF3l	Eukaryotic translation initiation factor 3 subunit L OS=Leishmania major GN=LMJF_36_0250 P	91
tr|Q4Q127|(	eIF3i	Eukaryotic translation initiation factor 3 subunit I OS=Leishmania major GN=LMJF_36_3880 P	79
tr|E9ACP3|E	eIF2 alpha	Putative elongation initiation factor 2 alpha subunit OS=Leishmania major GN=LMJF_03_0980	76
tr|Q4QIM7|(	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	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	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_	49
tr|Q4Q2S5|(	eIF3g	Eukaryotic translation initiation factor 3 subunit G OS=Leishmania major GN=LMJF_34_2700 P	46
tr|Q4QAL1|(	eIF1A	Putative translation factor sui1 OS=Leishmania major GN=LMJF_24_1210 PE=4 SV=1	33
tr|Q4QIB4|C	eIF2 beta	Translation initiation factor-like protein OS=Leishmania major GN=LMJF_08_0550 PE=4 SV=1	29
tr|Q4QF06|(	eIF1A	Putative eukaryotic translation initiation factor 1A OS=Leishmania major GN=LMJF_16_0140 P	27

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

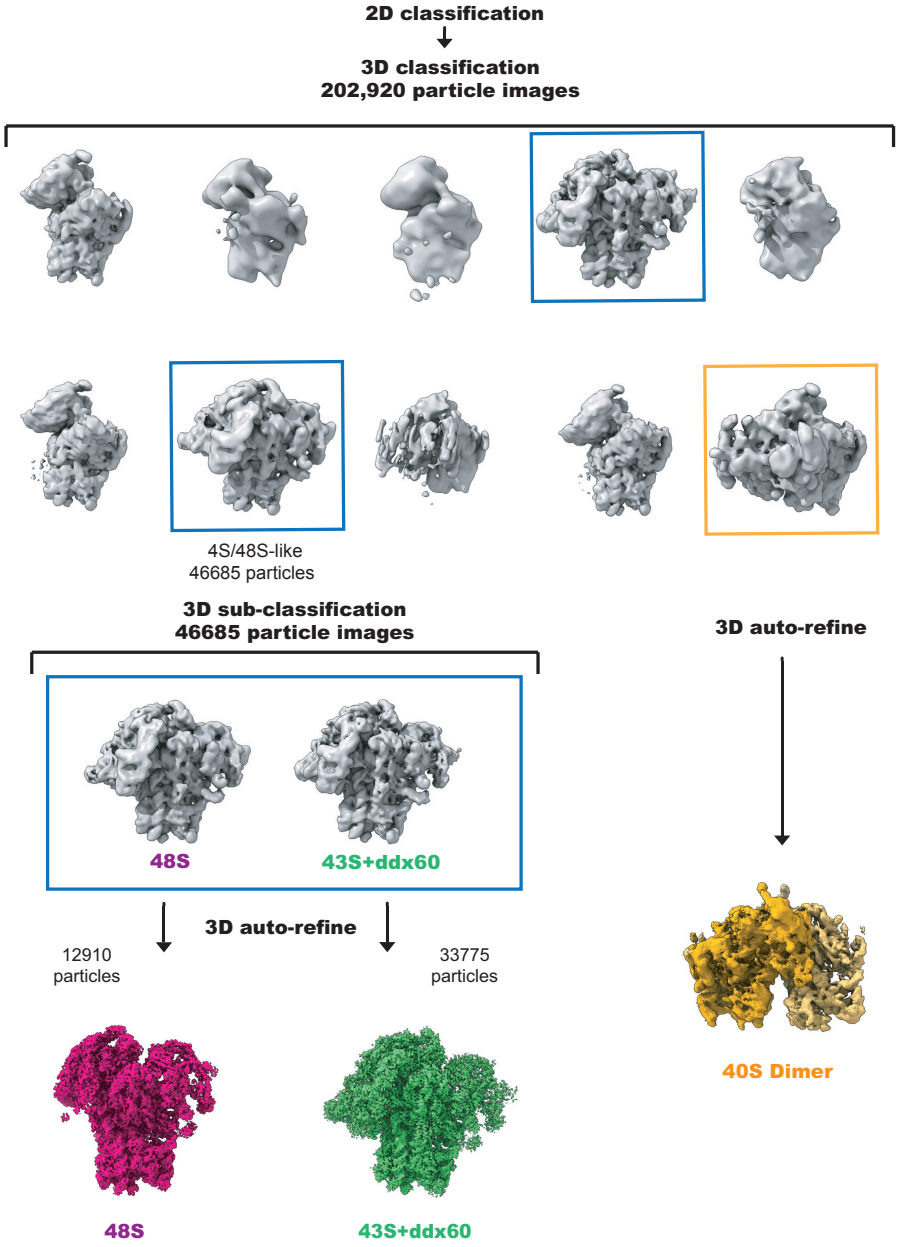


Fig. S3. Cryo-EM particle sorting and refinement of the *T. cruzi* 43S PIC. 2D classification of the 43S PIC particles yielded ~200 000 40S-like particles, after which a run of 3D classification (10 classes) allowed to single out 43S/48S ICs and 40S-dimers. A secondary run of 3D classification allowed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

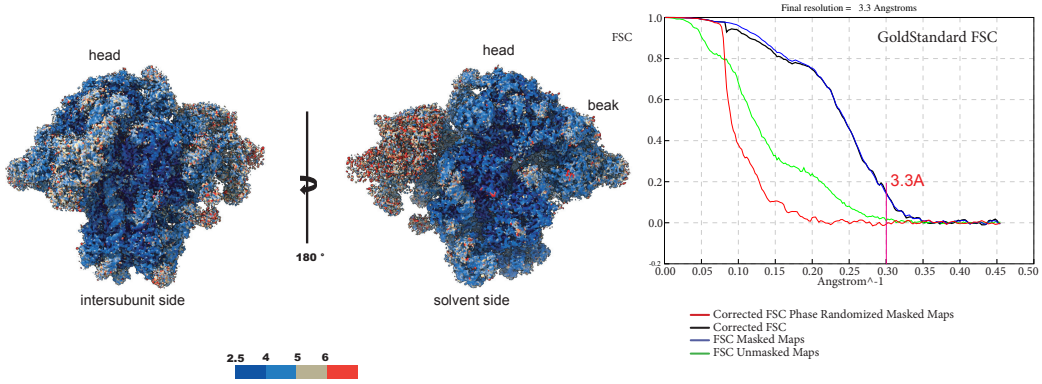


Fig. S4. Cryo-EM average and local resolution of the *T. cruzi* 43S PIC. The local resolution varies mainly on eIF3 (ranging from ~3 to ~6 Å), while is varies less on the rest of the structure (ranging from ~2.5 to ~3.5 Å for the 40S, k-DDX60, eIFs 1, 1A and 2b, and from ~3 to ~5 Å for eIFs 2a, 2g and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

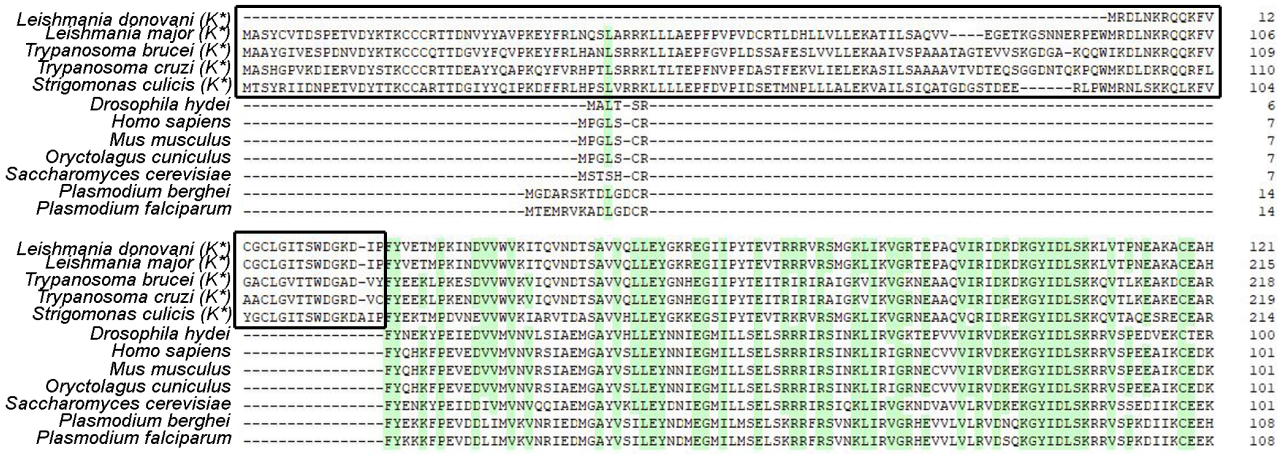


Fig. S5. Multiple sequence alignment of the eIF2a NTD among eukaryotes. Protein sequence alignment of eIF2α from various eukaryotic organisms was generated by Clone Manger (MultiWay, scoring matrix: Blosum 62). The Kinetoplastida order species are labeled with K*. The kinetoplastidian-specific eIF2α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

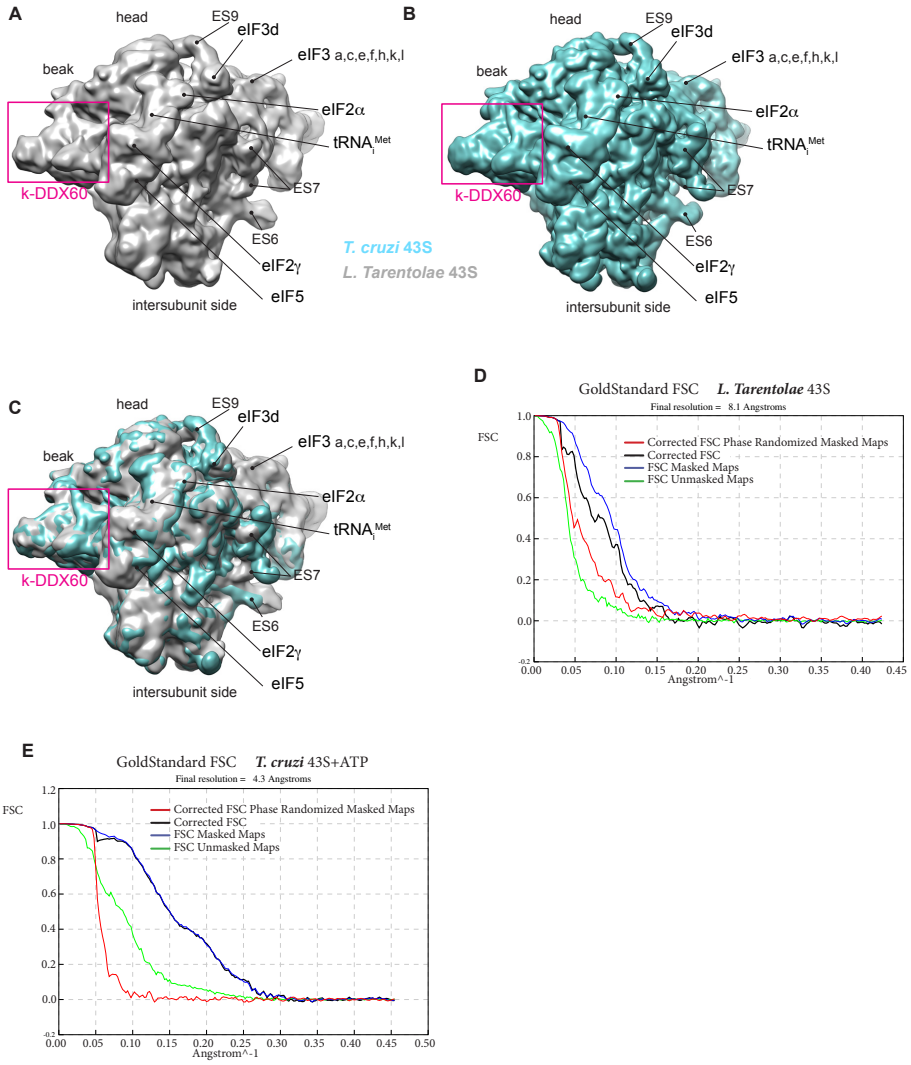


Fig. S6. Cryo-EM reconstruction of *L. tarentolae* 43S PIC compared to *T. cruzi* and average resolution. (**A**) Cryo-EM reconstructions of the *L. tarentolae* 43S PIC. (**B**) Cryo-EM reconstructions of the *T. cruzi* 43S PIC filtered at 8Å. (**C**) Superimposition of (A) and (B). (**D**) Average resolution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

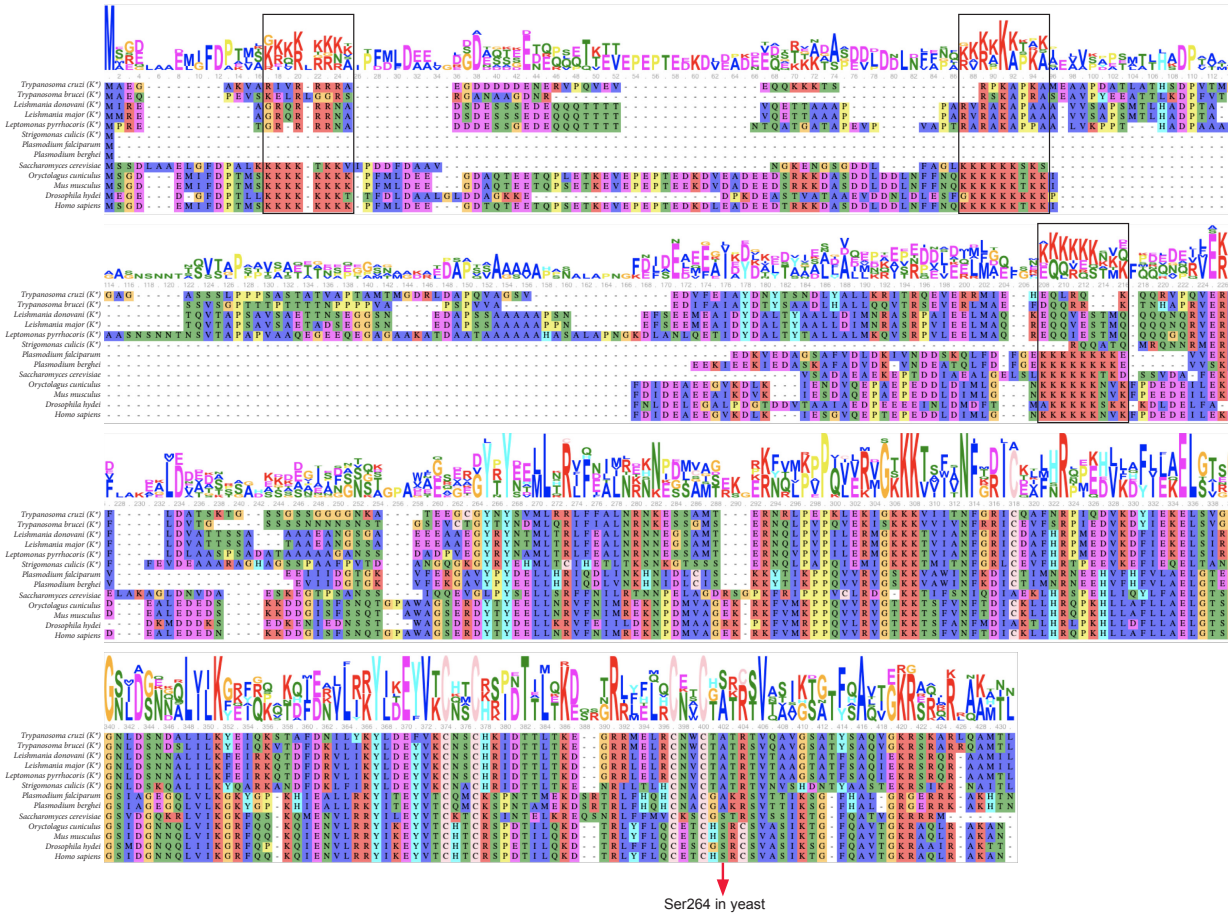


Fig. S7. Multiple sequence alignment of eIF2β among eukaryotes. Protein sequence alignment of eIF2β protein from various eukaryotic organisms. The Kinetoplastida order species are labeled with K*. Consensus is expressed as a sequence logo. The black boxes mark three conserved poly-lysine stretches (dubbed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

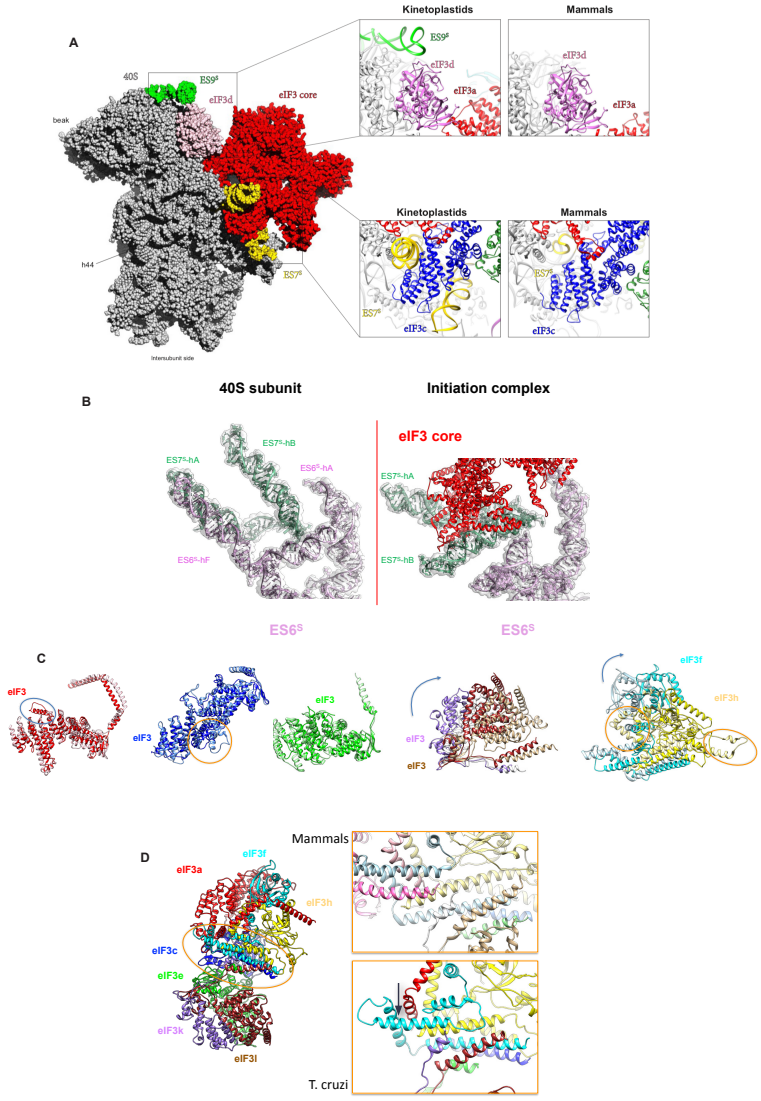


Fig. S8. Specific features of Kinetoplastidian eIF3 and its ribosome binding site. (A) Overall sphere representation of the *T.cruzi* 43S PIC showing kinetoplastidian specific rRNA oversized expansion segments (ESs) in contact with eIF3. Upper panel: comparison of the kinetoplastidian and mammalian eIF3d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

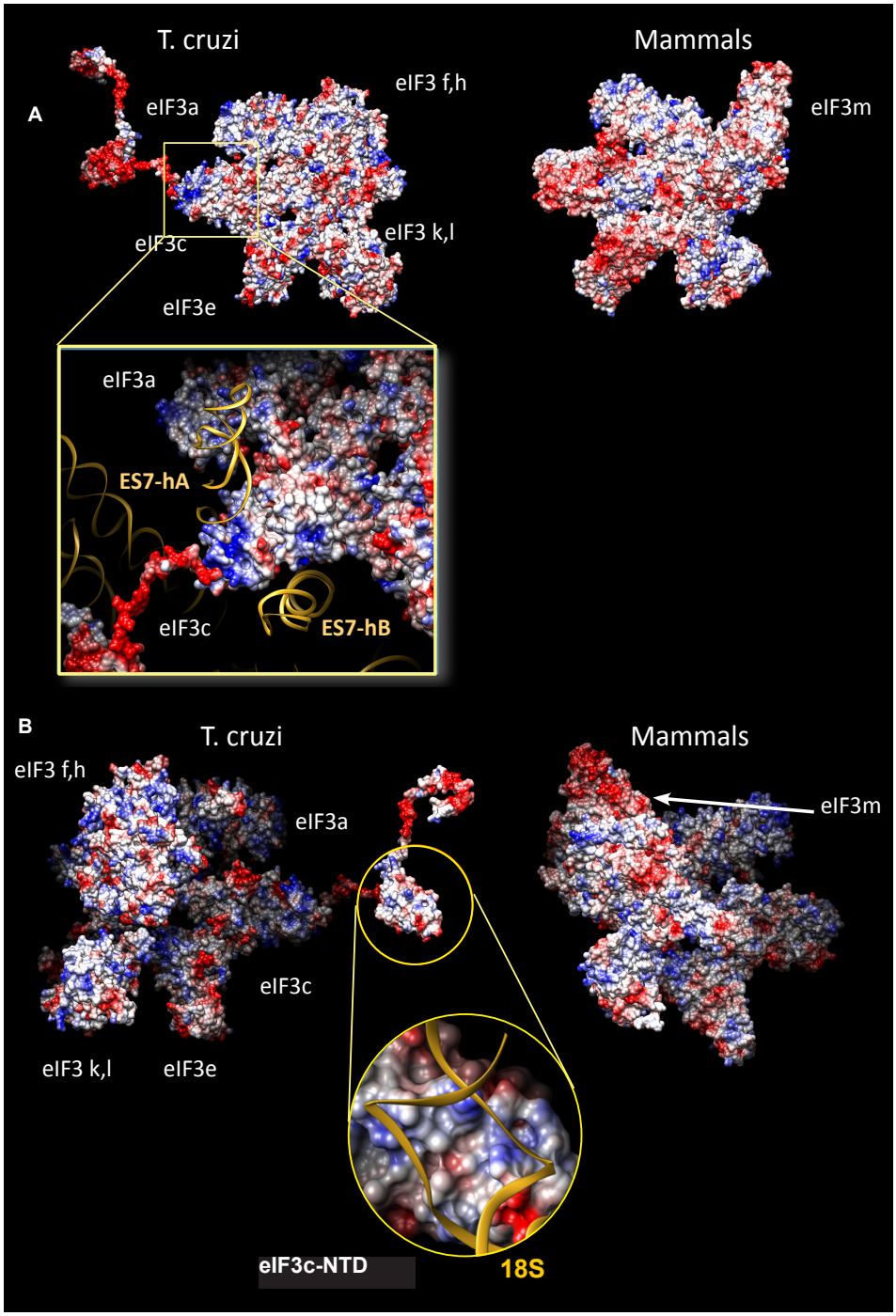


Fig. S9. Charge surface analysis of the T.cruzi and mammalian eIF3 structures. (A) Surface representation of the *T. cruzi* (left) and mammalian (right) eIF3 structure seen from the 40S platform side. Lower panel: close-up view of *T.cruzi* eIF3c and its interaction with the ES7^{s} helix A and helix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

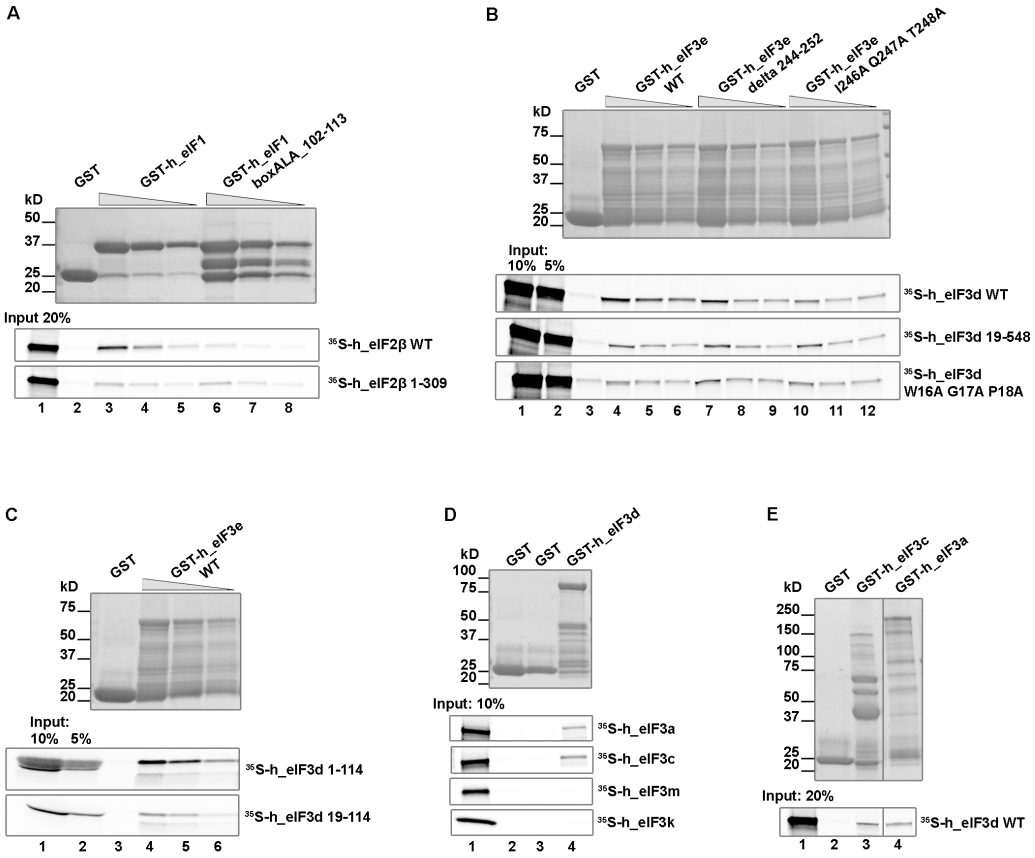


Fig. S10. *In vitro* analysis of eIF3 intersubunit interactions. (**A**) *In vitro* protein-protein binding analysis of the interaction between the *in vitro* translated human ^{35}S-labeled eIF2β and its C-terminal truncation (eIF2β 1-309) against wild type eIF1 or its mutated variant (eIF1-boxAla-102-113;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

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

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

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

Article 3: Interaction Networks of Ribosomal Expansion Segments in Kinetoplastids

In Brief

This chapter presents the interaction networks of ribosomal expansion segments in kinetoplastids. It compares the interaction of rRNA ES from four kinetoplastid parasites: *Leishmania major, Leishmania donovani, Trypanosoma brucei* and *Trypanosoma cruzi.* We analyze the interaction networks of their ESs in a comparative way to highlight conservation and differences of these unusual structured rRNA extensions. Our observations suggest the role of ESs during translation.

Chapter 22

Interaction Networks of Ribosomal Expansion Segments in Kinetoplastids

Quentin Vicens[1#], Anthony Bochler[1,3], Amy Jobe[2], Joachim Frank[2*], Yaser Hashem[1,3*]

[1] Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, 15 rue René Descartes, F-67000 Strasbourg, France.

[2] Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA.

[3] Inserm, Université de Bordeaux, Institut Européen de Chimie et Biologie, U1212, 33600 Pessac, France.

Present addresses:

[#] Department of Biochemistry and Molecular Genetics, University of Colorado Denver School of Medicine, Aurora, CO 80045, USA.

*Corresponding authors: J.F. +1 212 305 9510, jf2192@cumc.columbia.edu; Y.H. +33 5 40 00 88 22, yaser.hashem@inserm.fr

Abstract Expansion segments (ESs) are insertions of a few to hundreds of nucleotides at discrete locations on eukaryotic ribosomal RNA (rRNA) chains. Some cluster around 'hot spots' involved in translation regulation and some may participate in biogenesis. Whether ESs play the same roles in different organisms is currently unclear, especially since their size may vary dramatically from one species to another and very little is known about their functions. Most likely, ES variation is linked to adaptation to a particular environment. In this chapter, we compare the interaction networks of rRNA ES from four kinetoplastid parasites, which have evolved in diverse insect vectors and mammalian hosts: *Trypanosoma cruzi*, *Trypanosoma brucei*, *Leishmania donovani* and *Leishmania major*. Here, we comparatively analyze ribosome structures from these representative kinetoplastids and ascertain meaningful structural differences from mammalian ribosomes. We base our analysis on sequence alignments and three-dimensional structures of 80S ribosomes solved by cryo-electron microscopy (cryo-EM). Striking differences in size are observed between ribosomes of different parasites, indicating that not all ES are expanded equally. Larger ESs are not always matched by large surrounding ESs or proteins extensions in their vicinity, a particularity that may lead to clues about their biological function. ESs display different species-specific patterns of conservation, which underscore the density of their interaction network at the surface of the ribosome. Making sense of the conservation patterns of ESs is part of a global effort to lay the basis for functional studies aimed at discovering unique kinetoplastid-

specific sites suitable for therapeutic applications against these human and often animal pathogens.

Keywords Expansion segment; Kinetoplastid parasite; Ribosome structure.

Why expansion segments?

Because protein synthesis is a universal necessity for all organisms, ribosome function, overall shape and core architecture are conserved across species. Yet, ribosome composition varies considerably between different domains of life. For example, ribosomes from yeast and human are respectively ~40% and ~85% larger than bacterial ribosomes (Melnikov et al. 2012). In addition to comprising more proteins than bacterial ribosomes, eukaryotic ribosomes also contain elongated ribosomal RNAs (rRNA). This 'extra RNA' forms expansion segments (ES) (Clark et al. 1984; Hassouna et al. 1984; Ware et al. 1983).

Whether ES are all mere additions that are tolerated by the translation machinery, or whether they have important regulatory roles remains for the most part an open question. Most ES fall outside of the highly conserved ribosomal core (Bernier et al. 2018; Melnikov et al. 2012). Their size varies across species, although their patterns of insertion are conserved (Gerbi 1996). We know from early structural studies that ES and associated proteins portions comprise somewhat flexible substructures on the solvent-facing side, where they may interact with one another and with proteins to form an outer layer (Anger et al. 2013; Armache et al. 2010; Gao et al. 2005). This web-like organization is more rigid where contacts with the ribosome core are tighter, which argues against a superfluous nature. ES are generally best perceived as peripheral elements that facilitate folding, assembly, or activity. This view is to some extent similar to the recurring insertions of > 200 nt reported in other ribozymes (e.g., group I introns, RNase P, and hammerhead (Kachouri et al. 2005; Lehnert et al. 1996; Michel and Westhof 1990; Westhof and Massire 2004; Przybilski et al. 2005)).

What do we currently know about the function of ES? On the small ribosomal subunit (SSU) of protozoan ribosomes for instance, a network of three ES (ES3S, ES6S, ES7S) span > 150 Å in the vicinity of the mRNA exit channel (Gao et al. 2005; Hashem et al. 2013b). This cluster of ES could to be involved in translation initiation through the regulation of recruitment of initiation factors such as eIF3 (Hashem et al. 2013a; Hashem et al. 2013b). We currently hypothesize that this platform is required for translation of protozoan mRNAs, which have particular features compared to most know eukaryotic mRNAs (a hypermodified cap, presence of a 39 nt leader sequence that is trans-spliced to the 5' UTR of all kinetoplastid mature mRNAs) (Gao et al. 2005; Hashem et al. 2013b). ES may also participate in ribosome biogenesis, as illustrated by the finding that 12 out of the 16 ES on the large subunit (LSU) of the yeast ribosome are required for optimal growth and correct 25S rRNA processing (Ramesh and Woolford 2016). The roles associated with

certain ES in a particular organism may be conserved across various species. For instance, ES27L on the LSU was shown to be consistently alternating between two positions (near L1 or the peptide tunnel exit) from the structures of the human, wheat, fly and yeast ribosomes (Anger et al. 2013; Armache et al. 2010; Beckmann et al. 2001). This observation suggests a similar role, especially since replacing ES27L with that of another species restores function (Sweeney et al. 1994).

In this chapter, we focus on ribosomal ES from the Trypanosoma genus (ES in other eukaryotes have been discussed elsewhere (Anger et al. 2013; Li et al. 2017)). Tritryps kinetoplastid parasites, i.e. *Trypanosoma cruzi, Trypanosoma brucei* spp. and *Leishmania* spp. are human pathogens causing the Chagas disease, African trypanosomiasis and leishmaniasis, repectively. Chagas disease alone affects 8 million people in the Americas (PAHO 2017), and due to migration flows it has now emerged in non-endemic countries such as Spain and the United States (> 300,000 infected individuals in the US (Bern and Montgomery 2009)) (Gascon et al. 2010). Current treatments rely on toxic drugs that are often effective only during specific stages of the disease, further highlighting the urge to develop new strategies for targeting these parasites (Field et al. 2017). Being highly distinctive from those of mammalian ribosomes (Gao et al. 2005) (Hashem et al. 2013b), theses ES in kinetoplastids could potentially represent promising targets for therapeutic applications. Efficiently targeting kinetoplastid-specific ES requires a deep structural understanding of their organization and their function, which is the motivation behind the comparative analysis we present here.

Expansion segments in kinetoplastid ribosomes are large but structurally organized

Biochemical and structural studies on ribosomes from kinetoplastids pinpointed unique features for two *Trypanosoma* species. The first structure of the ribosome from *T. cruzi* was solved at 12 Å resolution (Gao et al. 2005) and that of *T. brucei brucei* (referred to as *T. brucei* in the following) at 5.6 Å resolution (Hashem et al. 2013b). We can now compare these structures to higher resolution structures of the 80S from *L. donovani* (Zhang et al. 2016; Shalev-Benami et al. 2016; Shalev-Benami et al. 2017) and *T. cruzi* (Liu et al. 2016, Querido et al. 2017), which we revised and/or expanded from the published models (Figure 22.1 a and b). We also include in this analysis a previously unpublished structure of the *L. major* 80S (Figure 22.1c) at ~5Å resolution (the structure will be deposited in the protein data bank (PDB). Images were collected on a FEI Polara electron microscope equipped with a Summit K2 Camera. After image processing, nearly 150 particle images were used in the final reconstruction). Sequence conservation between the *L. donovani* and *L. major* ribosomes is > 98% (rRNA) and > 97% (r-protein; except for eS24 ~ 93%). Our strategy was to (re)model the ES for these four species of kinetoplastids according to rRNA structural alignments that we generated using non-redundant sequences in SSU-Align (Nawrocki 2009), which we manually adjusted in Ugene

(Okonechnikov et al. 2012) and validated using R-scape (Rivas et al. 2017). ES were modeled outside of the cryo-EM map density boundaries of their respective ribosomes (Figure 22.1 a and b), in order to better visualize size differences across ribosomes, as had been done previously for other eukaryotic ribosomes (Anger et al. 2013). Because of their high level of flexibility, ES are usually poorly resolved by cryo-EM and other structural studies.

A side-by-side comparison of the ribosomes from *T. brucei*, *L. donovani*, *L. major*, and *T. cruzi* reveals that many ES are several folds larger than their counterparts in mammals (Figure 22.2). Although the 28S rRNA in the LSU of *T. brucei* and *T. cruzi* is halfway between yeast and human, their 18S in the SSU is > 20% longer than their human counterpart (Table 22.1). Most of this extra rRNA sequence is found within ES6S and ES7S that are inserted within helices 21 and 26. ES6S and ES7S are respectively 230 and 17 nucleotides long in *H. sapiens*, but 429 & 136 and 512 & 164 nucleotides long in *L. major* and *T. cruzi*. ES6S, ES7S, but also ES9S and ES31L adopt kinetoplastid-specific tertiary structures and they cluster around the mRNA channel exit. In particular, ES6S, ES7S and ES9S are located at a very strategic region where initiation factor 3 is known to bind in mammals, forming bridges between the LSU and SSU that are not found in other eukaryotic ribosomes (Hashem et al. 2013b).

From aligning > 80 sequences for the SSU and ~ 25 for the LSU, we found that all ES from kinetoplastids have a conserved structure, with dramatic extensions for some helical regions (Figures 22.3, 22.4). This confirms and expands prior findings obtained with fewer than 10 kinetoplastid sequences for all ES (Hashem et al. 2013b), and with 3,000 eukaryotic sequences for ES6S only (Wuyts et al. 2000). The proportion of conserved vs. variable regions varies among ES, with ES3S (~ 200 nt long) being the most conserved. Large size variations are observed even between kinetoplastids. For example, ES6S, ES7S, ES7L, ES19L and ES27L have a comparable size among the genera *Angomonas*, *Herpetomonas* and *Phytomonas*. The same ES are significantly larger in *Leishmania* and *Trypanosoma* species, the largest being found in *T. cruzi*. Although ES3S, ES9S, ES10S and ES31L are all larger in *Leishmania* spp. and in *T. brucei* than in *T. cruzi*, the size difference remains relatively modest (Table 22.1).

Expansion segments form kinetoplastid-specific interaction networks

The overall high degree of ES conservation can be explained by interaction networks at the ribosome surface. For example, in all cryo-EM structures of kinetoplastids, ES3S is tightly packed as it forms the left foot of the SSU (Figure 22.3a). ES3S interacts with ES2S, domains B and D of ES6S (Figure 22.4a), ES12S (forming the right foot of the SSU) and r-proteins eS4, eS6, eS8 and KSRP (Figure 22.3 a and b). The structure of the pseudoknot formed between ES3S and ES6S is

conserved across parasitic and eukaryotic ribosomes (of known structures), as reported earlier (Zhang et al. 2016; Hashem et al. 2013b; Armache et al. 2010; Alkemar and Nygard 2006). The kinetoplastid-specific r-protein (KSRP) that we identified in *T. cruzi* (Brito Querido et al. 2017) is similarly bound to ES3S and ES6S in *Trypanosoma* and *Leishmania,* through its two conserved RNA recognition motifs (RRM) and a C-terminal helix that also interacts with eS6 (Figure 22.3c). The only variable region within ES3S is the Ab stem, which nonetheless conserves its spatial arrangement as it bends around eS8 in the four kinetoplastids models (Figure 22.3d). It is important to highlight that ES3S simultaneously displays structural characteristics conserved in all eukaryotes (pseudoknot with ES6S) and unique features only found in kinetoplastids (KSRP binding).

Even for larger and more variable ES such as ES6S and ES7S, their core shows a high degree of conservation (Figure 22.4a), supporting a conserved topology. ES6S is the largest ES on the SSU (400–500 nt) and its highly conserved domains A–D interact tightly with r-proteins and KSRP at the SSU solvent-exposed surface. ES6S interacts also with ES7S that is of a similar size in *T. cruzi* and *T. brucei*, and ~ 25% larger in *T. cruzi* than in *L. donovani*. In contrast to stem Ab in ES3S, whose length changes only slightly across parasites, stems E1–E3 in ES6S roughly double in size from *T. brucei* through *Leishmania* to *T. cruzi* (Figure 22.4 a and b). As a result, the number of stem-loops protruding into the solvent is different among parasites (Figure 22.4b). Both the nature and the presence of these stems dictate the positioning of the conserved stems of ES7S, which are possibly more conformationally restricted in *Leishmania* and *T. cruzi* than in *T. brucei*. Note that although the lack of density prevented pinpointing the exact location of the expansions seen in E2 and E3, the interaction between E3 and ES7S seen in *T. brucei* cannot occur for the other three model parasites, where E3 is much longer (Figure 22.4b). These structural differences probably reflect the dynamic properties of this area that is a 'hotspot' for translation initiation regulation (Hashem et al. 2013a).

ES 9S, 10S and 12S on the other hand are among the smallest SSU ES. ES12S for example is an extension of h44 by 9–10 base pairs in all kinetolastids, while ES10S exists as a 5–9 base pair stem only in *T. brucei, T. congolense, T. simiae, and T. vivax* (Figure 22.5). The shorter ES seem to follow a similar trend as the longer ES, thus presenting conserved cores and variable stems.

On the LSU, ES are concentrated on two regions, the L1-stalk region, and the solvent-exposed face (Figure 22.2b). The ES at the solvent-exposed face appear less interconnected compared to the SSU ES. ES7L is isolated from the other LSU sizable ES. It presents a two-pronged fork shape (Figures 22.6a, 22.7c). Because of its numerous interactions with r-proteins (Figure 22.6a), its densities are relatively well resolved and present a good local resolution. It forms a pseudoknot with the small ES15L (Figure 22.6b and d). Another more isolated ES is 5L (also called ES42L) (Figure 22.6e). Although close to ES7L, no direct interactions were observed with

the latter. It presents a high level of sequence conservation and interacts with r-proteins eL8, eL15 and uL29 (Figure 22.6e).

On the other hand, LSU ES near the L1-stalk region are interconnected, where ES19L, ES27L and ES31L are clustered (Figures 22.2b, 22.6). The length of ES31L varies only within a range of ~ 20 nucleotides among kinetoplastids, while that of ES19L and ES27L increases by ~ 50% between *L. donovani* and *T. cruzi* (Table 22.1). As a result, stem F of ES31L adopts different orientations in the various structures, where it is constricted to various degrees by nearby ES (Figure 22.4b). ES31L is highly structured, as it comprises four nested three-way junctions but a high degree of sequence variability (Figure 22.6h). In spite of its sophisticated predicted tertiary structure, its cryo-EM corresponding densities are poorly resolved in all four kinetoplastid model ribosomes because of its high flexibility. Although ES31L interacts with r-proteins eL27 and eL34 (Figure 22.6h), the bulk of its rRNA is not stabilized by any other r-proteins, which explains its high level of flexibility. The cryo-EM reconstruction from several kinetoplatids strongly suggest its interaction with ES19L and ES27L because of its proximity to the latter ES (Figures 22.2b, 22.6g). ES19L appears as an elongated stem of a relatively well resolved density that shows its interactions with r-proteins eS27 and uL23, but also with srRNA1 (Figures 22.6f). It is interesting to highlight that ES31L and ES19L interaction occurs indirectly through r-protein eL27 (Figures 22.6 f and h). Last but not least, ES27L contains two three-way junctions conferring it an elongated shape with two grafted branches protruding towards the solvent (Figure 22.6i). It interacts directly with ES31L and r-proteins eL19 and eL22 (Figure 22.6i). Thus, all three, ES 19L, 27L and 31L form an interconnected bundle of rRNA near the L1-stalk on the LSU (Figure 22.2b).

Finally, it is worth mentioning that expansion segments may have a conserved structure in the absence of sequence conservation, as exemplified by ES7S (Figures 22.2, 22.4b), and subdomain A3 of ES7L (Figures 22.6a, 22.7c).

Compensatory interaction networks

Compensatory interaction networks are widespread on the LSU, as previously observed upon comparing ribosomes from yeast, wheat, fly, human and several parasites (Anger et al. 2013; Armache et al. 2010; Li et al. 2017; Hashem et al. 2013b). For instance, the larger ES31L in kinetoplastids (70 nt in yeast, 85 nt in human, but 220–237 nt in kinetoplastids) fills the space occupied by an extended helix from the 714 nt-long ES27L in humans, without altering the binding mode of eL27 and eL8 (Figure 22.7a). Another example is the conserved stem A1 of ES7L that packs against eL6, eL32 and eL33 but not eL28, in contrast to ES7L of other eukaryotic ribosomes (Anger et al. 2013; Ben-Shem et al. 2011; Hashem et al. 2013b; Li et al. 2017) (Figure 22.6 a and d). Moreover, although the pseudoknot formed between ES7L and ES15L is similar to that in *T. gondii* (Figure 22.6 b and d), the long-range

interaction involving the shallow groove of ES7L and a loop from ES39L in human is replaced by interactions with srRNA3 (Figure 22.6c).

Size increase of ES is often accompanied by extensions of r-proteins (Hashem et al. 2013b; Li et al. 2017). For example, the eL34 r-protein is extended in kinetoplastids by ~ 50 amino acids compared with human eL34, which can be explained by a larger ES31L (Figure 22.7b). It therefore came as a surprise that the length of uL4 remains the same across kinetoplastids, even though stem A3 of ES7L nearby dramatically varies (Figure 22.7c). Nevertheless, counterintuitive exceptions can be observed, such as in *P. falciparum* and *T. gondii* where ES7L is of a similar size compared to that of *T. brucei* (~ 290 nt). However, uL4 is elongated by ~ 40 amino acids in *P. falciparum* and *T. gondii*, so that it interacts with uL16 (Figure 22.7c), a late binding protein during ribosome assembly in yeast (Li et al. 2017; Ma et al. 2017), with no obvious correlation with the size of ES7L (Figure 22.7c).

Concluding remarks

In line with previous work that presented the structures and underscored the importance of expansion segments in other species (Armache et al. 2010; Beckmann et al. 2001; Ben-Shem et al. 2011), this chapter illustrates that ES are not random extensions with poor sequence conservation at the surface of ribosomes. ES are organized so that they tightly interact with core rRNA and r-proteins, recruit kinetoplastid-specific proteins, and act as dynamic gatekeepers. Some of these interaction networks are found in some but not all parasites, suggesting that the precise role of ES may vary even between species from the same phylum. Through highlighting the importance of ES networks in parasites, we contribute to better characterize kinetoplastid-specific ribosomal features, in the hope of stimulating the development of safer and more specific anti-parasitic therapeutic.

Authors contributions

QV and AB did the molecular modeling and cryo-EM maps (re)analysis. AJ and YH acquired *L. major* ribosome particles and processed the cryo-EM data. QV, YH and JF wrote the manuscript with input from all authors. QV and AB did the figures. YH and JF supervised the research.

Acknowledgments

We thank Robert A. Grassucci for assistance in data acquisition of *L. major*. This work was supported by the ANR grant ANR-14-ACHN-0024 @RAction program ‘‘ANR CryoEM80S’’ and the ERC-2017-STG #759120 “TransTryp” (to Y.H.).

Tables

Table 22.1. Size comparison of SSU and LSU expansion segments. The number of nucleotides is indicated for every conserved ES in seven eukaryotic species.

	L. donovani	*L. major*	*T. cruzi*	*T. brucei*	*S. cerevisiae*	*D. melanogaster*	*H. sapiens*
ES2S	22	22	22	22	18	16	21
ES3S	147	147	148	156	113	121	159
ES6S	429	429	512	406	222	300	230
ES7S	138	136	164	159	17	17	17
ES9S	76	76	76	89	45	151	52
ES10S	17	20	20	29	5	5	7
ES12S	25	25	26	26	38	41	43
Total for SSU	**854**	**855**	**968**	**887**	**458**	**651**	**529**
ES5L	42	42	47	47	45	50	50
ES7L	203	203	291	219	200	331	866
ES15L	32	30	39	32	15	41	189
ES19L	58	58	124	63	29	25	39
ES27L	213	213	323	234	159	222	714
ES31L	237	236	220	235	70	208	85
Total for LSU	**785**	**782**	**1044**	**830**	**518**	**877**	**1893**
Total number of ES ntds	**1639**	**1637**	**2012**	**1717**	**976**	**1528**	**2422**

Figures

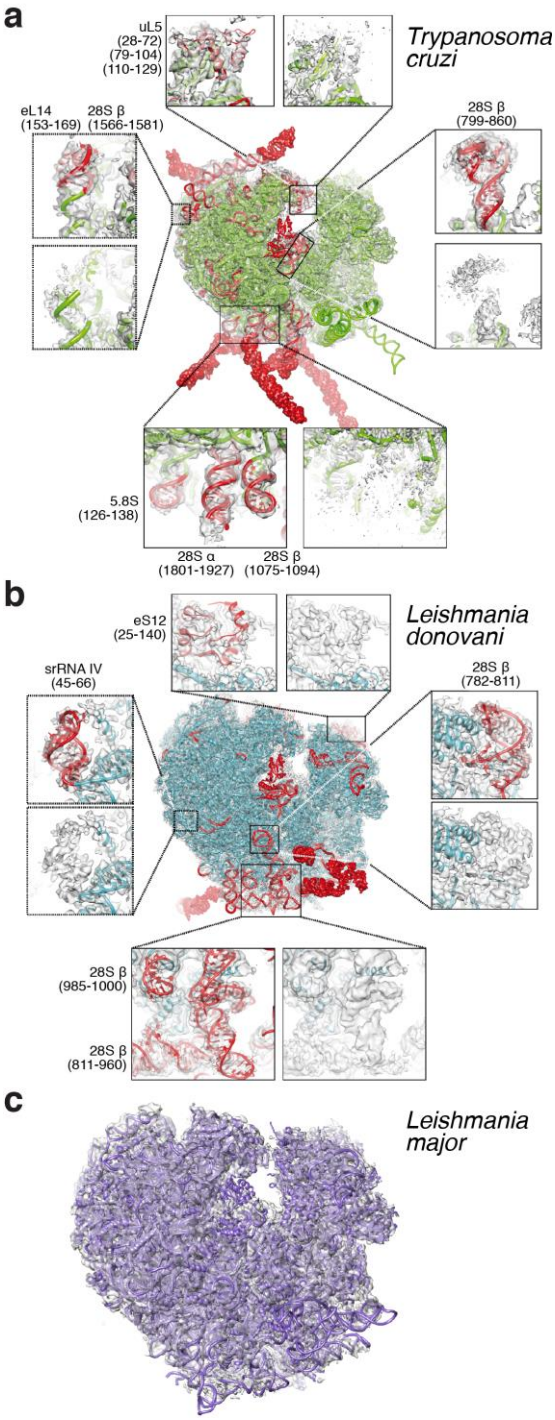

a
Trypanosoma cruzi
uL5 (28-72) (79-104) (110-129)
eL14 (153-169)
28S β (1566-1581)
28S β (799-860)
5.8S (126-138)
28S α (1801-1927)
28S β (1075-1094)
b
Leishmania donovani
eS12 (25-140)
srRNA IV (45-66)
28S β (782-811)
28S β (985-1000)
28S β (811-960)
c
Leishmania major

Figure 22.1. Strategy for modeling flexible espansion segments in ribosomes. (a) Modeling the complete *T. cruzi* 80S at 4 Å resolution. We modeled expansion segments in the SSU of our previously published model (PDB ID: 5OPT (Brito Querido et al. 2017; Pitula et al. 2002; Ayub et al. 2012)). We also included coordinates from a 2.5 Å-structure of the 60S particle from *T. cruzi* (PDB ID: 5T5H (Liu et al. 2016)). RNA and protein elements were added (a total of 1816 nucleotides and 810 aminoacids) are shown in red. Residues built within density limits (~8 to 15 Å) are shown in ribbon mode, while residues beyond are shown as surfaces. Insets illustrate representative regions of the original unfiltered maps. Dashed boxes point to regions located on the side of the ribosome hidden from view. (b) The quality of the *L. donovani* electron density map at 2.9 Å resolution (PDB ID: 5T2A (Zhang et al. 2016)) enabled us to add 873 nucleotides and 802 aminoacids to the original model (in red). (c) The revised model of the *L. donovani* ribosome was used as the basis for building the structure of the ribosome from *L. major* at 5 Å resolution. The *L. major* map and coordinates will be deposited in the PDB.

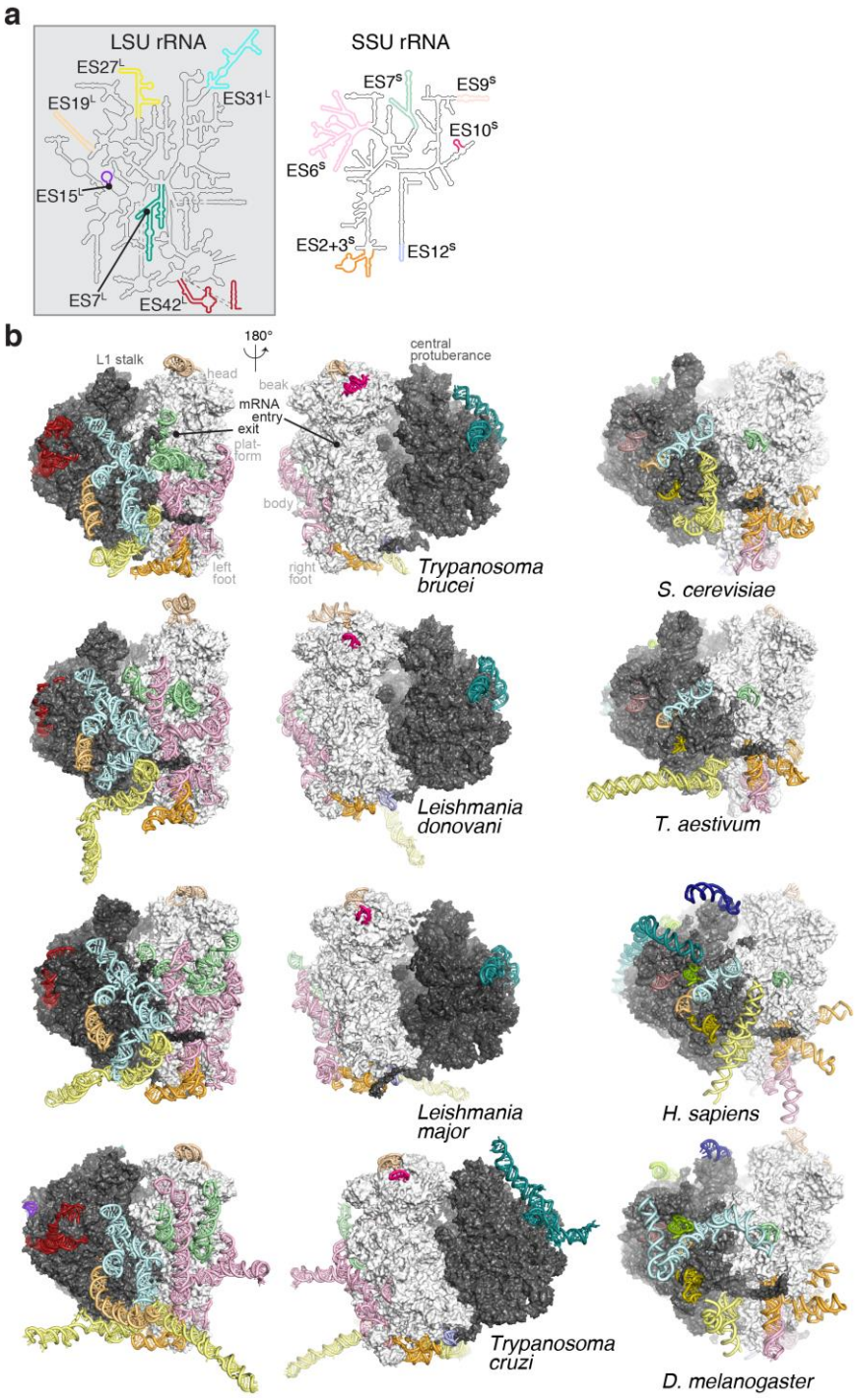

a
LSU rRNA
ES27L
ES19L
ES31L
ES15L
ES7L
ES42L
SSU rRNA
ES7S
ES9S
ES10S
ES6S
ES2+3S
ES12S
b
180°
L1 stalk
head
beak
central protuberance
mRNA entry
exit
plat-form
body
left foot
right foot
Trypanosoma brucei
S. cerevisiae
Leishmania donovani
T. aestivum
Leishmania major
H. sapiens
Trypanosoma cruzi
D. melanogaster

Figure 22.2. The complex outer layer of expansion segments at the surface of ribosomes from kinetoplastids. (a) Secondary structure schematic for the LSU and SSU rRNA of kinetoplastids. Expansion segments are highlighted in colors and labeled according to conventional numbering (Gerbi 1996). (b) Overall structure of the 80S ribosomes from four kinetoplastids, viewed from the mRNA L1-stalk side (left) and SSU beak sides (middle). Ribosomes are shown in surface view, except for ES rendered as ribbons and colored as in panel (a). (Right) Ribosome structures from other organisms are shown from the L1-stalk side as a reference (references in the main text).

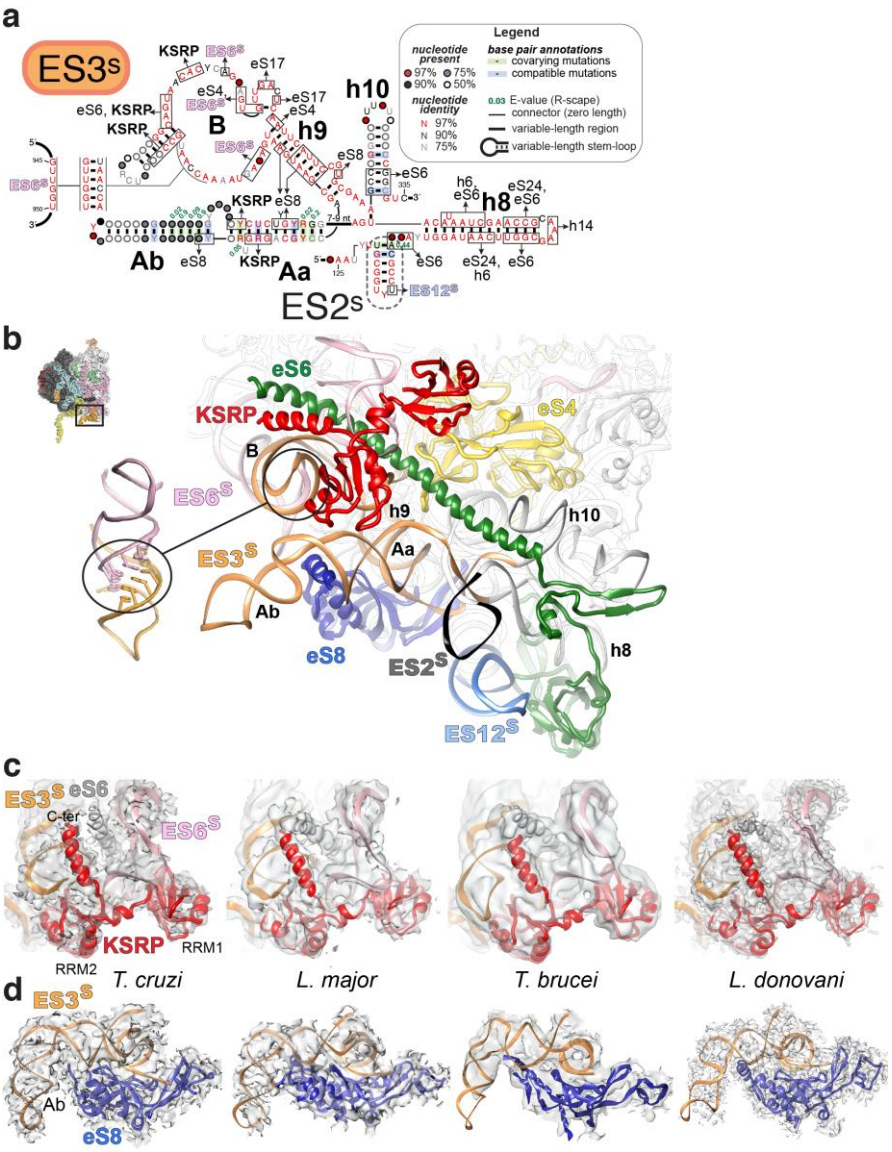


Figure 22.3. Conservation of interaction networks between kinetoplastids. (a) Interaction network and consensus secondary structure of ES3S. Note that ES2S forms a hairpin at the 5' of h8. Statistically significant covarying base pairs are shown in green (E-value < 0.05 (Rivas et al. 2017)). This and subsequent covariation models were generated using R2R (Weinberg and Breaker 2011). (b) Close-up of the left foot of the ribosome from *L. donovani*, showing the arrangement of several ES and r-proteins. (c) Conserved binding of KSRP, which interacts with ES3S,

ES6S and eS6. (d) Conserved spatial arrangement of stem Ab of ES3S and r-protein eS8. Sequence alignments for all ES are available directly from the authors upon request.

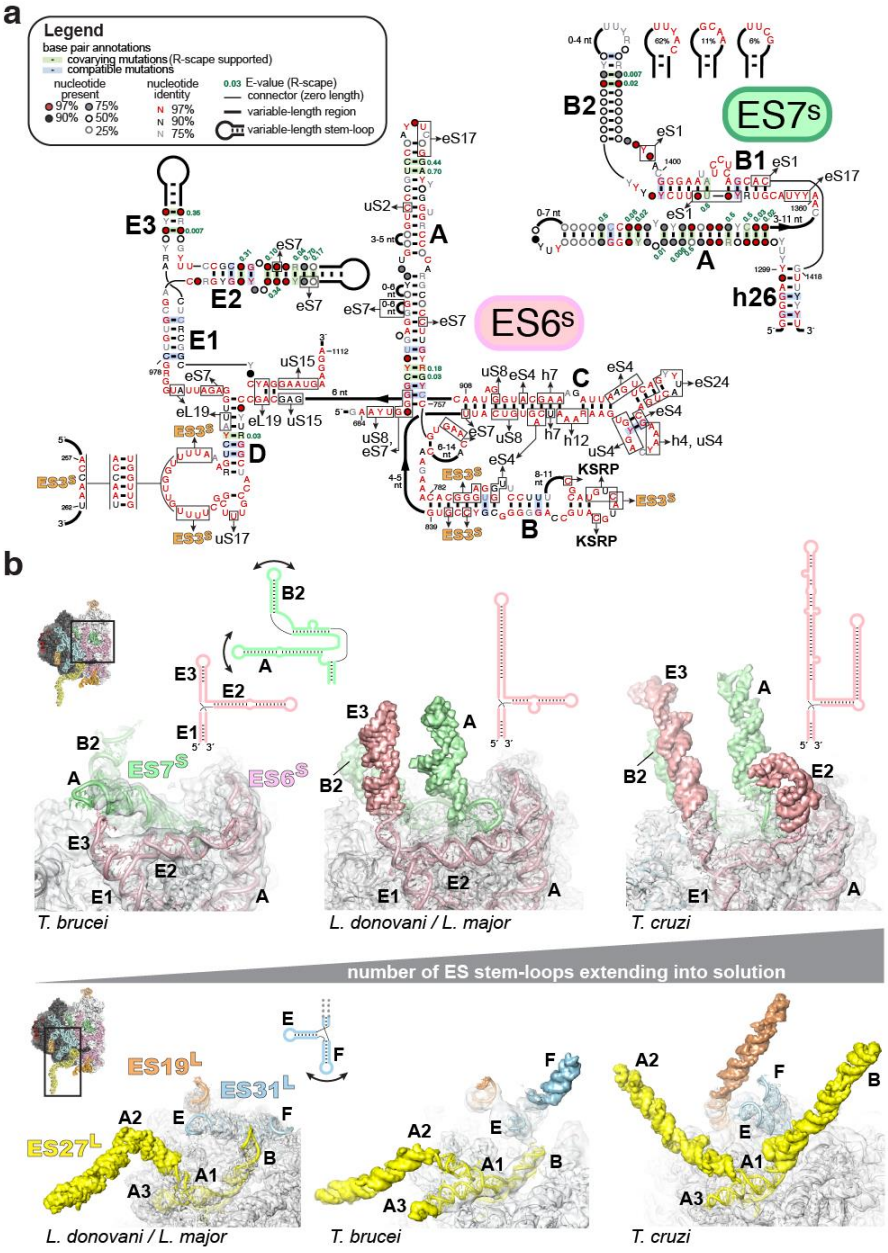


Figure 22.4. Sequence and structure variability among ES is significant but localized. (a) Interaction networks and consensus secondary structures of ES6S and ES7S (color-coding as in Figure 22.3). (b) Conformational variability as a function of the length of certain ES across kinetoplastids. Top: The size of ES7S barely changes across kinetoplastids, in contrast to that of ES6S. The increasing size of ES6S leads to displacement of ES7S helices. The regions covered by the electron density maps (grey) are shown as ribbons, while the sections modeled beyond the limit of the current experimental maps are shown as surfaces. Bottom: Conformational changes in particular of ES31L occur as the length of the surrounding ES19L and ES27L increases.

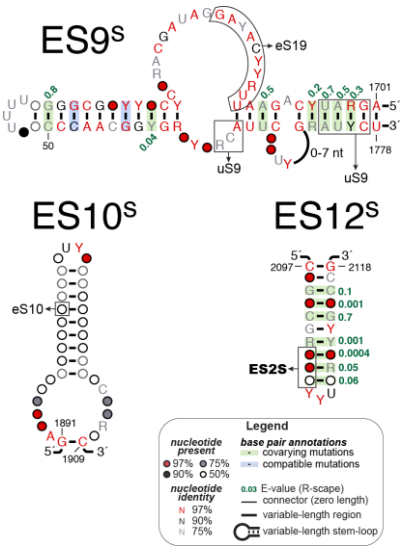
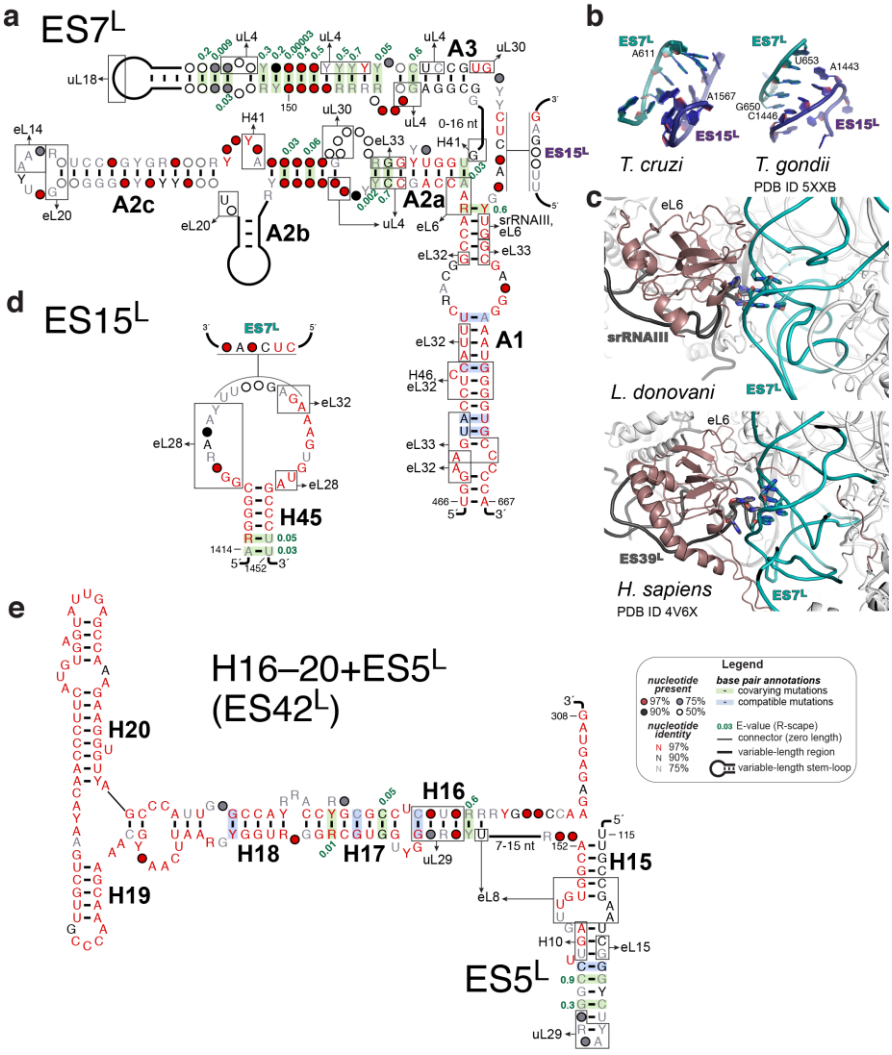


Figure 22.5. Covariation analysis of SSU expansion segments. Consensus sequences and secondary structures for ES9S, ES10S and ES12S, with interacting partners as on Figures 22.3–4.


a
ES7L
uL18
uL4
A3
uL30
H41
eL14
eL20
A2c
A2b
A2a
ES15L
eL6
srRNAIII,
eL6
eL32
eL33
A1
H46
eL32
eL33
eL32
466–667
5′ 3′
b
ES7L
A611
A1567
ES15L
T. cruzi
ES7L
U653
A1443
G650
C1446
ES15L
T. gondii
PDB ID 5XXB
c
eL6
srRNAIII
L. donovani
ES7L
eL6
ES39L
H. sapiens
PDB ID 4V6X
ES7L
d
ES15L
ES7L
eL32
eL28
eL28
H45
1414–1452
e
H16–20+ES5L
(ES42L)
H20
H19
H18
H17
H16
uL29
7-15 nt
eL8
H15
H10
eL15
ES5L
uL29
308–
115
152
Legend
nucleotide present
97% 75%
90% 50%
nucleotide identity
N 97%
N 90%
N 75%
base pair annotations
covarying mutations
compatible mutations
0.03 E-value (R-scape)
connector (zero length)
variable-length region
variable-length stem-loop

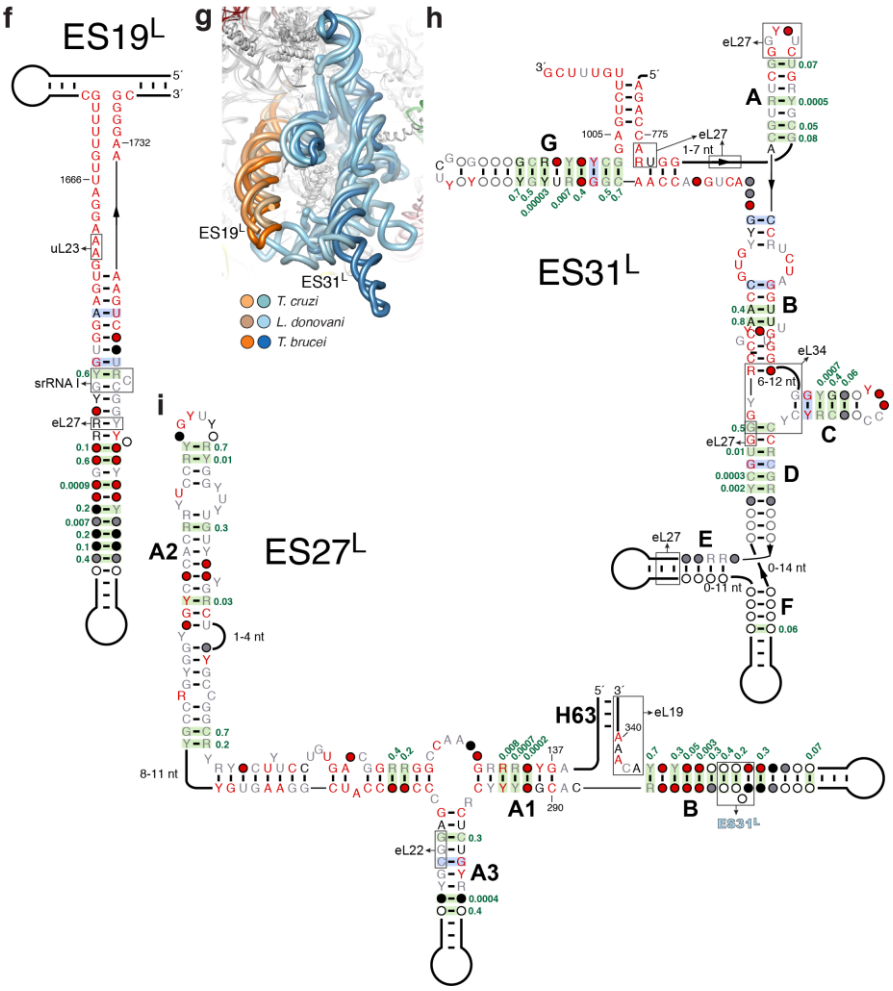


Figure 22.6. Covariation analysis and interaction networks of LSU expansion segments. (a) Consensus sequence and secondary structure for ES7L. (b) Close-up of the pseudoknot formed by ES7L and ES15L in *T. cruzi* and *T. gondii* (PDB ID: 5XXB (Li et al. 2017)). (c) Close-up of the region at the top of stem A1 of ES7L in *L. donovani* and *H. sapiens*. (d–f) Consensus sequences and secondary structures for ES15L, ES5L (also known as ES42L) and ES19L (interacting partners are not shown for H16–20). (g) Close-up of the region comprising ES19L and ES31L. (h–i) Same as (d–f) for ES31L and ES27L.

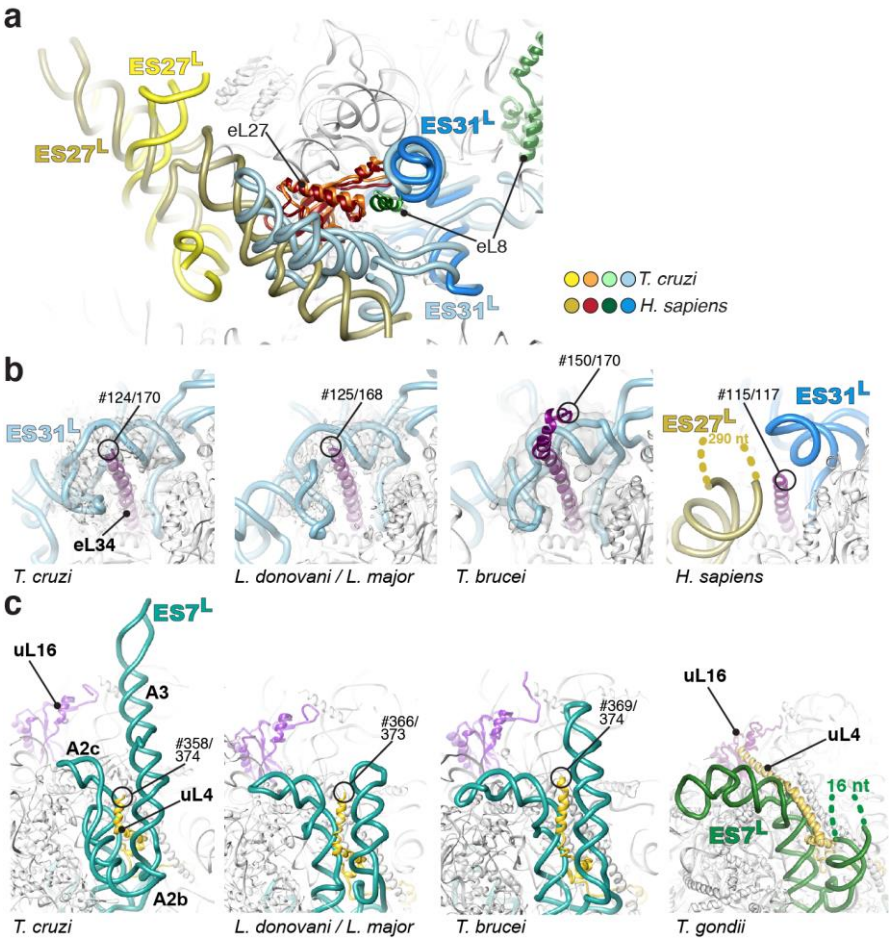


Figure 22.7. Compensatory interaction networks on the LSU. (a) Superimposition of the ribosomes from *T. cruzi* (light shades) and *H. sapiens* (PDB ID: 4V6X; dark shades (Anger et al. 2013)). The view is from the top of that shown in panel 22.4b, after removing regions modeled outside of density. (b) Close-up of the location where eL34 binds ES31L in kinetoplastids, and of the corresponding region in the human ribosome (PDB ID: 4V6X). For kinetoplastids, the number of the last residue modeled in density as well as the total length of the r-protein are indicated. (c) Close-up of ES7L bound to uL4 in kinetolastids, and of the corresponding region in the *T. gondii* ribosome (PDB ID: 5XXB (Li et al. 2017)), where an elongated uL4 interacts with uL16.

References

Alkemar G, Nygard O (2006) Probing the secondary structure of expansion segment ES6 in 18S ribosomal RNA. Biochemistry 45 (26):8067-8078. doi:10.1021/bi052149z

Anger AM, Armache JP, Berninghausen O, Habeck M, Subklewe M, Wilson DN, Beckmann R (2013) Structures of the human and Drosophila 80S ribosome. Nature 497 (7447):80-85. doi:10.1038/nature12104

Armache JP, Jarasch A, Anger AM, Villa E, Becker T, Bhushan S, Jossinet F, Habeck M, Dindar G, Franckenberg S, Marquez V, Mielke T, Thomm M, Berninghausen O, Beatrix B, Soding J, Westhof E, Wilson DN, Beckmann R (2010) Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-A resolution. Proc Natl Acad Sci U S A 107 (46):19748-19753. doi:10.1073/pnas.1009999107

Ayub AJ, Lapadula WJ, Hoebeke J, Smulski CR (2012) Ribosomes from Trypanosomatids: Unique Structural and Functional Properties. In: Biyani M (ed) Cell-Free Protein Synthesis. Intech, pp 3-28

Beckmann R, Spahn CM, Eswar N, Helmers J, Penczek PA, Sali A, Frank J, Blobel G (2001) Architecture of the protein-conducting channel associated with the translating 80S ribosome. Cell 107 (3):361-372

Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M (2011) The Structure of the Eukaryotic Ribosome at 3.0 A Resolution. Science, vol 334. doi:10.1126/science.1212642

Bern C, Montgomery SP (2009) An estimate of the burden of Chagas disease in the United States. Clin Infect Dis 49 (5):e52-54. doi:10.1086/605091

Bernier CR, Petrov AS, Kovacs NA, Penev PI, Williams LD (2018) Translation: The Universal Structural Core of Life. Mol Biol Evol 35 (8):2065-2076. doi:10.1093/molbev/msy101

Brito Querido J, Mancera-Martínez E, Vicens Q, Bochler A, Chicher J, Simonetti A, Hashem Y (2017) The cryo-EM structure of a novel 40S kinetoplastid-specific ribosomal protein. Structure in press

Clark CG, Tague BW, Ware VC, Gerbi SA (1984) Xenopus laevis 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications. Nucleic Acids Res 12 (15):6197-6220. doi:10.1093/nar/12.15.6197

Field MC, Horn D, Fairlamb AH, Ferguson MAJ, Gray DW, Read KD, De Rycker M, Torrie LS, Wyatt PG, Wyllie S, Gilbert IH (2017) Anti-trypanosomatid drug discovery: an ongoing challenge and a continuing need. Nat Rev Microbiol 15 (7):447. doi:10.1038/nrmicro.2017.69

Gao H, Ayub MJ, Levin MJ, Frank J (2005) The structure of the 80S ribosome from Trypanosoma cruzi reveals unique rRNA components. Proc Natl Acad Sci U S A 102 (29):10206-10211. doi:10.1073/pnas.0500926102

Gascon J, Bern C, Pinazo MJ (2010) Chagas disease in Spain, the United States and other non-endemic countries. Acta Trop 115 (1-2):22-27. doi:10.1016/j.actatropica.2009.07.019

Gerbi SA (1996) Expansion segments: regions of variable size that interrupt the universal core secondary structure of ribosomal RNA. In: Zimmermann RA, Dahlberg AE (eds) Ribosomal RNA: Structure, Evolution, Processing and Function in Protein Synthesis. Telford - CRC Press, pp 71-87

Hashem Y, des Georges A, Dhote V, Langlois R, Liao HY, Grassucci RA, Hellen CU, Pestova TV, Frank J (2013a) Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. Cell 153 (5):1108-1119. doi:10.1016/j.cell.2013.04.036

Hashem Y, des Georges A, Fu J, Buss SN, Jossinet F, Jobe A, Zhang Q, Liao HY, Grassucci RA, Bajaj C, Westhof E, Madison-Antenucci S, Frank J (2013b) High-resolution cryo-electron microscopy structure of the Trypanosoma brucei ribosome. Nature 494 (7437):385-389. doi:10.1038/nature11872

Hassouna N, Michot B, Bachellerie JP (1984) The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucleic Acids Res 12 (8):3563-3583. doi:10.1093/nar/12.8.3563

Kachouri R, Stribinskis V, Zhu Y, Ramos KS, Westhof E, Li Y (2005) A surprisingly large RNase P RNA in Candida glabrata. RNA 11 (7):1064-1072. doi:10.1261/rna.2130705

Lehnert V, Jaeger L, Michel F, Westhof E (1996) New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the Tetrahymena thermophila ribozyme. Chem Biol 3 (12):993-1009

Li Z, Guo Q, Zheng L, Ji Y, Xie YT, Lai DH, Lun ZR, Suo X, Gao N (2017) Cryo-EM structures of the 80S ribosomes from human parasites Trichomonas vaginalis and Toxoplasma gondii. Cell Res 27 (10):1275-1288. doi:10.1038/cr.2017.104

Liu Z, Gutierrez-Vargas C, Wei J, Grassucci RA, Ramesh M, Espina N, Sun M, Tutuncuoglu B, Madison-Antenucci S, Woolford JL, Jr., Tong L, Frank J (2016) Structure and assembly model for the Trypanosoma cruzi 60S

ribosomal subunit. Proc Natl Acad Sci U S A 113 (43):12174-12179. doi:10.1073/pnas.1614594113

Ma C, Wu S, Li N, Chen Y, Yan K, Li Z, Zheng L, Lei J, Woolford JL, Jr., Gao N (2017) Structural snapshot of cytoplasmic pre-60S ribosomal particles bound by Nmd3, Lsg1, Tif6 and Reh1. Nature Structural & Molecular Biology 24:214-220

Melnikov S, Ben-Shem A, Garreau de Loubresse N, Jenner L, Yusupova G, Yusupov M (2012) One core, two shells: bacterial and eukaryotic ribosomes. Nat Struct Mol Biol 19 (6):560-567. doi:10.1038/nsmb.2313

Michel F, Westhof E (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J Mol Biol 216 (3):585-610. doi:10.1016/0022-2836(90)90386-Z

Nawrocki EP (2009) Structural RNA Homology Search and Alignment Using Covariance Models. Washington University School of Medicine,

Okonechnikov K, Golosova O, Fursov M, team U (2012) Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28 (8):1166-1167. doi:10.1093/bioinformatics/bts091

PAHO (2017) Chagas disease. Pan American Health Organization. http://www.paho.org/hq/index.php?option=com_topics&view=article&id=10&Itemid=40743. Accessed October 20, 2017 2017

Pitula J, Ruyechan WT, Williams N (2002) Two novel RNA binding proteins from Trypanosoma brucei are associated with 5S rRNA. Biochem Biophys Res Commun 290 (1):569-576. doi:10.1006/bbrc.2001.6226

Przybilski R, Graf S, Lescoute A, Nellen W, Westhof E, Steger G, Hammann C (2005) Functional hammerhead ribozymes naturally encoded in the genome of Arabidopsis thaliana. Plant Cell 17 (7):1877-1885. doi:10.1105/tpc.105.032730

Ramesh M, Woolford JL, Jr. (2016) Eukaryote-specific rRNA expansion segments function in ribosome biogenesis. RNA 22 (8):1153-1162. doi:10.1261/rna.056705.116

Rivas E, Clements J, Eddy SR (2017) A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nat Methods 14 (1):45-48. doi:10.1038/nmeth.4066

Shalev-Benami M, Zhang Y, Matzov D, Halfon Y, Zackay A, Rozenberg H, Zimmerman E, Bashan A, Jaffe CL, Yonath A, Skiniotis G (2016) 2.8-A Cryo-EM Structure of the Large Ribosomal Subunit from the Eukaryotic Parasite Leishmania. Cell Rep 16 (2):288-294. doi:10.1016/j.celrep.2016.06.014

Shalev-Benami M, Zhang Y, Rozenberg H, Nobe Y, Taoka M, Matzov D, Zimmerman E, Bashan A, Isobe T, Jaffe CL, Yonath A, Skiniotis G (2017) Atomic resolution snapshot of Leishmania ribosome inhibition by the aminoglycoside paromomycin. Nat Commun 8 (1):1589. doi:10.1038/s41467-017-01664-4

Sweeney R, Chen L, Yao MC (1994) An rRNA variable region has an evolutionarily conserved essential role despite sequence divergence. Mol Cell Biol 14 (6):4203-4215

Ware VC, Tague BW, Clark CG, Gourse RL, Brand RC, Gerbi SA (1983) Sequence analysis of 28S ribosomal DNA from the amphibian Xenopus laevis. Nucleic Acids Res 11 (22):7795-7817. doi:10.1093/nar/11.22.7795

Weinberg Z, Breaker RR (2011) R2R--software to speed the depiction of aesthetic consensus RNA secondary structures. BMC Bioinformatics 12:3. doi:10.1186/1471-2105-12-3

Westhof E, Massire C (2004) Structural biology. Evolution of RNA architecture. Science 306 (5693):62-63. doi:10.1126/science.1104482

Wuyts J, De Rijk P, Van de Peer Y, Pison G, Rousseeuw P, De Wachter R (2000) Comparative analysis of more than 3000 sequences reveals the existence of two pseudoknots in area V4 of eukaryotic small subunit ribosomal RNA. Nucleic Acids Res 28 (23):4698-4708

Zhang X, Lai M, Chang W, Yu I, Ding K, Mrazek J, Ng HL, Yang OO, Maslov DA, Zhou ZH (2016) Structures and stabilization of kinetoplastid-specific split rRNAs revealed by comparing leishmanial and human ribosomes. Nat Commun 7:13223. doi:10.1038/ncomms13223

Article 4: Structural insights into the mammalian late-stage initiation complexes

In Brief

This paper presents two high-resolution structures of 48S initiation complexes from *Oryctolagus cuniculus* (rabbit) with two different messenger RNA (mRNA). The first structure is a late stage 48S initiation complex with native mRNA sequence at 3 Å. The second one is the same 48S initiation complex with a histone 4 (H4) mRNA at 3.5 Å. In these high-resolution structures we describe key interactions between the mRNA and the 18S ribosomal RNA (rRNA), ribosomal proteins (r-protein) and eukaryotic initiation factors (eIF) 1A, 2, 3. Our results suggest the impact of the sequence of mRNA on the structure of the 48S initiation complex.

Structural insights into the mammalian late-stage initiation complexes

Angelita Simonetti[1,3], Ewelina Guca[2,3], Anthony Bochler[1,2], Lauriane Kuhn[4] and Yaser Hashem[2,5] *

[1] Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR9002, Strasbourg 67000, France
[2] INSERM U1212 Acides nucléiques : Régulations Naturelle et Artificielle (ARNA), Institut Européen de Chimie et Biologie, Université de Bordeaux, Pessac 33607, France
[3] These authors contributed equally
[4] Proteomic Platform Strasbourg - Esplanade, Institut de Biologie Moléculaire et Cellulaire, CNRS, Université de Strasbourg, Strasbourg 67000, France.
[5] Lead contact
* Correspondence: yaser.hashem@u-bordeaux.fr

SUMMARY

In higher eukaryotes, the mRNA sequence in direct vicinity of the start codon, called the Kozak sequence (CRCCaugG, where R is a purine), is known to influence the rate of the initiation process. However, the molecular basis underlying its role remains poorly understood. Here, we present the cryo-electron microscopy (cryo-EM) structures of mammalian late-stage 48S initiation complexes (LS48S IC) in the presence of two different native mRNA sequences, β-globin and histone 4 (H4) at overall resolution of 3Å and 3.5Å, respectively. Our high-resolution structures unravel key interactions from the mRNA to eukaryotic initiation factors (eIF): 1A, 2, 3, 18S rRNA, and several 40S ribosomal proteins. In addition, we were able to study the structural role of ABCE1 in the formation of native 48S ICs. Our results reveal a comprehensive map of the ribosome/eIFs –mRNA and –tRNA interactions and suggest the impact of mRNA sequence on the structure of the LS48S IC.

INTRODUCTION

mRNA translation initiation in mammals is more complex than its bacterial counterpart. Indeed it includes more steps, more initiation factors and more regulation pathways. One can summarize the overall process in four steps, starting with pre-initiation. During pre-initiation, the ternary complex (TC) is formed by the binding of the heterotrimeric eukaryotic initiation factor 2 (eIF2) to one molecule of guanosine triphosphate (GTP) and the initiator methionylated tRNA ($tRNA_i^{Met}$). The TC then binds to the post-recycled ribosomal small subunit (SSU), also called 40S subunit. TC recruitment is partially mediated by eukaryotic initiation factors attached to the 40S, eIF1, eIF1A, and 13-subunit eIF3 complex. This leads to the formation of the 43S pre-initiation complex (PIC). The architecture of the 43S PIC has been investigated structurally at low to intermediate resolutions (Aylett et al., 2015; Erzberger et al., 2014; Hashem et al., 2013).

The second step consists on the recruitment of the 5' capped mRNA and leads to the formation of the 48S IC. This step is mediated by the cap-binding complex composed of eIF4F, eIF4A and eIF4B (Gross et al., 2003; Jackson et al., 2010; Marintchev et al., 2009; Rogers et al., 2001).

The third step is the scanning process for the start codon (AUG) in the 5' to 3' direction. This step was first investigated structurally in yeast (Hussain et al., 2014; Llacer et al., 2015) using *in vitro* reconstituted complexes. Upon start-codon recognition, the codon: anticodon duplex is formed between the mRNA and the $tRNA_i^{Met}$ aided by the eIF1A N-terminal tail (NTT) (Hinnebusch, 2011; Llácer et al., 2018; Lomakin and Steitz, 2013). The GTP is hydrolysed by eIF2γ, eIF1 dissociates from the P-site along with eIF1A C-terminal tail (Zhang et al., 2015) and the N-terminal domain (NTD) of eIF5 takes their place on the 40S (Llácer et al., 2018), before eIF5-NTD dissociates in turn at a stage that still remains to be elucidated. This results in the formation of the LS48S IC (that we describe in this work). The arrest of scanning and a cascade of structural rearrangements lead to the sequential dissociation of most eIFs upon the release of inorganic phosphate (Pi), generated from the GTP hydrolysis. eIF3 stays attached to the remaining complex probably through its peripheral subunits and leaves at a later stage during early elongation cycles (Beznosková et al., 2013, 2015). During all these steps the post-recycling factor ABCE1 can bind directly to the 40S and act as an anti-ribosomal subunits association factor (Heuer et al., 2017; Kiosze-Becker et al., 2016; Mancera-Martínez et al., 2017).

In the final fourth step, ABCE1 is replaced by the GTPase eIF5B on helix 14 of 18S rRNA, thus stimulating the joining of the 40S and 60S ribosomal subunits, forming an 80S complex (Fernandez et al., 2013) and eIF1A and eIF5B are released together (Fringer et al., 2007).

The sequences flanking the AUG start-codon region have been identified as crucial for start site selection by the IC (Kozak, 1986, 1987b, 1987a, 1989). The optimal sequence for translation initiation in eukaryotes was named after Marylin Kozak, who first defined the optimal sequence in vertebrates as CRCCaugG, where R stands for a purine (Kozak, 1984, 1989). In this motif, modification of certain positions have influence on translation efficiency, such as (-3) and (+4) (Kozak, 1984). As a result, a sequence can be dubbed "strong" or "weak" by

considering those positions. It was further shown that the substitution of A(-3) for pyrimidine, or mutations of the highly conserved G(+4), lead to a process known as “leaky scanning” with bypass of the first AUG and initiation of translation at the downstream start codon (Kozak, 1986, 1989; Lin et al., 1993). More recent studies observed a more extreme case of sequence-dependent translation initiation regulation, dubbed “cap-assisted” for certain cellular mRNAs, such as those encoding histone proteins and in particular histone 4 (H4) mRNA (Martin et al., 2011, 2016). Cap-assisted internal initiation of H4 mRNA implies a very minimalistic scanning mechanism, which is possible thanks to the presence of a tertiary structure on the mRNA at the channel entrance. This element assists in placing the start codon very close to the P-site almost immediately upon its recruitment through the cap-binding complex.

In spite of the tremendous recent advances in understanding this phase of translation, high-resolution structural studies of the initiation process have been conducted by *in vitro* reconstitution of the related complexes. This approach often requires biologically irrelevant molar ratios of the studied eIFs (Aylett et al., 2015; Erzberger et al., 2014; Hashem et al., 2013; Hussain et al., 2014; Llacer et al., 2015), thus limiting insight into more subtle regulatory pathways. Moreover, the structures of the mammalian (pre)ICs are still at intermediate resolutions approximating 6Å (Eliseev et al., 2018; des Georges et al., 2015; Simonetti et al., 2016). Finally, although the role of ABCE1 was experimentally demonstrated as a ribosomal subunit anti-association factor preventing premature binding of the 60S (Heuer et al., 2017), its impact on the 48S complex formation and conformation is still unclear.

Here, we present cryo-EM structures of LS48S IC formed after recognition of the start codon on two different native and abundant cellular mRNAs, β-globin and histone 4, presenting variants of the Kozak sequence. Both complexes were prepared and isolated in near-native conditions from rabbit reticulocyte lysate (Figure 1A). Although the initiation regulation may differ mechanistically between these two archetype mRNAs, our structures provide a high-resolution snapshot on the Kozak sequence-dependent variable interactions in the LS48S IC in mammals.

RESULTS

Overall structure of the mammalian 48S initiation complex

The complexes were prepared using a modified version of our approach (Simonetti et al., 2016) (see Methods) that consists on stalling the LS48S IC in rabbit reticulocytes lysate (RRL) using GMP-PNP (a non-hydrolysable analogue of GTP) on the two target cellular mRNAs: mouse histone H4 mRNA (suboptimal Kozak) and human β-globin mRNA (stronger Kozak) (Figure 1A). These mRNAs were transcribed and capped from BC007075 cDNA (β-globin) and X13235 cDNA (H4). The advantage of this approach is the ability to prepare ICs bound on different mRNAs of interest directly in nuclease-treated cell extract in the absence of the endogenous mRNAs, allowing for a study of regulatory aspects of the process in natural abundance levels of native eIFs at physiological molar ratios.

The composition of both complexes was investigated by mass spectrometry (Figure 1B). Our analysis reveals the incorporation in both complexes of all eIFs expected to be present after the start-codon recognition (eIF1A, eIF2α, eIF2β, eIF2γ, eIF3 complex, ABCE1). As expected, extremely poor numbers of peptides and spectra for eIF1 detected in either complex, corroborating that our complexes are at a late-stage after the start-codon recognition and eIF1 dissociation.

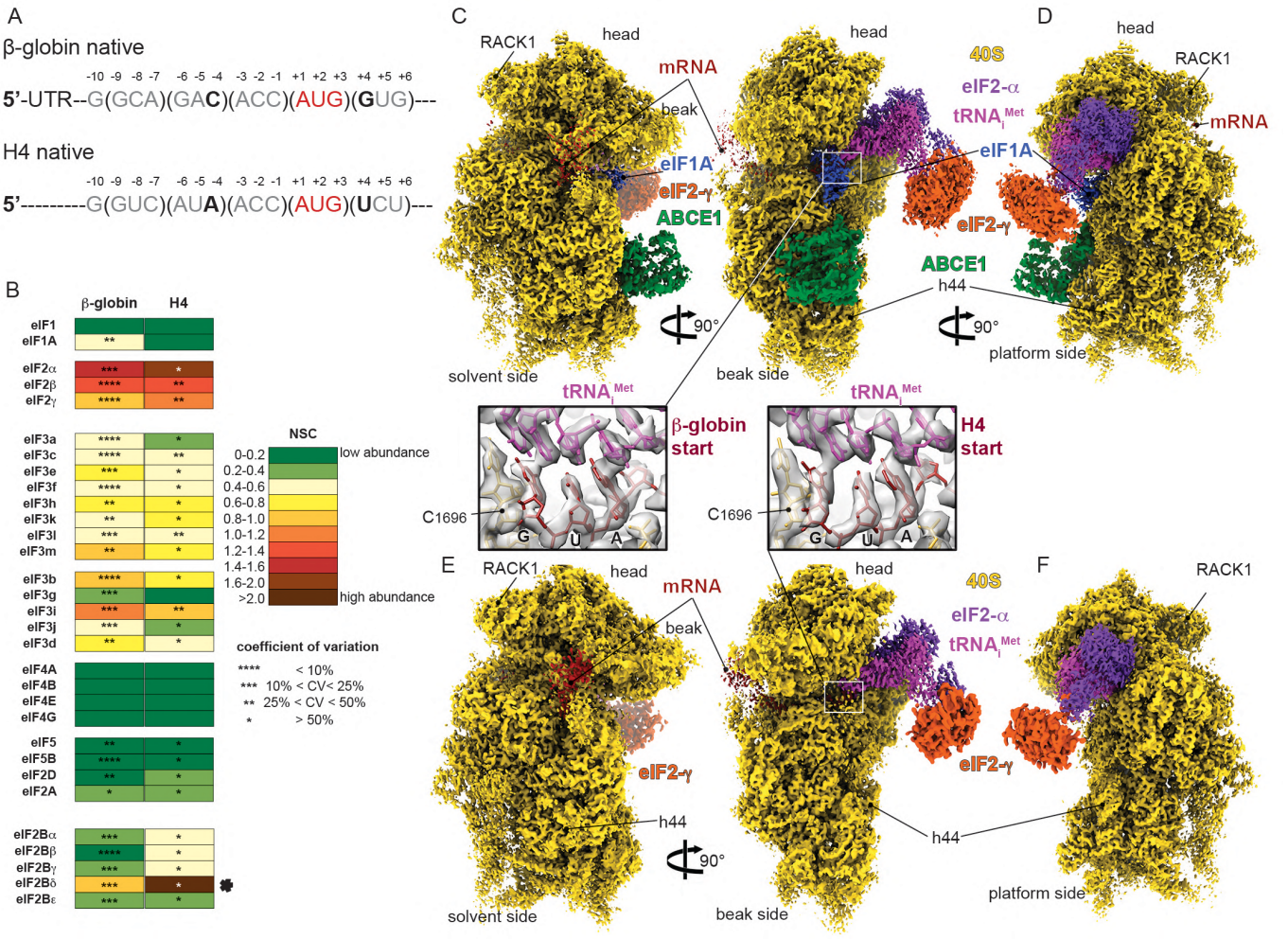


Figure 1. Overall structure of β-globin and H4 late-stage 48S initiation complexes.
(A) mRNA sequences used to form and purify the β-globin and the H4 ICs. Only the sequences near the AUG codon are represented and main differences in the Kozak sequence are indicated in bold. (B) Semi-quantitative mass spectrometry analysis of the eIFs in both ICs, indicating the abundance of each eIF based on the spectra count normalized. The two rounds of normalization were carried out using the total number of eIFs and estimated number of trypsin cleavage sites (see Methods). The normalized spectra counts (NSC) are presented as heat maps with cold colours indicating low abundance and warm colours indicating high abundance. The higher abundance of eIF2 proteins might be due to the excess of a free TC in the sample. The black star points out a high number of NSC for eIF2Bδ which is caused by the detection of three different isoforms of this protein. Small stars indicate the values of the coefficients of variation calculated for each NSC. In the analysis, the NSC for ABCE1 is not included, as it is a factor present also in other stages of translation than initiation. (C-D) Segmented cryo-EM reconstructions of the β-globin IC seen from (C) solvent, beak and (D) platform sides, respectively. The reconstruction shows 40S (in yellow), eIF2γ (in orange), eIF2α (in purple), $tRNA_i^{Met}$ (in magenta), mRNA (in red), eIF1A (in skyblue) and ABCE1 (in green). (E-F) Same as C-D but for the H4 IC. Boxed blowups represent the codon:anticodon duplexes in all shown reconstructions with their respective atomic models fitted in the corresponding electron densities.

In parallel, we have subjected our prepared complexes to structural analysis by cryo-EM. The structure of the β-globin LS48S IC (3.0Å, 29% of the total number of 40S particles, Figures 1C-1D and S1A-S1H) shows mRNA, 40S, eIF1A, TC and ABCE1. For the H4 mRNA 48S complex (3.5Å, 6.5% of the total number of particles, Figures 1E-F and S1P-W), the main reconstruction shows mRNA, 40S and the TC. We attribute the lower percentage of H4 LS48S IC formation to contamination with 60S subunits (~30%) (Figure S1W). Interestingly, both our cryo-EM structures and MS/MS analysis show that the H4 LS48S IC displays a significant reduction in the presence of eIF1A, leaving only residual density for its presence in the cryo-EM reconstruction (Figures 1E and S2A). Similar observation can be made for ABCE1 in the H4 LS48S IC. Our reconstructions show also another class of IC with eIF3 that is described later.

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

In both our reconstructions the codon:anticodon duplex is clearly formed, characterizing the cognate start-codon recognition (Figures 1C,E and 2A,B). AUG codons of both mRNAs face the $_{(34)}$CAU$_{(36)}$ of anticodon stem-loop (ASL) tRNA$_i^{Met}$, within hydrogen-bonding distances (~2.7Å). In the case of β-globin mRNA, the codon:anticodon interaction is stabilized further by the N-terminal tail (NTT) of eIF1A (Lys7 interacts with the ribose of G(+3) from mRNA, Figure 2D). The tail also interacts with the tRNA$_i^{Met}$ A(35) between Gly8 and Gly9. With few exceptions, this eIF1A NTT is highly conserved among eukaryotes (Figure 2D). Recent fluorescence anisotropy with yeast reconstituted PICs (Llácer et al., 2018) demonstrated that eIF1A binds with lesser affinity to a near-cognate start codon (UUG) compared to a cognate AUG. Along the same lines, only very residual density for eIF1A can be observed in the H4 LS48S IC structure (Figure S2A) (discussed below), which reflects its weaker binding affinity after the start-codon recognition at this late stage to the 48S complex.

C1696 of 18S rRNA is stacked on the C(34) base at the very tip of the tRNA$_i^{Met}$ ASL that is paired to G(+3) of both β-globin and H4 mRNA (Figure 2D, shown only for β-globin complex). This contact between C1696 and C(34) is also found in the yeast partial 48S pre-initiation complex (py48S IC) (Hussain et al., 2014) and it occurs even in the absence of any mRNA (des Georges et al., 2015). This stacking interaction may partly explain the difference in recruitment of initiator tRNA between bacteria and eukaryotes. In bacteria, the initiator tRNA is recruited directly at the P-site-accommodated start codon, whereas in eukaryotes, the tRNA$_i^{Met}$ is recruited at the pre-initiation stage of the complex before the attachment of mRNA into its channel. The tRNA$_i^{Met}$ ASL also interacts with the C-terminal tails of 40S ribosomal head protein uS19 (Figures 2E and S3B) through its Arg140 that contacts A(35).

We then compared the overall conformation of the 40S between both complexes and we observed that in the β-globin IC the head of the SSU is tilted downwards by ~2° and swivelled toward the solvent side by ~3° when compared to its counterpart in the H4 IC (Figure S2C). We attribute these subtle conformational changes to the dissociation of eIF1A in H4 LS48S IC, after the start-codon recognition, due to the loss of contacts between eIF1A and the 40S head.

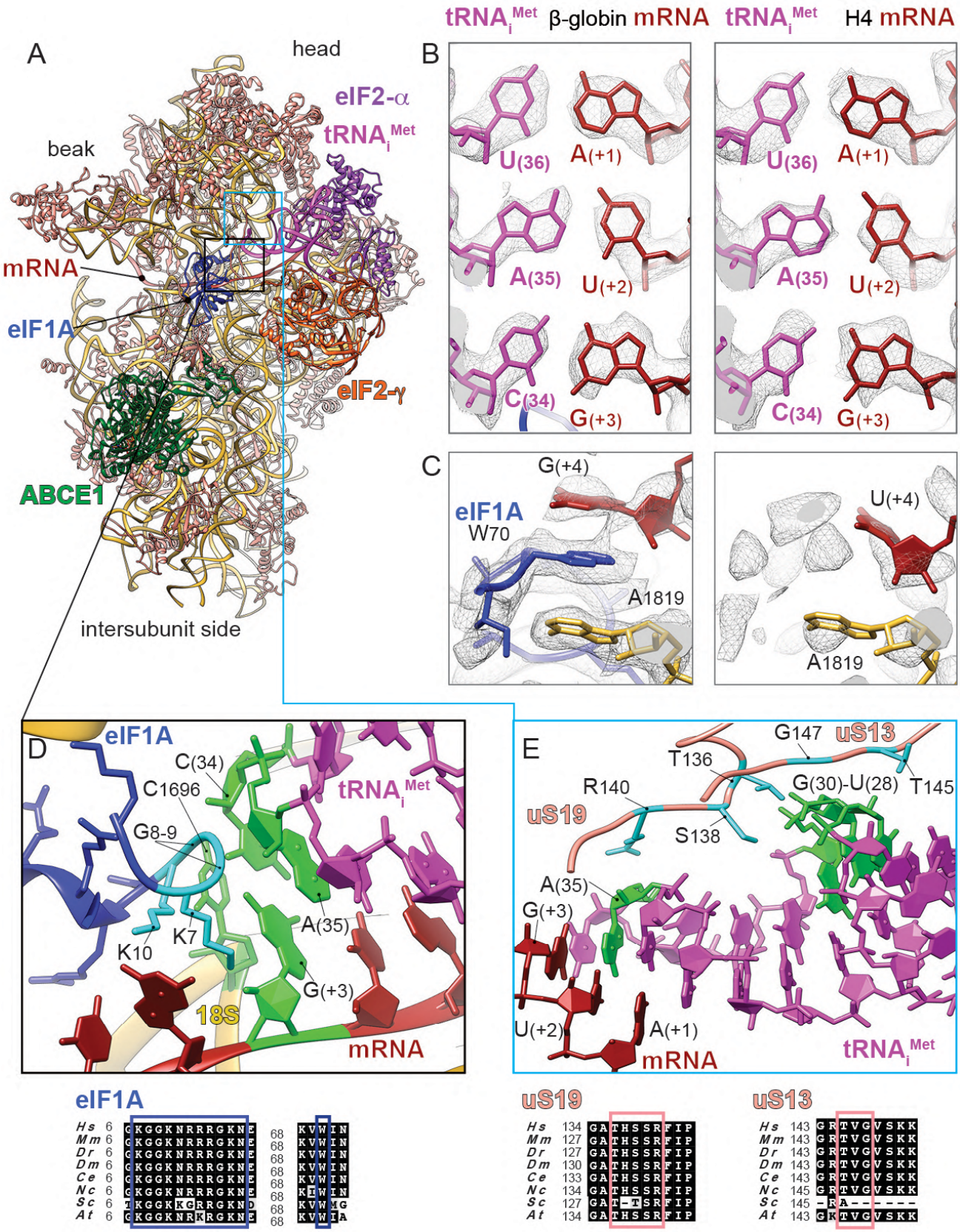


Figure 2. Key interactions surrounding the start-codon recognition sites in β-globin and H4 LS48S ICs.
(A) Ribbon representation of the atomic model of β-globin LS48S IC viewed from the intersubunit side. (B) Codon:anticodon base-pairing view in both mRNA complexes; left: β-globin, right: H4. (C) eIF1A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

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

The (+4) position, mainly occupied by a G in eukaryotic mRNAs, plays a pivotal role (Kozak, 1984, 1986). Our reconstructions demonstrate the structural importance of this position to both mRNAs. In the β-globin LS48S IC, the highly conserved Trp70 from eIF1A is trapped between the mRNA G(+4) position and the A1819 from h44 18S rRNA of the A-site by stacking interactions (Figure 2C). Interestingly, the interaction of the (+4) mRNA position with h44 was shown by cross-linking studies (Pisarev et al., 2006). Our β-globin LS48S IC structure also shows the proximity of uS19 C-terminal tail to (+4) mRNA position, which can also be corroborated by several cross-linking studies (Bulygin et al., 2005; Pisarev et al., 2008, 2006). In H4 mRNA a U is at position (+4), therefore the stacking interaction with eIF1A appears weaker than when a G is present. Moreover, nucleotides A1818 and A1819 have even more scant densities, indicating their undetermined conformations probably linked to this poor stacking (Figure 2C). Our reconstructions therefore suggest the structural importance of the (+4) position in the interaction with eIF1A.

Another crucial position in the Kozak consensus sequence is at (-3), often occupied by an adenine (Figure 1A). This nucleotide in both complexes shows several contacts with ribosomal proteins and initiation factors, including salt-bridge interaction between A(-3) base and a side chain of Arg55 from domain 1 (D1) of eIF2α (Figure 3A), which was reported previously in the py48S IC structure (Hussain et al., 2014; Llácer et al., 2018). However, in the yeast structure the A(-3) base is in the *syn* conformation and in both our mammalian ICs the adenine is in the *anti* conformation. Noteworthy, the near-cognate yeast mRNA present in the py48S IC structure (Llácer et al., 2018), contains adenines at the positions (-1) and (-2), which in principle could create an ideal stacking context for the A base in (-3), thus explaining this difference in conformation compared to our mRNAs where these positions are occupied by two cytosines. The (-3) position further interacts with the G957 nucleotide at the 40S platform (Figure 3A), highlighted in earlier studies (Demeshkina et al., 2000). In addition, cross-linking studies of reconstituted mammalian PIC previously demonstrated that eIF2α and uS7 interact with the (-3) nucleotide, and uS7 with the (-4) nucleotide (Pisarev et al., 2008, 2006). The interaction of uS7 through its β-hairpin was also suggested in the py48S IC structure, due to their proximity in space (Hussain et al., 2014; Llácer et al., 2018). However, in our structures, this interaction cannot be confirmed since the electron density at this specific region is

very disperse, probably because of the flexibility of this part of uS7 (see Discussion, Figure S5A) compared to its other parts.

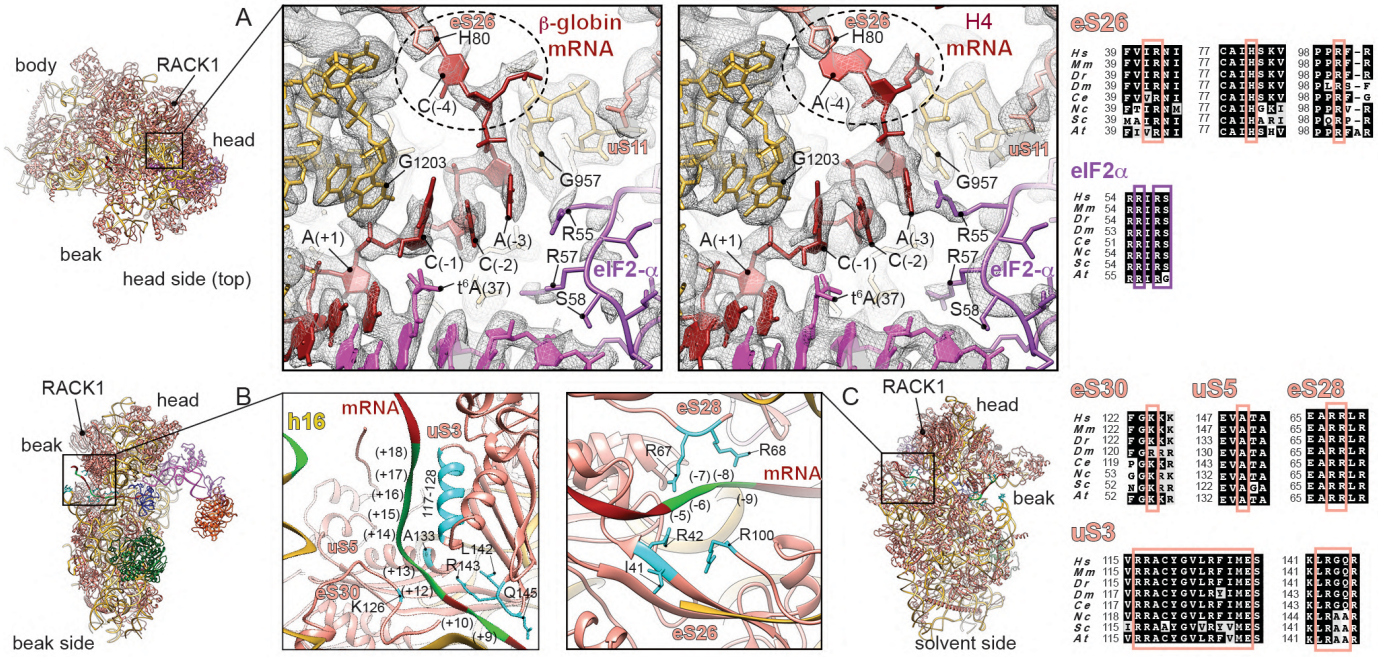


Figure 3. Kozak and beyond Kozak interaction networks in β-globin and H4 LS48S ICs.
(A) Close-up of the interactions of upstream start-codon nucleotides top-viewed from the head side in the β-globin (left panel) and H4 (right panel) ICs with ribosomal proteins eS26 and uS11, as well as eIF2α D1 domain, $tRNA_i$ and 18S rRNA. mRNA (-4) position contact with His80 of eS26 is highlighted in dashed line circle. The distances between atom N1 of His80 eS26 and amine groups of A(-4) (β-globin) and C(-4) (H4) are 3.7 Å and 3.2 Å, respectively. G1203 and G957 of 18S rRNA stacking and interaction with C(-1) and A(-3), respectively, of both mRNAs are shown. (B) mRNA entry channel seen from the beak side with close-up on the interactions with uS3, eS30 and h16 of 18S rRNA. (C) mRNA exit channel seen from the solvent side with close-up on the mRNA contacts with ribosomal proteins eS26 and eS28. (B) and (C) panels are shown on an example of β-globin LS48S IC. The nucleotides involved in the interactions are indicated in green and residues in cyan. Respective sequence alignments are shown in black boxes from eight representative eukaryotic species on the right of the figure panels.

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

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

eIF1A interaction with the β-globin mRNA sequence and the 18S rRNA

In addition to the above-mentioned contact with the start codon and G(+4), eIF1A can potentially establish several interactions at more distal positions in the β-globin mRNA sequence, closer to the mRNA entrance channel. Indeed, Arg12, Lys67 and Lys68 in eIF1A are in close proximity to C(+7), G(+6) and U(+5) (Figure 4A). eIF1A NTT also interacts with the 18S rRNA (Gly9 and Lys10 with C1696; and Lys16 and Asn17 with C1327) (Figures 2D and 4A). Other contacts involve the loops of the eIF1A OB domain with the 40S near the A-site (Figures 4B-4D): namely, Asn44 and Arg46 are in contact with A1817-A1819 and C1705 from h44 of 18S rRNA (Figure 4C); moreover Lys64 and Arg62 contact G604 and C605 of h18 18S rRNA (Figure 4D). In addition, Arg82, Tyr84 and Gln85 of eIF1A contact Glu58, Leu91 and Gly56 of ribosomal protein uS12 (Figure 4D); finally, Asp83 is in contact with Arg82 of eS30 (Figure 4D). Putting together, the above-mentioned interactions might depend on the mRNA sequence and perhaps they can have influence on the stability the cognate start-codon duplex with its anticodon by the NTT of eIF1A (residues Lys7, Gly8, Gly9 and K10).

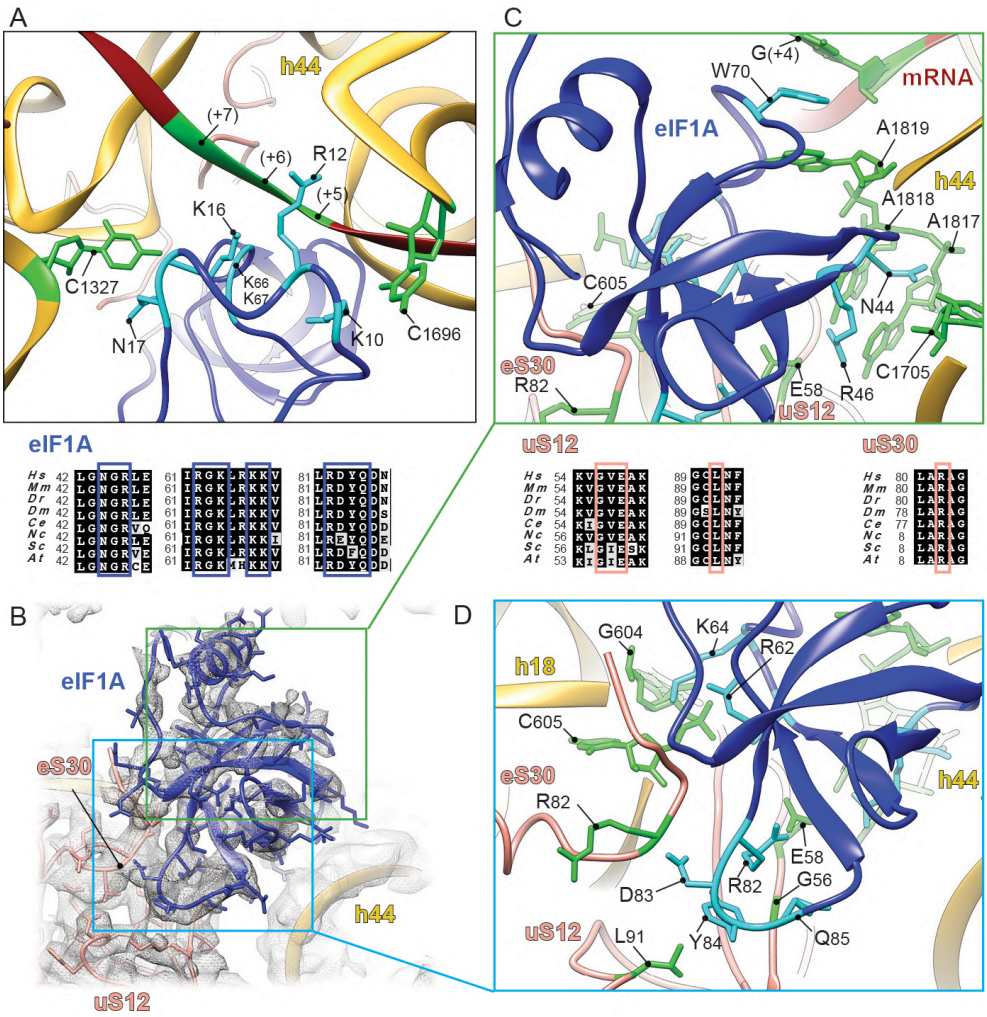


Figure 4. eIF1A interactions in the A-site of β-globin LS48S IC.
(A) eIF1A (in dark blue) N-terminal tail interactions with mRNA of downstream start-codon nucleotides and $tRNA_i$. The nucleotides involved in the interactions are indicated in green and residues in cyan. (B) eIF1A OB-domain interactions with mRNA and 40S. (C) Close-up on interactions of eIF1A (in dark blue) with h44 of 18S rRNA (nucleotides are coloured in green). (D) Zoom in on eIF1A (in dark blue) interactions with h18 of 18S rRNA (in gold) and ribosomal proteins uS12 and eS30 (in salmon). The nucleotides and residues of uS12 and eS30 involved in the interactions are indicated in green and eIF1A residues in cyan. Respective sequence alignments are shown in black boxes.

Noteworthy, eIF1A NTT was shown to interact with eIF5 (Luna et al., 2012, 2013), but because of its clear involvement in the start codon:anticodon duplex, we suggest that this eIF5 interaction occurs during the pre-initiation phase and very shortly after the recruitment of the mRNA.

mRNA interactions with the 48S beyond the Kozak sequence

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

At the entrance of the mRNA channel during this late stage of the initiation process, the mRNA extensively interacts with conserved residues of the 40S ribosomal proteins uS3 and eS30, and with rRNA h16 in positions spaning from +10 to ~ +20 through ionic and hydrophobic interactions (Figure 3B). For instance, the conserved Arg117 of the head protein uS3 contacts the mRNA at the channel entrance. This residue was recently indicated as important for stabilizing the P_{IN} closed state of the 48S in yeast IC (Llácer et al., 2018) and for the initiation accuracy in the presence of suboptimal Kozak sequence by *in vivo* assays in yeast (Dong et al., 2017). The contribution of this charged residue of uS3 contacting the mRNA is partially corroborated by cross-links in a previous study (Pisarev et al., 2008). More globally, charged amino acid residues from uS3 helix α (residues 117-128) are in close proximity to nucleotides from positions (+14) to (+18) (Figure 3B). Moreover, residues from a β-hairpin (residues 142-146) can potentially contact bases of the nucleotides C(+9) and C(+10) of the mRNA, forming hydrophobic and salt-bridge interactions. For ribosomal protein eS30, Lys126 is in close distance to the bases of G(+12) and A(+13). The proximity of A(+13) of mRNA to Ala133 of uS5 can also be noted (Figure 3B).

On the other side at the mRNA exit channel, we can observe the exit of both β-globin and H4 mRNAs from their respective 48S ICs (Figure 3C). The 5' untranslated region (5'UTR) for β-globin mRNA is substantially longer than for H4 (50 nt and 9 nt, respectively) (Figure 1A, see Methods). We compared the mRNA exit channels of both complexes below the ribosomal head protein RACK1, which unambiguously shows the expected larger 5' UTR for β-globin LS48S IC compared to H4 (Figure S2B). We were able to spot several possible contacts of ribosomal proteins at the exit site with mRNA nucleotides in both LS48S IC, including eS28 (Arg66 with A(-5), and Arg67 with A(-7)) as well as eS26 (Ile41 with C(-8), Arg42 with A(-9) and Arg100 with both these nucleotides) (Figure 3C), in agreement with previous cross-linking results (Pisarev et al., 2008, 2006).

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

The overall accommodation of the mammalian ASL resembles its yeast counterpart found in the P_{IN} state (Hussain et al., 2014; Llácer et al., 2018). The 48S IC-tRNA$_i^{Met}$ interaction network is summarized in the Figure S3B. In both ICs, we can observe a density attached to A(37), in which we can model the threonylcarbamyol group forming a t^6A modification (Figures 3A and 5A). This modification mediates the binding of t^6A(37) to the 2'OH of C(-1) in the mRNA, and therefore can further stabilize the start-codon recognition. It is tempting to suggest that C(-1):t^6A(37) interaction is required for efficient translation in mammals. This mRNA C(-1) position is conserved in higher eukaryotes, as revealed by quantitative sequence analysis (Grzegorski et al., 2014), and forms part of the Kozak sequence. Interestingly, the electron density for this modification is even stronger in the H4 mRNA complex than in β-globin, even at a lower resolution (3.5Å). We therefore suspect that this interaction could be more important in the case of suboptimal Kozak sequences, where this modification could compensate for the loss of some interactions with the 48S IC, compared to a stronger Kozak mRNA. The same interaction in the case of A at (-1) position is not excluded, however its nature and conformation will be different.

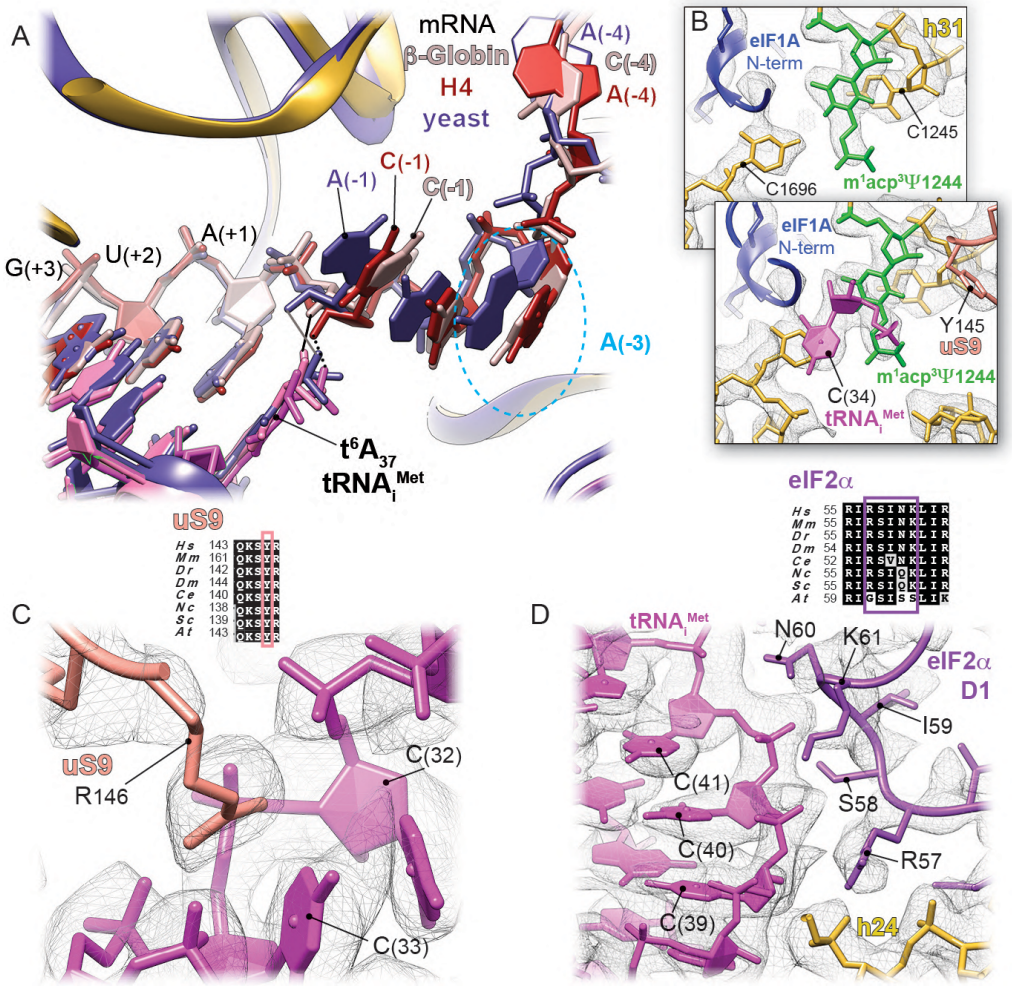


Figure 5. Initiator tRNA anticodon stem-loop (ASL) interactions with the LS48S IC. (A) Comparison of the tRNA$_i$ modified t^6A(37) interaction with mRNA (-1) position in mammalian β-globin (in light pink) and H4 (in red) and yeast (Llácer et al., 2018) (dark purple) initiation complexes. The contact with (-1) mRNA position is labelled by black solid line, black dots and grey solid line for β-globin, H4 and optimal yeast mRNA sequence, respectively. In dashed skyblue circle, comparison of the conformation of highly conserved A(-3) mRNA position: *anti* in mammalian IC and *syn* in yeast. (B) Interaction of the modified $m^1acp^3\Psi$1244 of the 18S rRNA (coloured in green, overlapping bottom panel) with the C(34) of the tRNA$_i$ and ribosomal protein uS9 (overlapping top panel). (C) Close-up on the interactions of C(32) and C(33) with Arg146 of uS9. (D) Close-up on the interaction of the ASL cytosines from the conserved G-C base-pairs with eIF2α domain D1. Respective sequence alignments are shown in black boxes and interacting residues in coloured frames.

Despite the universal presence of the t^6A hypermodified base in all organisms, and a crucial role in translation efficiency (Pollo-Oliveira and De Crécy-Lagard, 2019; Thiaville et al., 2016), it has only recently been shown that the modification directly contributes to the AUG recognition accuracy. In the py48S-eIF5 IC structure at a resolution of 3.5Å (Llácer et al., 2018), t^6A of the initiator $tRNA^{Met}$ was suggested to enhance codon:anticodon base-pairing by interacting with A(-1) and by a stacking on the downstream base-pair involving A(+1). Thanks to our mammalian LS48S IC at 3.0Å, we can clearly observe the threonylcarbamyol group in a different conformation, placing the carboxyl group within hydrogen-bonding distance (2.7Å) from the 2'OH group of C(-1) (Figure 5A). However, this modification does not appear to stack over the downstream base-pair as previously suggested. Noteworthy, in yeast there is a preference for an A at position (-1) (Dvir et al., 2013; Kozak, 1986), while it is a C(-1) in mammals.

Furthermore, C(34), that is a part of the anticodon, is stabilized by the modified U1244 (U1248 in human) of helix31 of rRNA ($m^1acp^3\Psi$, 1-methyl-3-(3-amino-3-carboxypropyl) pseudouridine) (Maden, 1990; Taoka et al., 2018) (Figure 5B), which was previously reported in the py48S-eIF5 (modified U1119) (Llácer et al., 2018). In addition, the neighbouring nucleotides, C(32) and C(33), are in contact with C-terminal arginine Arg146 of ribosomal head protein uS9 (Figure 5C).

Aside from its role in the codon:anticodon stabilization, uS19 together with uS13 were found to contact other parts of the $tRNA_i^{Met}$ through their highly conserved C-terminal tails (Figures 2E and S3B). Thr136 side chain of uS19 interacts with the guanine backbone of three conserved G–C base pairs in ASL that are crucial for stabilization of the initiation complex in eukaryotes (Dong et al., 2008). Residues Thr145 and Gly147 of uS13 are in contact with the phosphate groups of U28 and G29 (Figure 2E). Consistent with previous reports (Hussain et al., 2014; Llácer et al., 2018), several residues from domain D1 of eIF2α (Arg57, Ser58, Asn60 and Lys61) interact with the cytosines backbones of three conserved G–C base pairs of the ASL (phosphate groups of C39-41) (Figures 5D and S3B).

ABCE1 binding to the initiation complex is NTP-dependent

ABCE1 (named Rli1 in yeast) is a conserved NTP-binding cassette ABC-type multi-domain protein that plays a role in translation initiation as well as translation termination and ribosome recycling (Becker et al., 2012; Heuer et al., 2014; Khoshnevis et al., 2010; Pisarev et al., 2010; Shoemaker and Green, 2011; Young et al., 2015). It contains two nucleotide-binding domains (NBDs), where the two NTP molecules bind. Its N-terminal NBD contains two iron-sulphur clusters $[4Fe\text{-}4S]^{2+}$ (Barthelme et al., 2007, 2011; Karcher et al., 2008). In our β-globin LS48S IC structure, the NTP-binding cassette of ABCE1 displays lower local resolution (between 3.5 and 5Å, Figures 6A and S1D) compared to the average resolution, likely due to the flexibility of the NBDs. In H4 LS48S IC structure we observe only a residual density of ABCE1 (Figure S1L), which can be caused by a slightly different conformation of h44 18S rRNA in the absence of eIF1A.

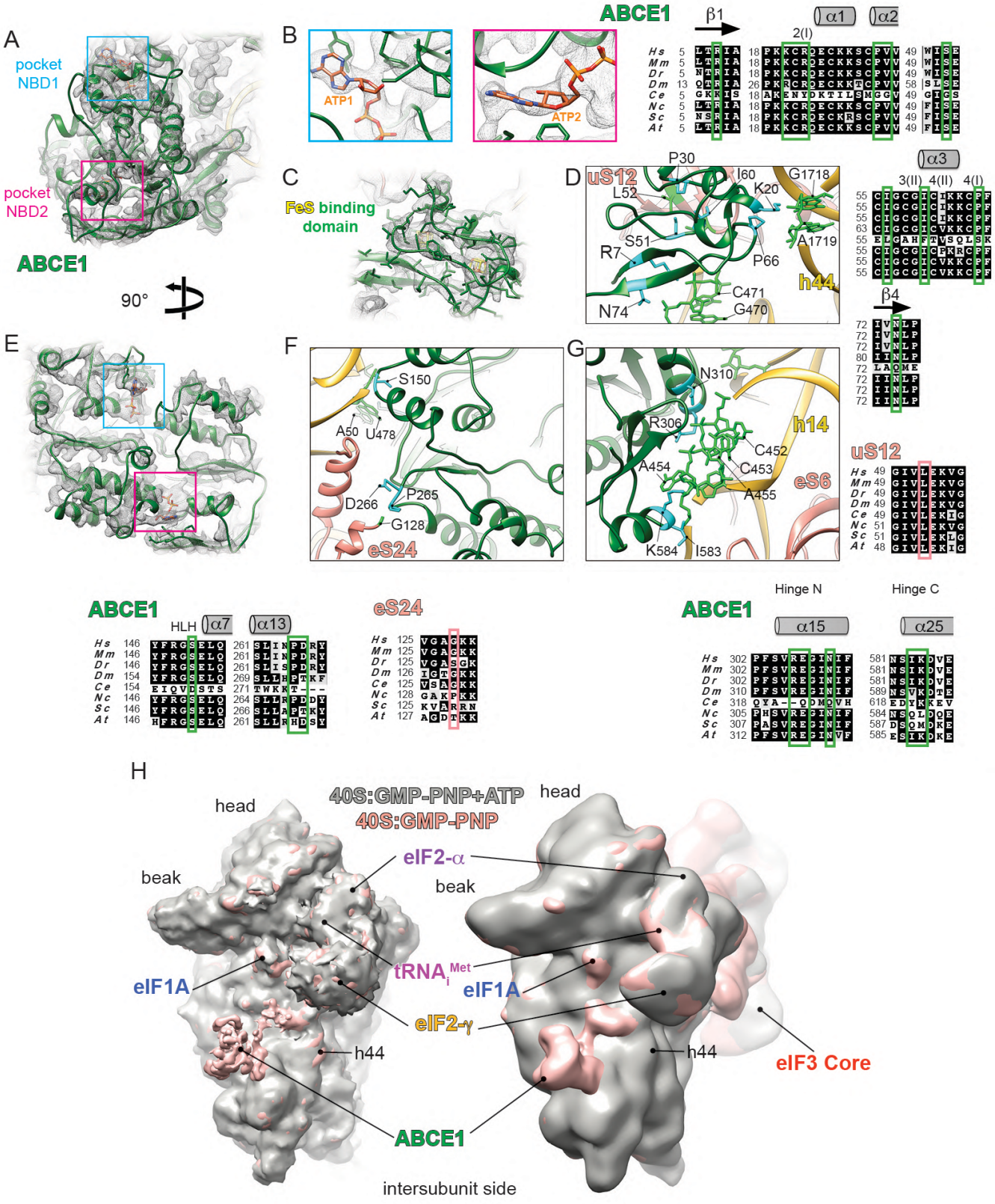


Figure 6. Atomic model of ABCE1 in the LS48S IC.
(A) and (E) Ribbon representations of ABCE1 (in green) in its electron density with NTP-binding pockets framed in blue (NBD1) and pink (NBD2) seen from different side views. (B) Blowups on the NTP pockets, NBD1 (left blue) and NBD2 (right pink). Although in the purification conditions GMP-PNP was used, ATP molecules were modelled in the electron densities obtained. (C) Mixed ribbon and stick representation of Fe-S binding domain atomic model fitted into its electron density. (D) Close-up of Fe-S binding domain interactions with 40S ribosomal protein uS12 and h44 of 18S rRNA. (F) Close-up on NBD1 interactions with the 40S. (G) Close-up on NBD2 interactions with the 40S. The nucleotides involved in the interactions are indicated in light green and protein residues in cyan. Respective sequence alignments are shown in black boxes. (H) Comparison between the β-globin (pink surface) and β-globin•GMP-PNP+ATP (grey surface) LS48S ICs reconstructions without (left superimposition) and with (right superimposition) eIF3. The panel shows that the addition of ATP triggers the dissociation of ABCE1, probably after its hydrolysis. No further conformational changes between both complexes, with and without ATP, can be detected.

The iron-sulphur (Fe-S) binding domain presents a higher local resolution than the NBDs (between 3 and 4Å, Figures 6B and S1D). Therefore, we can clearly distinguish the Fe-S clusters in the cryo-EM density map as well as the presence of two bound nucleotides that probably represent GMP-PNP that was used to stall the initiation complexes by blocking eIF2γ (Figures 6A-6C).

NBD1 contacts five nucleotides in the 18S rRNA helix14 (U478, A455, A454, C453 and C452) through residues located in a helix-loop-helix motif (Ser150) and in the hinge-N (Arg306 and Asn310) (Figures 6F-6G and S4A). NBD2 also makes contacts to nucleotides A455, A454, C453 through the hinge-C residues Lys584 and Ile583, conserved in higher eukaryotes (Figures 6G and S4A). Residues Arg566 and Arg567 from NBD2 are in close proximity to the 18S rRNA and can potentially be involved in the interaction with the latter, as suggested by their conservation (Figure S4A). Moreover, NBD1 residues Pro265 and Asp266 interact with residues from the C-terminal helix of ribosomal protein eS24 such as Gly128 (Figure 6F).

As for the Fe-S binding domain, it interacts with the 18S rRNA (Arg7 from strand β_1 to C471, Lys20 to G1718, Pro66 to A1719 and Asn74 to G470) (Figure 6D). In addition, ABCE1 interacts with ribosomal protein uS12 residues Ile50, Leu52 and Ile75 via a hydrophobic pocket formed by helices $\alpha_{2,}$ α_3 and cluster (II) (residues Pro30, Val31, Ile56, Ile60) (Figure 6D). This is consistent with previous structural and crosslinking studies in yeast and archeal complexes (Heuer et al., 2017; Kiosze-Becker et al., 2016; Nürenberg-Goloub et al., 2020).

In order to investigate the effect of ABCE1 binding on the structure and composition of the LS48S IC, we purified β-globin complexes using our choice strategy (Simonetti et al. 2016 and this work) supplemented with 10 mM of ATP, thus taking advantage of the ability of ABCE1 to hydrolyse ATP, in contrast to the obligate GTPase eIF2. The β-globin-ATP 48S IC was then analysed by cryo-EM and yielded two main LS48S IC reconstructions at ~14Å and ~10Å, that differ in the presence and absence of eIF3, respectively (Figures 6H and S1X,Y). The addition of ATP most likely causes the replacement of the GMP-PNP molecules in the NBD pockets by ATP molecules that was then hydrolysed by ABCE1. Our structures clearly reveal the dissociation of ABCE1 as a result of ATP addition, consistent with recent structural and biophysical studies (Gouridis et al., 2019; Heuer et al., 2017; Kiosze-Becker et al., 2016). Aside of the absence of ABCE1, the global structure of the β-globin-ATP LS48S IC is identical to its higher-resolution counterpart without ATP (Figure 6H), thus very likely excluding a direct active role of ABCE1

in the assembly of the initiation complex. It is reasonable to assume that in the cell ABCE1 undergoes on/off cycles to the IC in an ATP-dependent manner, as we have previously suggested (Mancera-Martínez et al., 2017). However, these results do not contradict the demonstrated function for ABCE1 as an anti-ribosomal-subunit association factor.

eIF3 in the late-stage 48S initiation complex

Nearly 15% (~5% of the total particles count) of the particles corresponding to the β-globin LS48S IC structures contain a density for eIF3 at the solvent side. After extensive particle sorting and refinement, a reconstruction of the β-globin LS48S IC showing eIF3 was obtained at a resolution of ~3.6Å (Figures S1I-S1O), thus allowing verification of the recognition of the start codon for this specific class (Figures 7A-7B). eIF3 a and c subunits, i.e. those that bind directly to the 40S, are mostly resolved at a resolution ranging from 3.5 to 4.5Å, enabling a model of the exact residues in interaction with the 40S subunit (Figures 7C-7G). We find that eIF3a in β-globin LS48S IC shows several contacts with 40S ribosomal body proteins, with eIF3a residues Asn10, Lys13, Arg14 and Phe18 interacting with eS1 residues Asp77, Asn76, Asp191 and Pro190, respectively (Figures 7F-7G). In addition, eIF3c residues Asn388, Arg340, Asn384, Gly341, Lys343 and Arg450 contact the 40S ribosomal protein eS27 via residues Glu75, Thr61, Gln65 and Cys59 (Figure 7E). Moreover, residues Lys342, Lys343, Thr391 and Tyr392 interact with nucleotides G925, C1112 and U1116, two latter being a part of the apical loop of expansion segment 7 (ES^7).

The eIF3d subunit structure is at lower resolution as compared to the eIF3 octamer core. Nevertheless, secondary structure elements can clearly be depicted when filtered to a lower resolution (6Å), which enabled the fitting of its partial crystal structure in our density (Lee et al., 2016). The modelled eIF3d displays contacts with several ribosomal head proteins: helix α_{12} contacts the N-terminal loop of uS7; the loop between β_9 and β_{10} contacts RACK1 (loop located between strands β_{6D} and β_{7A}); and β-strand loops and the “RNA gate” insertion contacts eS28 (strand β_3, loops between helices α_8 and α_9).

In spite of the low local-resolution of this particular subunit, our structure provides clues to the demonstrated interactions of eIF3d and eIF3a with the mRNA 5’UTR (Figure 7D). Thus, shown by the residual electron density traces, the 5’UTR of numerous mRNAs such as β-globin can possibly interact with different parts of eIF3d: N-terminal loop and strand β_2 (residues S166 to E172) and a loop between β_{11} and β_{12} (residues Asn513 and Lys514) (Figure 7D). These residues of eIF3d are better conserved in higher eukaryotes, indicating a possible species-specific regulation. An interaction with 5’ RNA terminus recognition motif was also previously reported (Lee et al., 2016). More insight into the interaction patterns of eIF3d with different mRNA 5’UTRs in the context of translation initiation will be an important goal in future studies.

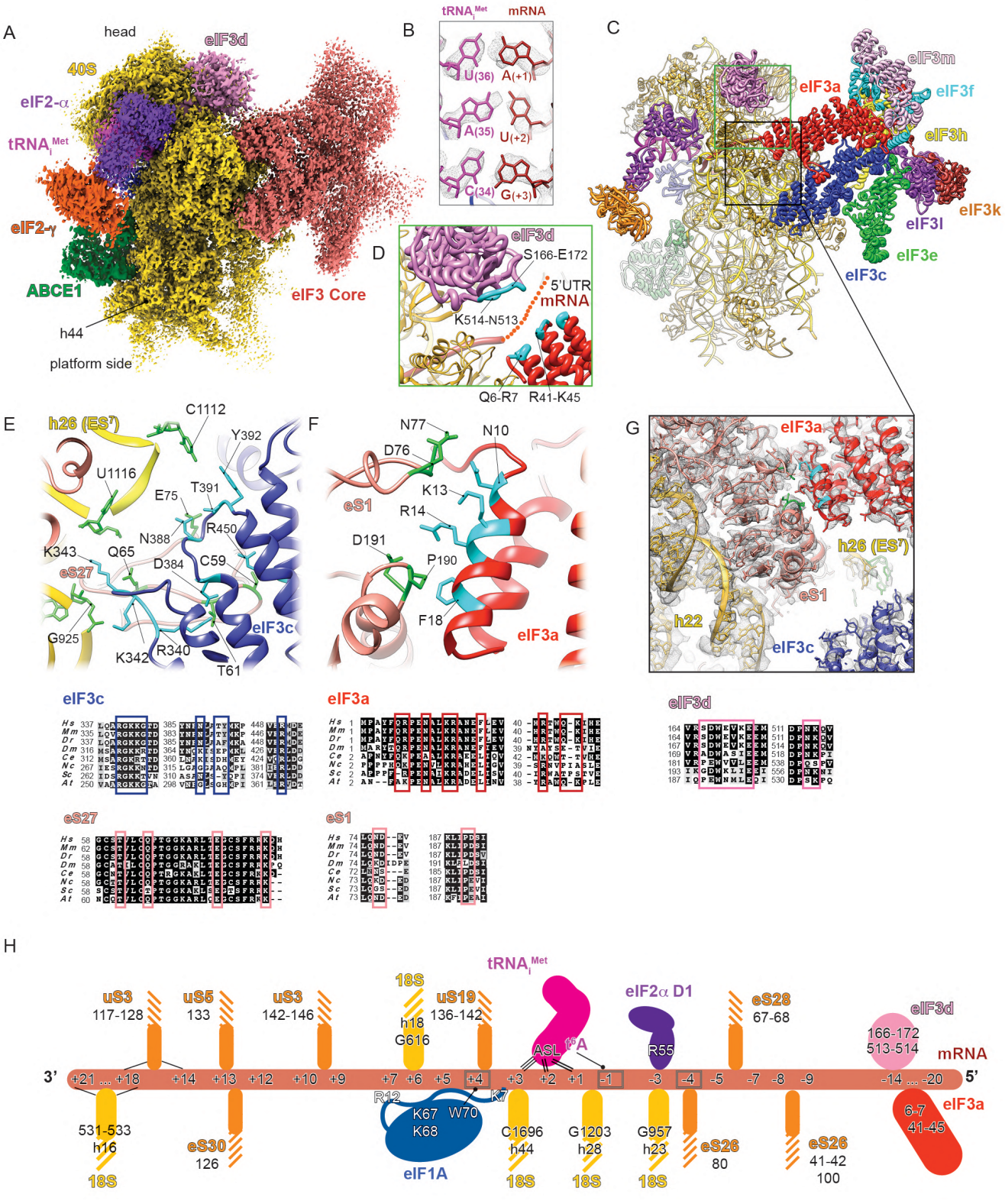


Figure 7. Cryo-EM reconstruction of the eIF3-containing class of the β-globin LS48S IC.

(A) Segmented map showing electron density of the eIF3 core (in rose) attached to the 48S viewed from the platform side. (B) Codon:anticodon base-pairing view in LS48S-eIF3 ICs (identical to both β-globin and H4). (C) Ribbon representation of the atomic model of LS48S-eIF3 IC seen from the platform side. (D) Blowup on the mRNA channel exit, seen from the platform side. mRNA 5'UTR cannot be modelled because of the low local-resolution of the cryo-EM reconstruction in this region, therefore we propose an extrapolation of the mRNA 5'UTR trajectory showing possible interactions with eIF3 d and a subunits (residues coloured in cyan). (E) Close-up on the eIF3c (in navy blue) interaction with 40S: h26 (ES^7) of 18S rRNA and eS27. (F) Close-up on the eIF3a (in coral) interaction with eS1. (G) Mixed ribbons and sticks representation of eIF3 core interactions with 40S. The nucleotides involved in the interactions are indicated in chartreuse and protein residues in cyan. Respective sequence alignments are shown in black boxes with interacting residues highlighted by coloured frames. (H) Summary of the mRNA interactions with mammalian LS48S IC. The ribosomal proteins are coloured in orange and 18S rRNA elements in yellow. The mRNA contacts critical for recognition of optimal or suboptimal Kozak context are highlighted by grey frames.

Finally, the mRNA 5'UTR of β-globin can also interact with eIF3a (residues Gln6, Arg7, Arg41, Gln44 and Lys45) (Figure 7D). Similarly to eIF3d, these residues are mainly conserved among higher eukaryotes (Figure 7G). It was previously shown that the eIF3a–PCI domain (a domain with a common fold for proteasome, COP9, initiation factor 3) is critical for stabilizing mRNA binding at the exit channel (Aitken et al., 2016). However, because of the low local resolution of the β-globin mRNA 5'UTR in our reconstruction, we do not exclude other patterns of interaction. H4 LS48S IC also shows the residual presence of eIF3 core, however particle sorting reveals a reconstruction containing eIF3 (Figure S1W) at only intermediate resolution because of the low number of particles.

DISCUSSION

Our cryo-EM structures reveal in detail the accommodation of two native mRNA sequences encoding either β-globin or histone 4 in the context of the late-stage mammalian IC. In the presented IC structures, we did not identify any density corresponding to eIF1. Combined with codon:anticodon complex formation and several conformational changes characteristic of the stage after the start-codon recognition, we have dubbed our complexes "late-stage 48S IC (LS48 IC)". Initiation on β-globin and H4 mRNA may undergo different regulatory processes as previously reported (Martin et al., 2011, 2016), however in our structures we only analyse the mRNA nucleotide interactions of the Kozak sequences (Figure 7H and Table S2) without dwelling on the exact regulation mechanism that may be in part influenced by the different interaction patterns that we observe. We therefore believe that our two archetype mRNA sequences are representative of native cellular mRNAs incorporating different Kozak sequences, as their observed interactions are purely the result of sequence differences, and unrelated to specific regulatory pathways.

In the Kozak sequence, the position (+4) appears to play a role in both mRNAs, where it is a G in β-globin and U in H4. At this position, the crucial interaction with Trp70 of eIF1A appears to be weaker in the case of H4 IC when compared to β-globin IC, as indicated by our MS/MS normalized spectral counts and the cryo-EM reconstructions. We suggest that the poor abundance of eIF1A in the H4 LS48S IC cryo-EM reconstruction (Figures 1, S1L and S2A) might not suggest a negligible role for this initiation factor in suboptimal Kozak consensus mRNAs during scanning. Rather, it simply shows its weaker interaction in the complex after start-codon

recognition, supported by our semi-quantitative MS/MS analysis. As expected, the biggest decrease in spectral counts was observed for eIF1A in the H4 LS48S IC compared to its β-globin counterpart (Figure 1B). The weak above-mentioned stacking interaction with mRNA purine at (+4) position is likely to be the reason of the affinity drop and subsequent weaker interaction between eIF1A and the 48S in the case of H4, after scanning and the recognition of the start codon. This observation is consistent with the suggested special translation initiation mechanism for H4 (Martin et al., 2011, 2016). It was suggested that H4 mRNA undergoes an unconventional "tethering mechanism", where the ribosomes are tethered directly on the start codon without scanning (Martin et al., 2011). This particular mechanism was proposed to occur thanks to the presence of two secondary structure elements present downstream of the start codon in H4 mRNA (position +19), contacting h16 of rRNA (Martin et al., 2016). In this region, our H4 LS48S IC structure shows disperse densities that cannot be interpreted. We believe that eIF1A is present in the H4 complex during pre-initiation and the short scanning process, and only after the recognition of the start codon the affinity can be affected by the mRNA Kozak context.

Regarding the N-terminal tail of eIF1A, it is present in the A-site starting from the scanning process, as shown by our structure and also yeast 43S PIC and closed-48S ICs (Llacer et al., 2015; Llácer et al., 2018). It is important to emphasize the binding of eIF1A on the H4 IC at a certain stage, because we can observe residual electron density for this initiation factor in the H4 LS48S IC (Figures S1L and S2A). This observation tends to validate the unconventional very short scanning mechanism proposed for H4 mRNA (Martin et al., 2011), yielding in the faster accommodation of the start codon as compared to β-globin mRNA. Nevertheless, in the early initiation steps eIF1A may interact with eIF5, as demonstrated by biophysical studies of *in vitro* purified proteins (Luna et al., 2013).

The importance of positions (+4) and (-3) of mRNA has been pointed out in previous studies (reviewed in (Kozak, 1989)), however the significant involvement of position (-4) in mammals was not highlighted. Similarly to py48S-eIF5 IC (Llácer et al., 2018), our structures show that when this position is A, it can be stabilized by residue His80 of eS26. This residue is highly conserved in eukaryotes (Figure 3A), therefore showing a universal mode of interaction. The role of eS26 (and eS28) in the accommodation of the 5' UTR was also highlighted by chemical cross-linking studies performed with 80S ribosome assembled on H4 mRNA (Martin et al., 2016). Remarkably, in the case of the yeast IC, it was shown that the interaction between the A nucleotide at the position (-4) and His80 of eS26 does occur through stacking (Llácer et al., 2018), in contrast to our study where we can show C and mainly A (-4) stacked below the His80 of eS26. This variation is probably due to the different sequence between our mRNA and those of yeast, where positions -1 to -3 are occupied by A, that favours more stacking between the nucleotides bases and consequently more twist (Figure 5A). Indeed, the kink in E/P site in py48S-eIF5 IC (Llácer et al., 2018) is sharper than in the case of mammalian 48S IC (Figure S3A). Interestingly, a previous cross-linking study in the context of the human 80S ribosome highlighted the eS26 binding to G and U nucleotides in the region (-4) to (-9) of the mRNA (Graifer et al., 2004). In a more recent biochemical study in yeast, it was shown that the mRNAs bound to the ribosomes depleted in eS26 (RpS26 in Saccharomyces) translate poorly when compared to

those enriched in eS26 (Ferretti et al., 2017). It was also reported in the same study that RpS26 is necessary for preferential translation of mRNAs with A at (-4) position and not G, showing that the interaction is very specific and not simply purine/pyrimidine dependent. Therefore, the His80 eS26 recognition is likely optimal for mRNA sequences containing A at this position. In yeast, however, the nucleotide context surrounding the start codon is less critical but shares with mammals the importance of the (-3) position (Cavener and Ray, 1991; Kozak, 1986). Further studies on translation of mRNAs containing mutations at these positions will help unveiling the mechanism of scanning in mammals and will shed light on the leaky scanning mechanism.

The tails of uS13 and uS19 have been previously shown to make direct interactions with the ASL of the peptidyl-site-exit-site (P/E) tRNA in presence of elongation factor G-ribosome complex in a pre-translocation state in prokaryotes (Zhou et al., 2013). To the best of our knowledge, these proteins have not yet been reported to be particularly involved in the initiation process in eukaryotes. Nevertheless, our structures are supported by earlier crosslinking studies of human 80S ribosome showing that the tail of uS19 is located closer to the decoding site than that of prokaryotic S19 (Graifer et al., 2004). In the case of *S. cerevisiae* uS13 and uS19, the C-terminal parts are not conserved, compared to human protein homologues (Figure 2E) and they have never been observed to interact with the $tRNA_i^{Met}$ (Hussain et al., 2014; Llácer et al., 2018). Other fungi, such as *N. crassa,* possess very similar sequences to mammalian counterparts and probably would demonstrate similar interactions to $tRNA_i^{Met}$ as shown by our structures. Moreover, and in contrast to yeast, *N. crassa* possesses a mammalian-like eIF3. This is in line with the recent genome-wide mapping of Kozak impact in the fungal kingdom, showing the particularity of start-codon sequence context in *S. cerevisiae* compared to other fungi (Wallace et al., 2019).

The ribosomal protein uS7 is located close in space to the position (-3) of mRNA, but we cannot confirm its interaction with this nucleotide due to the lack of the density for this part of the protein. Noteworthy, this is the only flexible region in this protein structure (Figure S5A), containing the crucial β-hairpin in the case of bacterial and yeast initiation (Visweswaraiah and Hinnebusch, 2017; Wimberly et al., 1997). This region in yeast contains the glycine-stretch GGGG (residues 150-153), whereas in human it is GRAG (residues 129-132). Genetic experiments on single-point mutants of this β-hairpin demonstrated almost unchanged phenotype for human-like G151R and G152A mutations (Visweswaraiah et al., 2015), but the G151S mutation was lethal. More recent work showed by using genetic and biochemical approaches that uS7 modulates start-codon recognition by interacting with eIF2α domains in yeast (Visweswaraiah and Hinnebusch, 2017). The residues implicated in the described interactions are highly conserved and are also present in mammals (Figure S5B). Therefore, we speculate that the effect of the studied substitutions of uS7 might result in similar phenotypes in mammals.

In comparison to the py48S-eIF5N structure (Llácer et al., 2018), the electron density of eIF5-NTD was not observed in our complexes, although mass spectrometry analysis revealed some residual presence of eIF5 only in the β-globin LS48S IC. This may suggest that the presented LS48S ICs were trapped between eIF1 dissociation and eIF5-NTD binding (which would represent the structure corresponding to the intermediate state between p48S-closed and p48S-5N, according to Llácer et al., 2018). Another reason might be the weaker affinity of eIF5 to the

initiation complex at this stage, as both of our complexes are purified directly from RRL without supplementation of any factors, which is overcome when the complex is assembled *in vitro* with higher molar ratios.

After the dissociation of the TC (GTP-driven binding), eIF5B is recruited to the IC at the exact binding site of ABCE1. As shown by (Wang et al., 2019), the time between binding of eIF5B and association of 60S is very short (~0.59 s) and due to its dynamic nature, we most likely would not be able to stall the 48S-eIF5B complex using our protocol. Indeed, only a small number of MS spectra for eIF5B was recorded.

When compared to the yeast 40S ribosome-ABCE1 post-splitting complex (Heuer et al., 2017), we do not observe any large structural differences. However, NTP pocket 1 appears to be more open than NTP pocket 2 (Figures 6A and 6E, coloured frames) and it is similar to the “open state” found by X-ray crystallography (Karcher et al., 2008). Recent single-molecule-based fluorescence resonance energy transfer (smFRET) study on archeal ABCE1 showed that two NTP sites are in asymmetric dynamic equilibrium and both NTP sites can exist in different conformations (Gouridis et al., 2019). Therefore, we propose that in LS48S IC the ABCE1 is present in its asymmetric conformer, where NTP pocket 1 is in the open state, whereas pocket 2 is in the closed state. The position of ABCE1 in the IC suggests steric incompatibility with the human re-initiation factor, eIF2D (Weisser et al., 2017). Indeed, the winged helix (WH) domain of eIF2D was found to interact with the central part of h44 ribosomal RNA in the absence of ABCE1, at the exact position of the ABCE1 Fe-S cluster (I) in our LS48S IC (Figure S4B). This cluster also shows sequence similarity to the C-terminal SUI domain of eIF2D, found to be located in the re-initiation complex at the top of h44 rRNA (Figure 4A) (Weisser et al., 2017).

Regarding the relatively low abundance of eIF3 in our complexes, we believe that after the LS48S complex formation, eIF3 simply detaches from the 40S, probably during the grid preparation as has been consistently observed in structural studies of analogous complexes. The superposition of the eIF3 octamer to our previous structure of the *in vitro* reconstituted 43S PIC (des Georges et al., 2015) showed high structural similarity (RMSD 1.3 Å over all atoms). Consistent with previous study (des Georges et al., 2015), our structure reveals a sizable unassigned density at the mRNA channel exit, interacting mainly with eIF3 a and c (Figures S6B and S6C). Because of its location, it is tempting to attribute this unassigned density to the 5’UTR of mRNA, however its presence in 43S PIC (des Georges et al., 2015), which does not contain any mRNA, strongly contradicts this assignment. Thus, following our previous suggestion, we believe that this density belongs mainly to flexible segments of eIF3d. Finally, the eIF3 b-i-g module is not visible in our structure, however in the case of py48S-eIF5 IC structure (Llácer et al., 2018), it was demonstrated that these subunits relocate together to the solvent site upon start-codon recognition.

CONCLUSIONS

Our cryo-EM structure at 3.0 Å represents the highest resolution reconstruction of a mammalian translation initiation complex till date. It refines our understanding of the architecture of late-stage IC and provides structural insights into the Kozak sequence role in canonical cap-dependent translation initiation. The data presented here demonstrate different interaction networks of the mRNA within the initiation complex based on its sequence. Furthermore, our results demonstrate that the binding of ABCE1 does not impact the conformation of the 48S IC. Finally, our structure reveals the molecular details of the mammalian eIF3 core interactions with the 40S at near-atomic resolution.

ACKNOWLEDGEMENTS

We thank Gabor Papai and Julio Ortiz Espinoza (IGBMC, Strasbourg, France) for assistance in data acquisition and Franck Martin for providing the plasmid for β-globin mRNA. We also thank the High-Performance Computing Centre of the University of Strasbourg (funded by the Equipex Equip@Meso project) for IT support and the staff of the proteomic platform of Strasbourg-Esplanade for conducting the nanoLC-MS/MS analysis (funded by LABEX: ANR-10-LABX-0036 NETRNA). The mass spectrometry instrumentation was granted from Université de Strasbourg (IdEx 2015 Equipement mi-lourd). We thank Alan G. Hinnebusch for critical reading of the manuscript and Cameron Mackereth for numerous useful comments to it. We acknowledge Israel S. Fernández for providing us the molecular model of eIF3 core. This work was supported by the ANR grant ANR-14-ACHN-0024 @RAction program ‘‘ANR CryoEM80S’’ and the ERC-2017-STG #759120 “TransTryp” (to Y.H.).

AUTHOR CONTRIBUTIONS

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

DECLARATION OF INTERESTS

The authors declare no competing interests.

MATERIALS and METHODS

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Guanosine 5′-[β,γ-imido]triphosphate (GMP-PNP)	Merck	Car#148892-91
Protease inhibitor cocktail tablets	Roche	Cat#11873580001
RNasin® Ribonuclease Inhibitors	Promega	Cat#N251B
Rabbit Reticulocyte Lysate nuclease-treated	Promega	Cat#L416A
ScriptCap™ m^{7}G Capping System	Epicentre	Cat#SCCE0625
Deposited Data		
Structure of β-globin LS48S IC	This Paper	PDB ID: 6YAL
Structure of β-globin LS48S+eIF3 IC	This Paper	PDB ID: 6YAM
Structure of H4 LS48S IC	This Paper	PDB ID: 6YAN
Cryo-electron microscopy map of β-globin LS48S IC	This Paper	EMDB ID: EMD-10760
Cryo-electron microscopy map of β-globin LS48S+ eIF3 IC	This Paper	EMDB ID: EMD-10761
Cryo-electron microscopy map of H4 LS48S IC	This Paper	EMDB ID: EMD-10762
Cryo-electron microscopy map of β-globin LS48S IC + ATP	This Paper	EMDB ID: EMD-10763
Cryo-electron microscopy map of β-globin LS48S+eIF3 IC + ATP	This Paper	EMDB ID: EMD-10764
Recombinant DNA		
plasmid for β-globin mRNA	(Simonetti et al. 2016)	N/A
plasmid for H4 mRNA	This Paper	N/A
Software and Algorithms		
SCIPION	(de la Rosa-Trevín et al., 2016)	http://scipion.i2pc.es
Molecular Dynamic Flexible Fitting	(Trabuco et al., 2008)	https://www.ks.uiuc.edu/Research/mdff/
CTFFIND4	(Rohou and Grigorieff, 2015)	http://grigoriefflab.janelia.org/ctf
RELION	(Scheres, 2012)	https://www2.mrc-lmb.cam.ac.uk/groups/scheres/impact.html
UCSF Chimera	(Pettersen et al., 2004).	https://www.cgl.ucsf.edu/chimera/
MotionCor	(Zheng et al., 2017)	https://emcore.ucsf.edu/ucsf-motioncor2
PHENIX 1.9.1692 Phenix.ERRASER	(Adams et al., 2010) (Chou et al., 2016)	http://www.phenix-online.org/
Visual Molecular Dynamics	(Humphrey et al., 1996)	https://www.ks.uiuc.edu/Research/vmd/
Scalable Molecular Dynamics 2	(Phillips et al., 2005)	http://www.phenix-online.org/

Lead Contact and Materials Availability

Further information and requests for reagents should be sent to the Lead Contact, Yaser Hashem (yaser.hashem@u-bordeaux.fr).

In vitro transcription and capping

Human β-globin mRNA was prepared as previously described (Simonetti et al., 2016). The template for mouse H4–12 mRNA (375 nt; accession number X13235) was generated by PCR amplification extended on its 3′-end with a 5′-$(CAA)_9$CAC-3′ tail from plasmid containing the gene synthetized by Proteogenix. The PCR product purification and in vitro transcription of mouse H4–12 mRNA were performed as described for the preparation of β-globin mRNA. The pure transcripts were capped using the ScriptCap™ m^7G Capping System (Epicentre). Radiolabelled transcripts were obtained by substituting the GTP from the kit with [α^{32}P]GTP.

Sample preparation for Cryo-EM

β-globin 48S IC and H4 48S IC were isolated from nuclease-treated rabbit reticulocyte lysate (RRL) (Promega L4960), as previously described (Simonetti et al., 2016) with the main difference that a lower concentration of Mg^{2+} was used for ribosome complex assembly as detailed below. Prior to complexes assembly, the reaction mix in a final volume of 83 µl has been prepared by adding 21 µl of the Amino Acid Mixture (for a final concentration of 0.13 mM) and 80 units of RNasin (Promega Ref. number N2511) to 60 µl of RRL. The mix was incubated at 30°C for 5 minutes to reactivate the ribosomes. In a final volume of 157 µl, complex assembly has been obtained by adding 13 µg of mRNA, guanylyl imidodiphosphate (GMP-PNP) to final concentration of 5 mM and $Mg(OAc)_2$ to keep the final concentration of free Mg^{2+} at 0.5 mM. The reaction is incubated for further 5 minutes at 30°C. β-globin and H4 mRNA assembled complexes were separate on 5%–25% linear sucrose gradient (in buffer containing 25 mM HEPES-KOH [pH 7.6], 79 mM KOAc, 0.5 mM $Mg(OAc)_2$, and 1 mM DTT) by centrifugation at 36,000 rpm in a SW41Ti rotor for 4.5 h at 4°C. Moreover, we have optimized the gradient fraction collection using BioComp Piston Gradient Fractionator™ devise. The formation of translation initiation complexes has been monitored following the ribosome profile via the UV absorbance (optical density [OD] at 260 nm) and the radioactivity profile of the ^{32}P-labeled globin or H4 mRNA. Fractions containing β-globin/48S IC or H4/48S IC were centrifuged at 108,000 rpm (S140AT Sorvall-Hitachi rotor) for 1 h at 4°C and the ribosomal pellet was dissolved in a buffer containing 10 mM HEPES-KOH pH 7.4, 50 mM KOAc, 10 mM NH_4Cl, 5 mM $Mg(OAc)_2$ and 2 mM DTT.

Grids preparation and data collection parameters

The grids were prepared by applying 4 µL of each complex at ~70 nM to 400 mesh holey carbon Quantifoil 2/2 grids (Quantifoil Micro Tools). The grids were blotted for 1.5 sec at 4°C, 100% humidity, using waiting time 30 s, and blot force 4 (Vitrobot Mark IV). The data acquisitions were performed for the β-globin•GMP-PNP and H4•GMP-PNP ICs on a Titan Krios S-FEG instrument (FEI) operated at 300 kV acceleration voltage and at a nominal underfocus of Δz = ~ 0.5 to ~ 3.5 µm using the CMOS Summit K2 direct electron detector 4,096 x 4,096 camera and automated data collection with SerialEM (Mastronarde, 2003) at a nominal magnification of 59,000 x. The K2 camera was used at super-resolution mode and the output movies were binned twice resulting in a pixel size of 1.1Å at the specimen level (the calibrated magnification on the 6.35 µm pixel camera is 115,455 x). The camera was setup to collect 20 frames and frames 3 to 20 were aligned. Total collected dose is ~26 $e^-/Å^2$. In addition, Cryo-EM images of the β-globin•GMP-PNP+ATP were collected on a Polara Tecnai F30 cryo-transmission electron microscope (FEI instruments) operated at 300 keV acceleration voltage and at a nominal underfocus of Δz = ~ 0.5 to ~4.0 µm, using a direct electron detector CMOS (Falcon I) 4,096 x 4,096 camera calibrated at a nominal magnification of 59,000 x, resulting in a pixel size of 1.815 Å.

Image processing

SCIPION (de la Rosa-Trevín et al., 2016) package was used for image processing and 3D reconstruction. MotionCor (Zheng et al., 2017) was used for the movie alignment of 8238 movies from the β-globin complex and 8520 movies for the H4 complex. CTFFIND4 (Rohou and Grigorieff, 2015) was used for the estimation of the contrast transfer function of an average image of the whole stack. Particles were selected in SCIPION. Approximately 1,067,000 particles were selected for the β-globin•GMP-PNP IC, 666,000 particles for the H4•GMP-PNP and 200,000 particles for the β-globin•GMP-PNP+ATP IC. RELION (Scheres, 2012) was used for particle sorting through 3D classification via SCIPION, (please refer to Figure S1 for particle sorting details for all three complexes). Selected classes were refined using RELION's 3D autorefine and the final refined classes were then post-processed using the procedure implemented in RELION applied to the final maps for appropriate masking, B factor sharpening, and resolution validation to avoid over-fitting.

Model building, map fitting and refinement

The four initiation complexes were modelled based on the previous initiation *Oryctolagus cuniculus* complex (PDBID: 5K0Y) (Simonetti et al., 2016) resolved at 5.8 Å. Adjustments of RNA and proteins were done using the visualization and modelling software UCSF Chimera version 1.12 (build 41623) (Pettersen et al., 2004). Sequences of modelled factors from *Oryctolagus cuniculus* were retrieved using BLAST (Altschul et al., 1990) tools in the NCBI database (NCBI Resource Coordinators, 2017) using respective template sequence described below. Templates structures were extracted from the PDB (Berman et al., 2006). ABCE1 from *Saccharomyces cerevisiae* 40S

complex (PDB ID: 5LL6 chain h) (Heuer et al., 2017) was used as template to thread ABCE1 of *Oryctolagus cuniculus* in Swiss-model (Biasini et al., 2014) webservice. The core of initiation factor 3 (eIF3) composed of subunits A, C, E, F, H, K, L and M was extracted from corresponding mammalian eIF3 (PDB ID: 5A5T) (des Georges et al., 2015) and further rmodelled using (Neupane et al., 2019). Initiation factor 3D (eIF3d) from *Nasonia vitripennis* (PDB ID: 5K4B) (Lee et al., 2016) was the template to thread eIF3d of *Oryctolagus cuniculus* in Swiss-model. Eukaryotic Initiation factor 1A (eIF1A) template was extracted from *Saccharomyces cerevisiae* 48S pre-initiation complex (PDB ID: 3JAP chain i) (Llacer et al., 2015) and thread into *Oryctolagus cuniculus* in Swiss-model. The ternary complex (TC) was affined from *Oryctolagus cuniculus* initiation complex (PDB ID: 5K0Y) (Simonetti et al., 2016). Both messenger RNAs (globin and H4) were modelled using modelling tools of Chimera. Refinements were done on all four complexes in their corresponding maps. The refinement workflow followed four major steps that applied to all initiation complexes. First, a Molecular Dynamic Flexible Fitting (MDFF) (Trabuco et al., 2008) ran for 200000 steps with gscale of 1 (potential given to the density map to attract atoms in their density). The trajectories reached a plateau of RMSD curve around frame 160 for the four complexes. A minimization followed the trajectories to relax the system. MDFF ran on VMD (Humphrey et al., 1996) 1.9.2 coupled with NAMD2 (Phillips et al., 2005) v.1.3. software. Next steps of refinement required the usage of several specialized tools for RNA and proteins geometry included as modules in PHENIX (Adams et al., 2010) version 1.13-2998-000 software. Phenix.ERRASER (Chou et al., 2016) is a specialized tool for RNA refinement and Phenix.real_space_refine is specialized for proteins geometry and density fitting refinement. Finally, a last step of minimization using VMD and NAMD2 was applied. Assessment and validation of our models were done by Molprobity (Chen et al., 2010) webservice. Validation statistics are in Table S1.

Mass spectrometry analysis and data post-processing

Protein extracts were precipitated overnight with 5 volumes of cold 0.1 M ammonium acetate in 100% methanol. Proteins were then digested with sequencing-grade trypsin (Promega, Fitchburg, MA, USA) as described previously (Khusainov et al., 2016). Each sample was further analyzed by nanoLC-MS/MS on a QExactive+ mass spectrometer coupled to an EASY-nanoLC-1000 (Thermo-Fisher Scientific, USA). Data were searched against the rabbit UniprotKB sub-database with a decoy strategy (UniprotKB release 2016-08-22, taxon 9986 *Oryctolagus cuniculus*, 23086 forward protein sequences). Peptides and proteins were identified with Mascot algorithm (version 2.5.1, Matrix Science, London, UK) and data were further imported into Proline v1.4 software (http://proline.profiproteomics.fr/). Proteins were validated on Mascot pretty rank equal to 1, and 1% FDR on both peptide spectrum matches (PSM score) and protein sets (Protein Set score). The total number of MS/MS fragmentation spectra was used to quantify each protein from three independent biological replicates (Spectral Count relative quantification). Proline was further used to align the Spectral Count values across all samples. The average of three experiments was normalized in respect to the total number of spectral counts (NSC) for all initiation factors (eIFs) (2807 for β-globin and 2126 for H4). Therefore, the H4 results were multiplied by the

normalization factor of 1.32. Then, the multiplicands for different eIFs were added and the results underwent a second normalization according to the number of trypsin sites (>70% of probability) predicted by PeptideCutter provided by the ExPaSy server (https://web.expasy.org/peptide_cutter). The heat maps were generated in respect to the NSC. The coefficients of variations were calculated as standard of deviation between the normalized NSC and are presented in percentage (see Figure 1B).

Alignments

The alignments of 8 eukaryotic species (*Homo sapiens*, *Mus musculus*, *Danio rerio*, *Drosophila melanogaster*, *Ceanorhabditis elegans*, *Neurospora crassa*, *Saccharomyces cervisiae*, *Arabidopsis thaliana*) shown in the figures, were done using Constraint-based Multiple Alignment Tool (COBALT) (Papadopoulos and Agarwala, 2007) from NCBI and visualized using BoxShade Server (ExPASy).

Data and Code Availability

The atomic coordinates of the β-globin LS48S IC, β-globin LS48S+eIF3 IC and H4 LS48S IC have been deposited in the Protein Data Bank (PDB) under the accession numbers: 6YAL, 6YAM and 6YAN, respectively. The cryo-EM maps of β-globin LS48S IC, β-globin LS48S+eIF3 IC, H4 LS48S IC, β-globin LS48S IC + ATP and β-globin LS48S+eIF3 IC + ATP have been deposited in the Electron Microscopy Data Bank (EMDB) with the accession codes: EMD-10760, EMD-10761, EMD-10762, EMD-10763 and EMD-10764, respectively.

REFERENCES

Adams, P.D., Afonine, P. V, Bunkóczi, 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. Sect. D Biol. Crystallogr. *66*, 213–221.

Aitken, C.E., Beznosková, P., Vlčkova, V., Chiu, W.L., Zhou, F., Valášěk, L.S., Hinnebusch, A.G., and Lorsch, J.R. (2016). Eukaryotic translation initiation factor 3 plays distinct roles at the mRNA entry and exit channels of the ribosomal preinitiation complex. Elife *5*, 1–37.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Aylett, C.H.S., Boehringer, D., Erzberger, J.P., Schaefer, T., and Ban, N. (2015). Structure of a Yeast 40S – eIF1 – eIF1A – eIF3 – eIF3j initiation complex. Nat. Struct. Mol. Biol. *22*, 9–12.

Barthelme, D., Scheele, U., Dinkelaker, S., Janoschka, A., MacMillan, F., Albers, S.V., Driessen, A.J.M., Stagni,

M.S., Bill, E., Meyer-Klaucke, W., et al. (2007). Structural organization of essential iron-sulfur clusters in the evolutionarily highly conserved ATP-binding cassette protein ABCE1. J. Biol. Chem. *282*, 14598–14607.

Barthelme, D., Dinkelaker, S., Albers, S.-V., Londei, P., Ermler, U., and Tampe, R. (2011). Ribosome recycling depends on a mechanistic link between the FeS cluster domain and a conformational switch of the twin-ATPase ABCE1. Proc. Natl. Acad. Sci. *108*, 3228–3233.

Becker, T., Franckenberg, S., Wickles, S., Shoemaker, C.J., Anger, A.M., Armache, J.P., Sieber, H., Ungewickell, C., Berninghausen, O., Daberkow, I., et al. (2012). Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. Nature *482*, 501–506.

Berman, H., Henrick, K., Nakamura, H., and Markley, J.L. (2006). The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res. *35*, D301–D303.

Beznosková, P., Cuchalová, L., Wagner, S., Shoemaker, C.J., Gunišová, S., von der Haar, T., and Valášek, L.S. (2013). Translation Initiation Factors eIF3 and HCR1 Control Translation Termination and Stop Codon Read-Through in Yeast Cells. PLoS Genet. *9*. DOI: 10.1371/journal.pgen.1003962

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

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., et al. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. *42*, W252–W258.

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

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

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 12–21.

Chou, F.C., Echols, N., Terwilliger, T.C., and Das, R. (2016). RNA Structure Refinement Using the ERRASER-Phenix Pipeline. In Nucleic Acid Crystallography. Methods in Molecular Biology, p. vol.1320.

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

Dong, J., Nanda, J.S., Rahman, H., Pruitt, M.R., Shin, B.-S., Wong, C.-M., Lorsch, J.R., and Hinnebusch, A.G.

(2008). Genetic identification of yeast 18S rRNA residues required for efficient recruitment of initiator $tRNA^{Met}$ and AUG selection. Genes Dev. *22*, 2242–2255.

Dong, J., Echeverría, C., Thakur, A., Shin, B., Lorsch, J.R., and Hinnebusch, A.G. (2017). Rps3/uS3 promotes mRNA binding at the 40S ribosome entry channel and stabilizes preinitiation complexes at start codons. PNAS *114*, E2126–E2135.

Dvir, S., Velten, L., Sharon, E., Zeevi, D., Carey, L.B., Weinberger, A., and Segal, E. (2013). Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. Proc. Natl. Acad. Sci. *110*, E2792–E2801.

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.

Erzberger, J.P., Stengel, F., Pellarin, R., Zhang, S., Schaefer, T., Aylett, C.H.S., Cimermančič, P., Boehringer, D., Sali, A., Aebersold, R., et al. (2014). Molecular architecture of the 40S · eIF1 · eIF3 translation initiation complex. Cell *158*, 1123–1135.

Fernandez, I.S., Bai, X.-C., Hussain, T., Kelley, A.C., Lorsch, J.R., Ramakrishnan, V., and Scheres, S.H.W. (2013). Molecular Architecture of a Eukaryotic Translational Initiation Complex. Science (80-.). *342*, 1240585–1240585.

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

Fringer, J.M., Acker, M.G., Fekete, C.A., Lorsch, J.R., and Dever, T.E. (2007). Coupled Release of Eukaryotic Translation Initiation Factors 5B and 1A from 80S Ribosomes following Subunit Joining. Mol. Cell. Biol. *27*, 2384–2397.

des Georges, A., Dhote, V., Kuhn, L., Hellen, C.U.T., Pestova, T. V, Frank, J., and Hashem, Y. (2015). Structure of mammalian eIF3 in the context of the 43S preinitiation complex. Nature *525*, 491–495.

Gouridis, G., Hetzert, B., Kiosze-Becker, K., de Boer, M., Heinemann, H., Nürenberg-Goloub, E., Cordes, T., and Tampé, R. (2019). ABCE1 Controls Ribosome Recycling by an Asymmetric Dynamic Conformational Equilibrium. Cell Rep. *28*, 723–734.

Graifer, D., Molotkov, M., Styazhkina, V., Demeshkina, N., Bulygin, K., Eremina, A., Ivanov, A., Laletina, E., Ven'yaminova, A., and Karpova, G. (2004). Variable and conserved elements of human ribosomes surrounding the mRNA at the decoding and upstream sites. Nucleic Acids Res. *32*, 3282–3293.

Gross, J.D., Moerke, N.J., von der Haar, T., Lugovskoy, A.A., Sachs, A.B., Mccarthy, J.E.G., and Wagner, G. (2003). Ribosome Loading onto the mRNA Cap Is Driven by Conformational Coupling between eIF4G and eIF4E. Cell *115*, 739–750.

Grzegorski, S.J., Chiari, E.F., Robbins, A., Kish, P.E., and Kahana, A. (2014). Natural variability of Kozak

sequences correlates with function in a zebrafish model. PLoS One *9*, e108475.

Hashem, Y., Georges, A., Dhote, V., Langlois, R., Liao, H.Y., Grassucci, R.A., Hellen, C.U.T., Pestova, T. V, and Frank, J. (2013). Structure of the Mammalian Ribosomal 43S Preinitiation Complex Bound to the Scanning Factor DHX29. Cell *153*, 1108–1119.

Heuer, A., Berninghausen, O., Eyler, D.E., Green, R., Preis, A., Hauser, A., Beckmann, R., Barrio-Garcia, C., and Becker, T. (2014). Cryoelectron Microscopic Structures of Eukaryotic Translation Termination Complexes Containing eRF1-eRF3 or eRF1-ABCE1. Cell Rep. *8*, 59–65.

Heuer, A., Gerovac, M., Schmidt, C., Trowitzsch, S., Preis, A., Kötter, P., Berninghausen, O., Becker, T., Beckmann, R., and Tampé, R. (2017). Structure of the 40S-ABCE1 post-splitting complex in ribosome recycling and translation initiation. Nat. Struct. Mol. Biol. *24*, 453–460.

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

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. J. Mol. Graph. *14*, 33–38.

Hussain, T., Llacer, J.L., Fernandez, I.S., Munoz, A., Martin-Marcos, P., Savva, C.G., Lorsch, J.R., Hinnebusch, A.G., and Ramakrishnan, V. (2014). Structural Changes Enable Start Codon Recognition by the Eukaryotic Translation Initiation Complex. Cell *159*, 597–607.

Jackson, R.J., Hellen, C.U.T., and Pestova, T. V. (2010). The Mechanism Of Eucaryotic Translation Initiation And Principles Of Its Regulation. Nat Rev Mol Cell Biol. *11*, 113–127.

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

Khoshnevis, S., Gross, T., Rotte, C., Baierlein, C., Ficner, R., and Krebber, H. (2010). The iron-sulphur protein RNase L inhibitor functions in translation termination. EMBO Rep. *11*, 214–219.

Khusainov, I., Vicens, Q., Bochler, A., Grosse, F., Myasnikov, A., Ménétret, J.-F., Chicher, J., Marzi, S., Romby, P., Yusupova, G., et al. (2016). Structure of the 70S ribosome from human pathogen *Staphylococcus aureus*. Nucleic Acids Res. *44*, 10491–10504.

Kiosze-Becker, K., Heuer, A., Beckmann, R., Beck, M., Becker, T., Gerovac, M., Rashid, U.J., Nürenberg-Goloub, E., Ori, A., and Tampé, R. (2016). Structure of the ribosome post-recycling complex probed by chemical cross-linking and mass spectrometry. Nat. Commun. *7*, 1–9.

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

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

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

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

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

de la Rosa-Trevín, J.M., Quintana, A., del Cano, L., Zaldívar, A., Foche, I., Gutiérrez, J., Gómez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V., et al. (2016). Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. J. Struct. Biol. *195*, 93–99.

Lee, A.S.., Kranzusch, P.J., Doudna, J.A., and Cate, J.H.D. (2016). eIF3d is an mRNA cap-binding protein required for specialized translation initiation. Nature *536*, 96–99.

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

Llácer, J.L., Hussain, T., Marler, L., Lorsch, J.R., Hinnebusch, A.G., and Ramakrishnan, V. (2015). Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Article Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Complex. Mol. Cell *59*, 399–412.

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*, 1–33.

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

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 eIF2β. Cell Rep. *1*, 689–702.

Luna, R.E., Arthanari, H., Hiraishi, H., Akabayov, B., Tang, L., Cox, C., Markus, M.A., Luna, L.E., Ikeda, Y., Watanabe, R., et al. (2013). The interaction between eukaryotic initiation factor 1A and eIF5 retains eIF1 within scanning preinitiation complexes. Biochemistry *52*, 9510–9518.

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

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

Marintchev, A., Edmonds, K.A., Marintcheva, B., Hendrickson, E., Oberer, M., Suzuki, C., Herdy, B., Sonenberg, N., and Wagner, G. (2009). Topology and Regulation of the Human eIF4A/4G/4H Helicase Complex in Translation Initiation. Cell *136*, 447–460.

Martin, F., Barends, S., Jaeger, S., Schaeffer, L., Prongidi-Fix, L., and Eriani, G. (2011). Cap-Assisted Internal Initiation of Translation of Histone H4. Mol. Cell *41*, 197–209.

Martin, F., Ménétret, J.F., Simonetti, A., Myasnikov, A.G., Vicens, Q., Prongidi-Fix, L., Natchiar, S.K., Klaholz, B.P., and Eriani, G. (2016). Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation. Nat. Commun. *7*, 1–7.

Mastronarde, D.N. (2003). SerialEM: A Program for Automated Tilt Series Acquisition on Tecnai Microscopes Using Prediction of Specimen Position. Microsc. Microanal. *9*, 1182–1183.

NCBI Resource Coordinators (2017). Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. *45*, D12–D17.

Neupane, R., Pisareva, V.P., Rodríguez, C.F., Pisarev, A. V, and Fernández, I.S. (2019). A complex IRES at the 5'-UTR of a viral mRNA assembles a functional 48S complex via an uAUG intermediate. BioRxiv 863761. DOI: 10.1101/863761

Nürenberg-Goloub, E., Kratzat, H., Heinemann, H., Heuer, A., Kötter, P., Berninghausen, O., Becker, T., Tampé, R., and Beckmann, R. (2020). Molecular analysis of the ribosome recycling factor ABCE 1 bound to the 30 S post-splitting complex. EMBO J. *e103788*. DOI: 10.15252/embj.2019103788

Papadopoulos, J.S., and Agarwala, R. (2007). COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics *23*, 1073–1079.

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.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. J. Comput. Chem. *26*, 1781–1802.

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

Pisarev, A. V., Skabkin, M.A., Pisareva, V.P., Skabkina, O. V., Rakotondrafara, A.M., Hentze, M.W., Hellen, C.U.T., and Pestova, T. V. (2010). The Role of ABCE1 in Eukaryotic Posttermination Ribosomal Recycling. Mol. Cell *37*, 196–210.

Pisarev, A. V, Kolupaeva, V.G., Pisareva, V.P., Merrick, W.C., Hellen, C.U.T., and Pestova, T. V (2006). Specific functional interactions of nucleotides at key − 3 and + 4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. Genes Dev. 624–636.

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

Rogers, G.W., Richter, N.J., Lima, W.F., and Merrick, W.C. (2001). Modulation of the Helicase Activity of eIF4A by eIF4B, eIF4H, and eIF4F. J. Biol. Chem. *276*, 30914–30922.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. *192*, 216–221.

Scheres, S.H.W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530.

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

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.

Taoka, M., Nobe, Y., Yamaki, Y., Sato, K., Ishikawa, H., Izumikawa, K., Yamauchi, Y., Hirota, K., Nakayama, H., Takahashi, N., et al. (2018). Landscape of the complete RNA chemical modifications in the human 80S ribosome. Nucleic Acids Res. *46*, 9289–9298.

Thiaville, P., Legendre, R., Rojas-Benitez, D., Baudin-Baillieu, A., Hatin, I., Chalancon, G., Glavic, A., Namy, O., and de Crecy-Lagard, V. (2016). Global translational impacts of the loss of the tRNA modification t6A in yeast. Microb. Cell *3*, 29–45.

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.

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

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

Wallace, E., Maufrais, C., Sales-Lee, J., Tuck, L., Oliveira, L. De, Feuerbach, F., Moyrand, F., Natarajan, P., Madhani, H.D., and Janbon, G. (2019). Start codon context controls translation initiation in the fungal kingdom. BioRxiv, 654046. DOI: 10.1101/654046.

Wang, J., Johnson, A.G., Lapointe, C.P., Choi, J., Prabhakar, A., Chen, D.H., Petrov, A.N., and Puglisi, J.D. (2019). eIF5B gates the transition from translation initiation to elongation. Nature *573*, 605–608.

Weisser, M., Schäfer, T., Leibundgut, M., Böhringer, D., Aylett, C.H.S., and Ban, N. (2017). Structural and Functional Insights into Human Re-initiation Complexes. Mol. Cell *67*, 447-456.e7.

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

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

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

Zheng, S.Q., Palovcak, E., Armache, J.-P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods *14*, 331–332.

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

SUPPLEMENTAL INFORMATION

SUPPLEMENTAL FIGURES

Figure S1, related to Figure 1. Particle classification outputs, average and local resolutions of the different β-globin and H4 LS48S ICs cryo-EM reconstructions.

Figure S2, related to Figure 2 and Figure 4. Structural differences between β-globin and H4 LS48S ICs.

Figure S3, related to Figure 3. Differences between mammalian and yeast mRNA trajectories in the mRNA channel and the $tRNA_i^{Met}$ interactions.

Figure S4, related to Figure 6. ABCE1 sequence comparison between eukaryotes and a sterical incompatibility with eIF2D.

Figure S5, related to Figure 3. Atomic model of uS7 in mammalian LS48S IC.

Figure S6, related to Figure 7. Comparison between eIF3 of ICs from near-native conditions and *in vitro* assembly.

Figure S1, related to Fig. 1

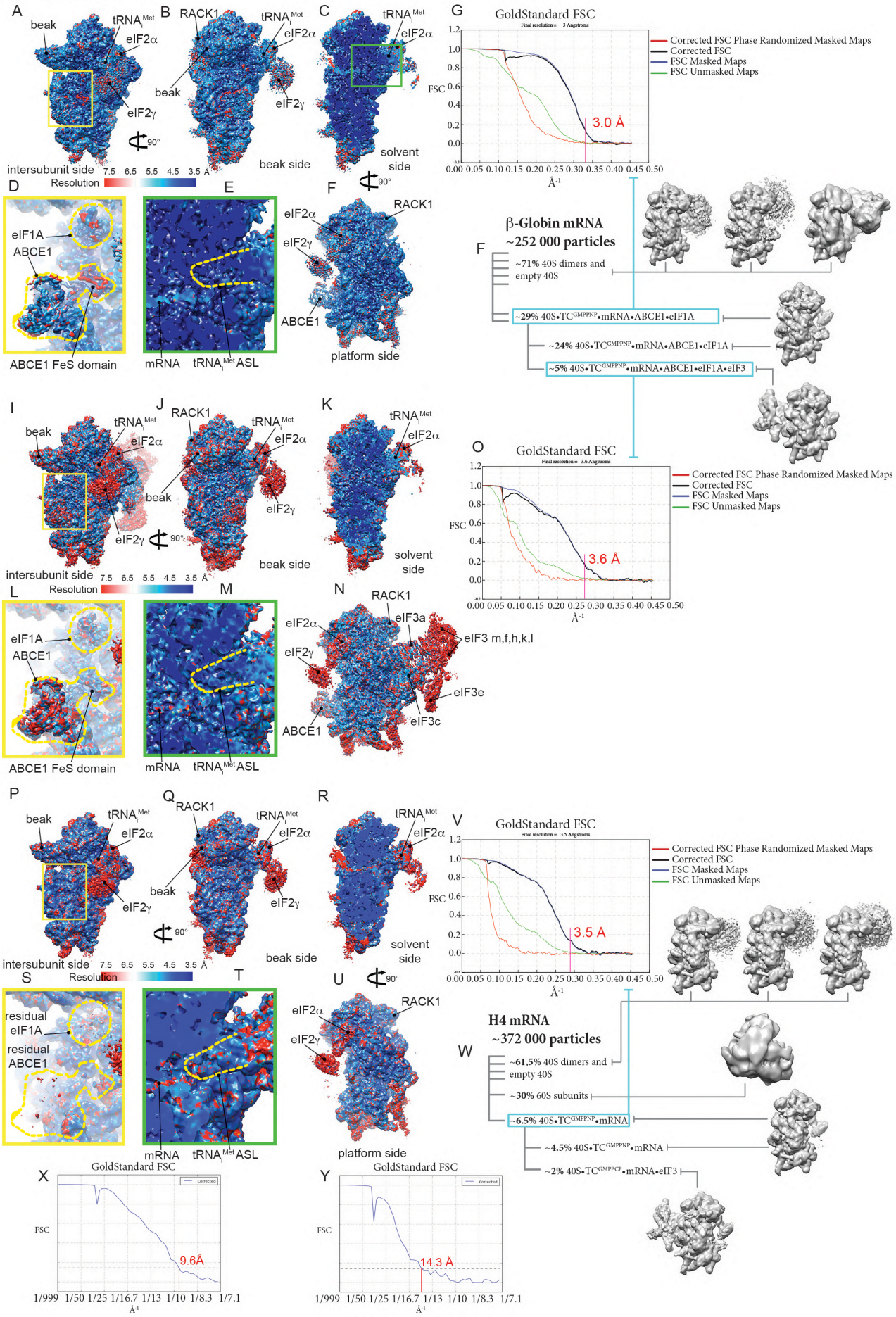


Figure S1, related to Figure 1. Particle classification outputs, average and local resolutions of the different β-globin and H4 LS48S ICs cryo-EM reconstructions. (A-F) Local resolution of the β-globin late-stage IC reconstruction representing a class lacking eIF3 seen from the intersubunit side (A), beak side (B), solvent side (C), platform side (F). (D-E) Insets showing local resolution of eIF1A and ABCE1 (yellow frame) and of mRNA and $tRNA_i^{Met}$ anticodon stem-loop (ASL) (green frame). (G) Average resolution of the reconstruction of the class without eIF3. (H) Particle classification output of the β-globin LS48S IC. (I-N) Local resolution of the β-globin LS48S IC reconstruction representing a class showing eIF3 octamer seen from intersubunit side (I), beak side (J), solvent side (K), platform side (N). (L-M) insets showing local resolution for eIF1A and ABCE1 (yellow frame) and for mRNA and $tRNA_i^{Met}$ ASL (green frame). (O) Average resolution of the reconstruction of the class of β-globin LS48S IC with eIF3. (P-U) Local resolution of the H4 LS48S IC reconstruction representing a class lacking eIF3 seen from the intersubunit side (P), beak side (Q), solvent side (R), platform side (U). The local resolution of eIF3 is not shown, as its average resolution is relatively low. (S-T) Insets showing local resolution of eIF1A and ABCE1 (yellow frame) and of mRNA and $tRNA_i^{Met}$ ASL (green frame). (V) Average resolution of the reconstruction of the class. (W) Particle classification output of the H4 LS48S IC. (X-Y) Average resolutions of the reconstruction of β-globin+ATP LS48S IC reconstructions after their particle sorting (X). Similarly to the counterpart complex with ATP, two major classes stand out, with and without eIF3 (Y).

Figure S2, related to Fig. 2

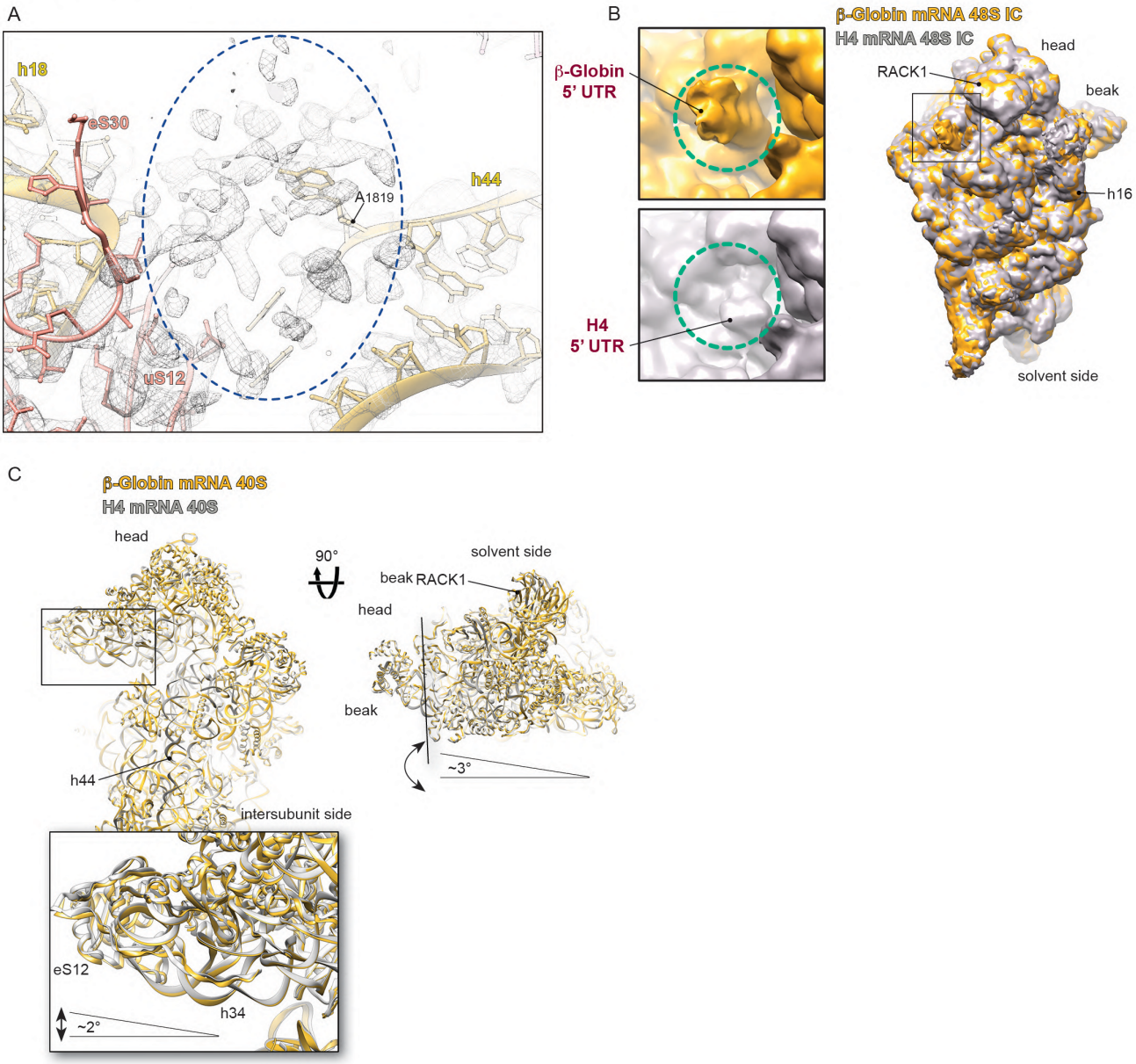


Figure S2, related to Figure 2 and Figure 4. Structural differences between β-globin and H4 LS48S ICs. (A) Residual density for eIF1A in H4 LS48S IC in the A-site, highlighted by dashed blue oval. (B) Comparison of mRNA channel exit between β-globin (orange surface) and H4 (grey surface) 40S reconstructions. (C) Comparison between β-globin (gold ribbons) and H4 (grey ribbons) ICs 40S atomic models. Blowups highlight the conformational changes between both types of complexes at the mRNA channel entrance, the beak and the A, P and E –sites. Dashed coloured ovals indicate the conformational changes in several ribosomal proteins (uS3, eS30 and uS7).

Figure S3, related to Fig. 3

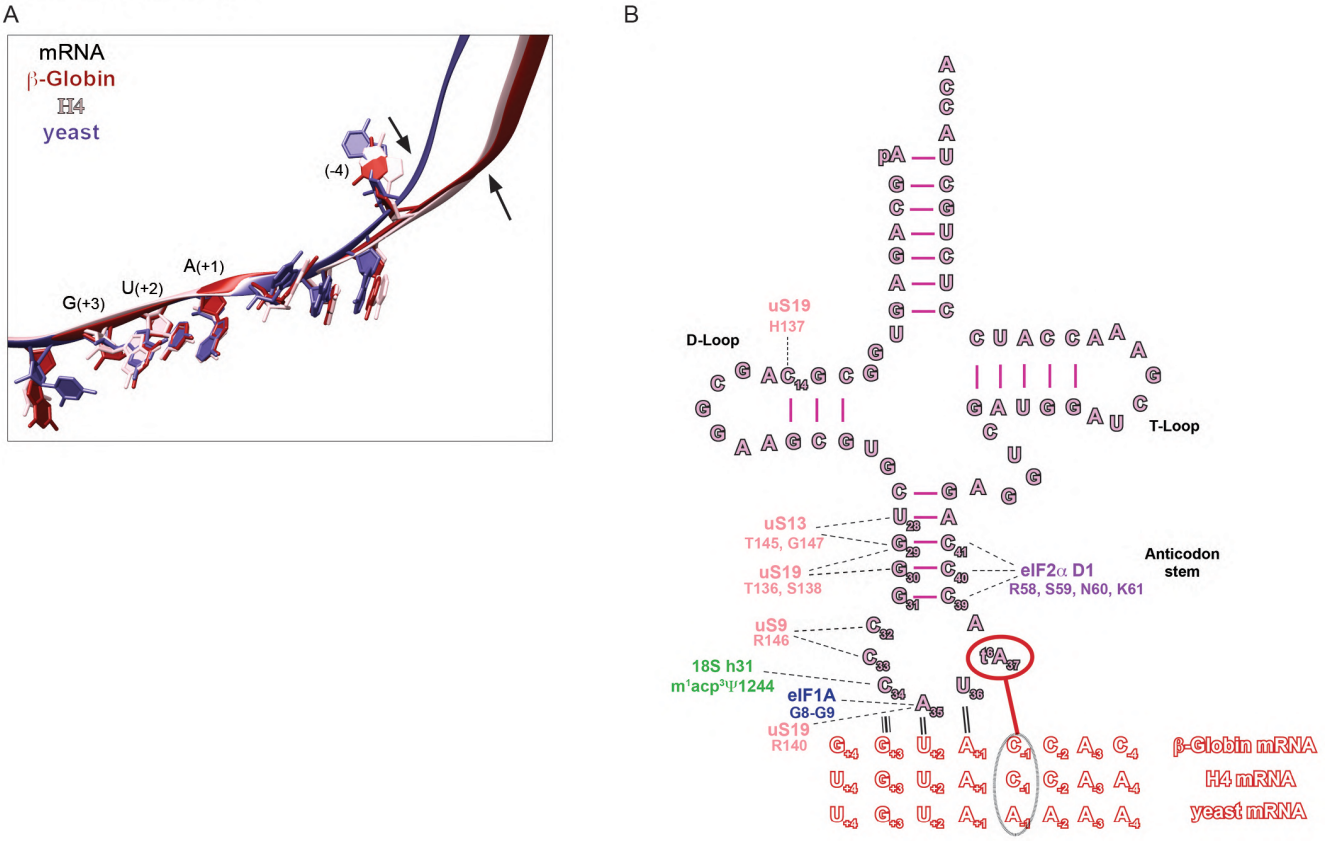


Figure S3, related to Figure 3. Differences between mammalian and yeast mRNA trajectories in the mRNA channel and the $tRNA_i^{Met}$ interactions. (A) Superimposition of yeast optimal mRNA Kozak consensus sequence in py48S-eIF5 IC structure (Llácer et al., 2018) and mammalian mRNAs showing a smoother P/E kink in the case of β-globin and H4 ICs, indicated by arrows. (B) The cloverleaf representation of the $tRNA_i^{Met}$ summarizing the interactions within the LS48S IC described in the Results section. The interaction of $t^6A(37)$ modification with (-1) mRNA position is highlighted by dashed-line oval.

Figure S4, related to Fig. 6

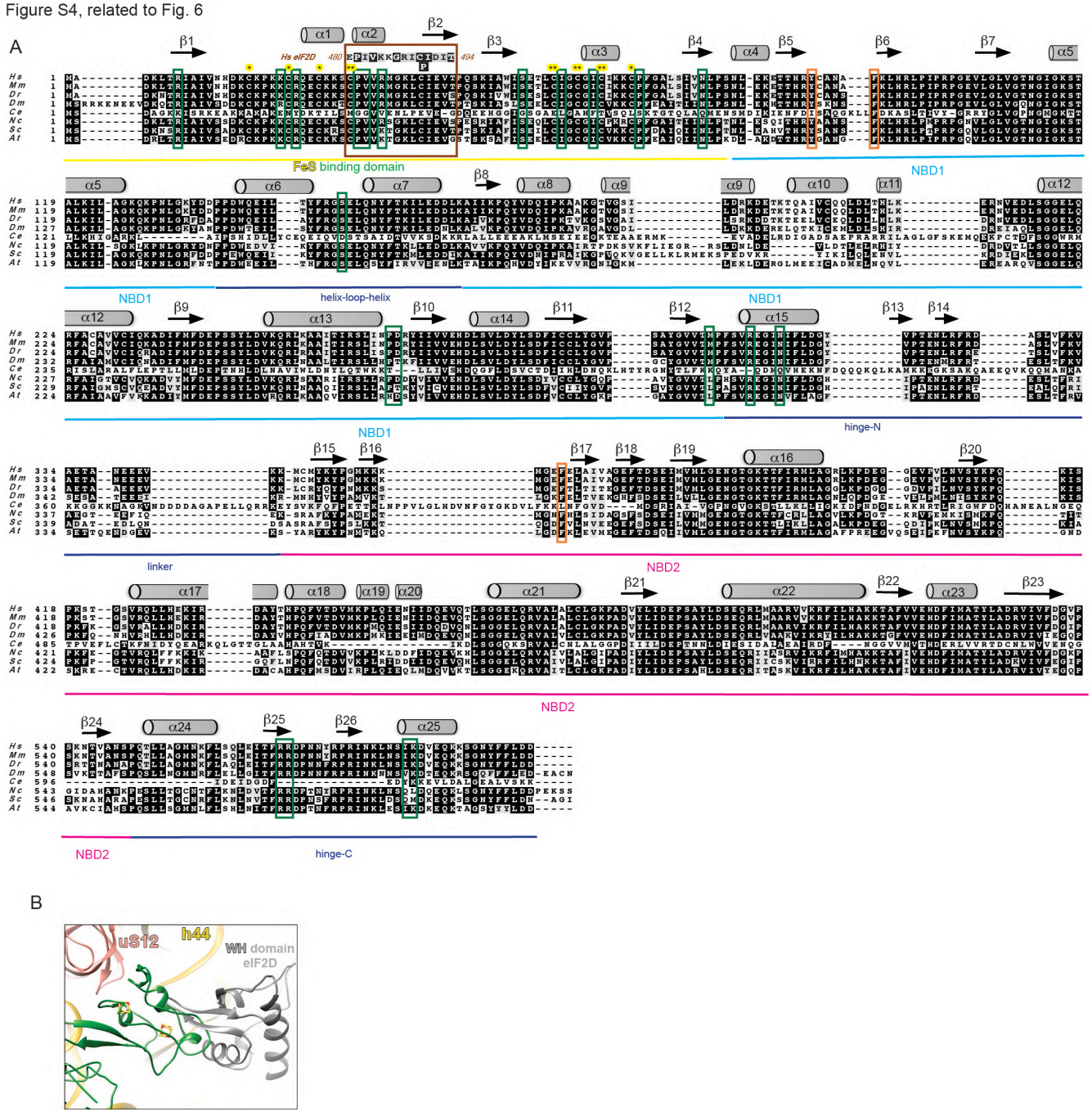


Figure S4, related to Figure 6. ABCE1 sequence comparison between eukaryotes and a sterical incompatibility with eIF2D. (A) Alignment of ABCE1 among eight representative eukaryotic species with secondary structure elements labelled, helices α and β-sheets (according to (Karcher et al., 2008)), and functional domains, indicated by bottom coloured lines. The residues involved in the interactions described in Results section are framed in green. In brown frame, the sequence of Fe-S cluster that showed similarity to the N-terminal part of eIF2D SUI domain (~40% sequence identity). The coordinated cysteines of cluster (I) (*) and (II) (**) are labelled in yellow. In orange frames, the residues responsible for proper ATP/GTP binding in nucleotide binding domains (NBD1 and NBD2) pockets. The accession numbers used for the alignment: *Hs*: NP_002931.2, *Mm*: NP_056566.2, *Dr*: NP_998216.2, *Dm*: NP_648272.1 (pixie), *Ce*: NP_506192.1 ABC transporter class F, *Nc*: XP_963869.3, *Sc*: AJV19484.1, *At*: OAP04903.1). (B) Superimposition of ABCE1 in β-globin LS48S IC (in green) with eIF2D (PDB ID: 5OA3) (Gouridis et al., 2019), showing the sterical clashing of Fe-S cluster (I) with winged-helix (WH) domain of eIF2D (in grey).

Figure S5, related to Fig. 3

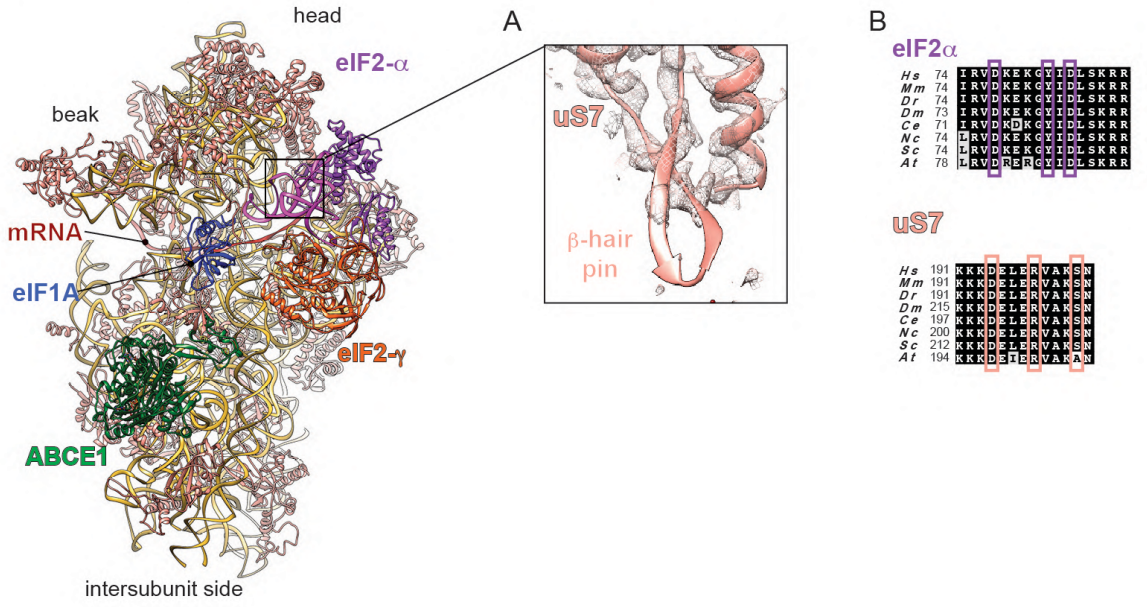


Figure S5, related to Figure 3. Atomic model of uS7 in mammalian LS48S IC. (A) Ribbon representation of uS7 in its electron density, showing the lack of density only in the mRNA-contacting β-hairpin, indicating the flexibility of this part of the protein. (B) Sequence conservation of the interacting residues in eIF2α and uS7, showed by (Visweswaraiah and Hinnebusch, 2017), among eukaryotic species showed in coloured frames.

Figure S6, related to Fig. 7

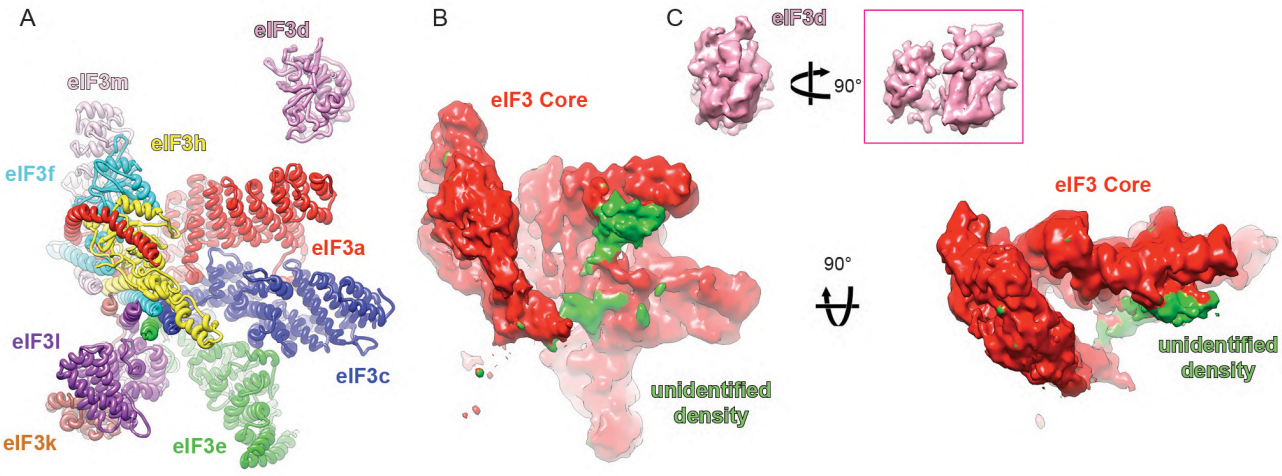


Figure S6, related to Figure 7. Comparison between eIF3 of ICs from near-native conditions and *in vitro* assembly. (A) Atomic model of eIF3 octamer core along with eIF3d subunit (in coloured ribbons). (B) Segmented cryo-EM densities of eIF3 octamer core along with eIF3d (filtered to 8 Å), showing unidentified density (green surface) in contact with eIF3 a and c subunits. (C) Fitting of eIF3d partial crystal structure into its cryo-EM density.

SUPPLEMENTAL TABLES

Table S1, related to Figure 1. Refinement and validation statistics for the three LS48S IC structures from *O.cuniculus*.

Validation[a]	**LS48S IC with β-globin mRNA**	**LS48S IC+eIF3 with β-globin mRNA**	**LS48S IC with H4 mRNA**
PDB ID	**6YAL**	**6YAM**	**6YAN**
Clashscore[b]	11.92 (63rd percentile)	17.51 (40th percentile)	3.67 (97th percentile)
Molprobity score	2.56 (44th percentile)	2.77 (33rd percentile)	1.95 (78th percentile)
Favored rotamers	92.03%	91.49%	93.14%
Allowed rotamers	5.27%	5.41%	5.14%
Poor rotamers	2.70%	3.10%	1.72%
Ramachandran favored	86.02%	85.15%	87.39%
Ramachandran allowed	10.07%	10.83%	9.03%
Ramachandran outliers	3.91%	4.02%	3.58%
Correct sugar puckers	99.68%	99.57%	99.46%
Correct backbone conformation	84.88%	84.52%	83.81%

[a] Compiled using MolProbity (Chen et al., 2010)
[b] Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1,000 atoms.

Table S2, related to Figures 2-7. Summary of the interaction found in the β-globin LS48S and H4 LS48S ICs, described in this paper. The contacts found in this work are highlighted in blue. Stars indicate the interactions with the nucleotide bases.

protein factors	mRNA position β-globin	mRNA position H4	tRNA position	40S component	function
eIF1A					
Lys7	G(+3)				start-codon recognition by the eIF1A NTT
Gly8-Gly9			A(35)		
Trp70	G(+4)*			18S rRNA A1819	Kozak dependent
Arg12	**C(+7)**				
Lys67	G(+6)				
Lys68	U(+5)	**eIF1A absent**			
Gly9, Lys10				18S rRNA C1696	influence on the stability the cognate codon:anticodon duplex by the eIF1A NTT
Lys16, Asn17				C1327	
Asn44, Arg46				C1705 A1817-A1819,	might depend on the mRNA sequence
Lys64 , Arg 62				G604, C605	
Arg82, Tyr84, Gln58				uS12 Leu91, Gly56	
Asp83				eS30 Arg82	
	G(+6)	**U(+6)**		18S rRNA **G616**	
	A(+1)*, C(-1)*			18S rRNA G1203	stabilizing the ASL in the P-site
			C(34)	$m^1acp^3\Psi$1244	
	G(+4)			uS19	
			A(35)	**Arg140**	
			G(29), G(30)	**Thr136**	
				uS13	
			U(28), G(29)	**Thr145, Gly147**	
				uS9	
			C(33), C(32)	**Arg 146**	
				eS26	reinforcement of start-codon recognition in case of suboptimal Kozak sequence
	C(-4)*	**A(-4)***		**His80**	
	(-8)			Ile41	stabilizing the mRNA channel exit site
	(-9)			Arg42	
	(-8), (-9)			Arg100	
				eS28	
	(-5)			Arg66	
	(-7)			Arg67	

protein factors	mRNA position		tRNA position	40S component	function
	β-globin	H4			
eIF2α Arg55 Arg57, Ser58, Asn60, Lys61	C(-3)*		C(39-41)		stabilizing the ASL in the P-site
	mRNA entry channel (+14) to (+18) (+9), (+10) (+12)*, (+13)* A(+13)			uS3 Arg117 helix α (117-128) β-hairpin (142-146) eS30 Lys126 uS5 Ala133	stabilizing the mRNA
ABCE1 HLH and NBD1: Ser150, Arg306, Asn310 NBD2: Lys584, Ile583, Arg566, Arg567 NBD1: **Pro265, Asp266** Fe-S cluster: Arg7, Lys20, Pro66, Asn74 Pro30, Val31, Arg33, Ile56, Ile60				18S rRNA U478, A455, A454, C453, C452 A455, A454, C453 eS24 **C-terminal helix** **Gly128** 18S rRNA C471 G1718 A1719 G470 uS12 Ile50, Leu52, Leu62, Glu63, Ile75	binding to 40S
eIF3c **Asn388** **Arg340, Asn384** **Gly341, Lys343** **Arg450** **Lys342** **Thr391, Tyr392** **Lys343** eIF3a **Asn10** **Lys13** **Arg14** **Phe18**				eS27 **Glu75** **Thr61** **Gln65** **Cys59** 18S rRNA **G925** **C1112** **U1116** eS1 **Asp77** **Asn76** **Asp191** **Pro190**	binding to the 40S platform site
Gln6, Arg7, Arg41, Gln44, Lys45	**5'UTR**				binding to mRNA 5'UTR
eIF3d **S166–E172, Asn513, Lys514**	**5'UTR**				

4. CONCLUSION AND PERSPECTIVES

Our team is investigating the translation initiation complexes of several species by cryo-electron microscopy. One point of interest of the team is to depict the structure of kinetoplastids translation complexes in order to compare them to their counterparts in their mammalian hosts to highlight structural specificities and differences. Therefore my PhD project was to model several translation initiation complexes, including a native 40S of *T.cruzi*, a 80S ribosome of *T.cruzi* and a 43S preinitiation complex of *T.cruzi*, based on cryo-EM data.

On the 40S native *T.cruzi* complex, we identified an empty density located on the left foot of the small subunit 40S. By searching in the mass spectrometry data, we identified a ribosomal protein that was never modeled before. We modeled it with a mix of *ab initio* and homology modeling (for its two RNA Recognition Motifs) to create the first model of the kinetoplastid-specific ribosomal protein (KSRP). Also, we found out that the model fitted perfectly on several kinetoplastid ribosomes including 80S of *L.major*, *L.donovani, T.brucei* and *T.cruzi*. KSRP is present in both pre-initiation complex and full ribosome and is bound to the left foot of 40S and interacts through its two RNA Recognition Motifs (RRM) with the expansion segments ES6S and ES3S. The first hypothesis about the role of KSRP is that its binding suggests KSRP to acts as a rRNA chaperone in a zone that is completely specific to parasites compared to mammalian complex. Its depletion is lethal to the parasite, indicating that it also plays a crucial role in the mRNA translation regulation. The short-term mandatory perspective I see for this project is to undergo functional and mutagenesis studies on KSRP. This will help gaining deeper understanding about KSRP and its interactions with the ribosomes, by revealing residues that are essential for its binding to the ribosome. In term, such knowledge could be the starting point for further investigating how to block it from binding to the 40S. For now, we only know that its knockdown induces the death of the parasite. The need to push further the understanding of its purpose on the ribosome and if it plays a specific role in the kinetoplastid translation initiation is a prerequisite. The long-term perspective is to know at which step blocking this protein is fatal to the parasite and to improve the resolution in order to pursue drug design study on KSRP.

The goal of modeling a 80S ribosome of *T.cruzi,* and completing several 80S ribosome structure including *L.donovani* and *L.major,* was to reveal the structure of all kinetoplastid-specific features. The large RNA expansion segments (ES) have already been

modeled at an intermediate resolution (5.57 Å) from the structure of the *T.brucei* 80S ribosome (PDB : 4V8M (Hashem et al., 2013a)), which was our starting point for the expansion segments model. After modeling and analyzing several 80S kinetoplastids ribosomes, we concluded that expansion segments are not random extensions at the surface of ribosomes but formed an organized layer interacting with core rRNA and r-proteins, recruit kinetoplastid-specific proteins and act as dynamic gatekeepers. Furthermore, we can observe that some expansion segments are different among different species of kinetoplastids, for instance ES6S of *T.brucei* lacks one stem loops of 106 nucleotides compared to *T.cruzi*., suggesting that ESs might play different/additional roles between species of the same phylum. An obvious follow-up of this work would be the investigation of the impact of deletion and interchange of each expansion segments and observe their effect on the translation process. It will be a long-term perspective but finding the exact role of each expansion segment and if they play a role in translation or in another process is crucial to understanding why kinetoplastids ribosomes have evolved such large and structured RNA extensions on both ribosome subunits.

In the 43S preinitiation complex (PIC) of *T.cruzi*, we discovered a new helicase that we have termed k-DDX60 that binds in the mRNA channel entry side. K-DDX60 has a domain that is completely inserted in the mRNA channel, which is intriguing and can be a clue to its function in initiating translation. Also, due to k-DDX60 extended and strong interactions within the 43S preinitiation complex and the small subunit, several so far structurally uncharacterized interactions have been observed, some of which were partially revealed by other biochemical techniques in different eukaryotic species, some others are completely new. We believe that this was achieved thanks to the gained stability of the whole 43S complex through the interaction of k-DDX60. This complex also highlighted the potential role of expansion segments ES6S and ES7S as a docking platform for eIF3. Indeed, I compared this model with the 80S ribosome of *T.cruzi* that I built during my first year of PhD. The result was most intriguing because ES6S and ES7S were in completely different conformation because of the eIF3 accommodation to the 43S PIC suggesting the role of these expansion segments in the translation initiation. Finally, I built the first eIF3 model from a kinetoplastid specie. It was already known that the subunit eIF3m was lacking in kinetoplastids compared to mammalian, but we observed the structural rearrangement and compensation of kinetoplastid eIF3 subunits. For this PIC of *T.cruzi*, my work and the modelling of the whole complex that I provided raises several questions and projects. One logical follow-up would be to improve the purification protocol of this complex and rederive its structure at an improved resolution in order to obtain the structure of the missing domains

of k-DDX60 and better understand its interaction with other initiation factors. Here again, similarly to KSRP, functional studies are needed on this big helicase k-DDX60 to understand its positioning at the mRNA entry site and its intriguing insert domain going directly into the P-site of the ribosome. From my point of view, this insert domain blocks the mRNA channel to avoid any translation from occurring, which raises the question of the exact role of this helicase. A more long-term question about this complex would be where to place it in the translation initiation process landscape? is it a kinetoplastid-specific step or is it maybe a hibernation state of the complex, mediated by k-DDX60 blocking the mRNA channel? Finally, a future question to answer would be to know how this helicase releases from the PIC because. Again, if it appears that k-DDX60 plays a key role in kinetoplastid translation initiation, it could be a potential therapeutic target and blocking it could lead to the death of the parasites.

Overall, through the modeling of kinetoplastid translation initiation complexes, I highlighted and modeled the first 3D structure of several kinetoplastid structural specificities: a new kinetoplastid-specific ribosomal protein (KSRP), a new kinetoplastid-specific helicase k-DDX60, all the expansion segments of *T.cruzi* 80S ribosome, the first model of eIF3 from kinetoplastids. Despite having the same general translation mechanism, the kinetoplastid translation initiation displays several unique structural features compared to their mammalian hosts. Several major questions remain about these structural features roles in translation to fully understand translation in kinetoplastids.

To finish, all the perspectives point to a common goal: finding specific weaknesses in the kinetoplastid translation through these specific structural features that could be targeted specifically. This would lead to the development of safer anti-parasitic therapeutic strategies with less side effects and cytotoxicity. I sincerely hope that all my work and structures will serve as a solid basis for the fight against these impactful, yet neglected, diseases.

REFERENCES

Adrian, M., Dubochet, J., Lepault, J., McDowall, A.W., 1984. Cryo-electron microscopy of viruses. Nature 308, 32–36. https://doi.org/10.1038/308032a0

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D., 2012. Towards automated crystallographic structure refinement with *phenix.refine*. Acta Crystallogr. D Biol. Crystallogr. 68, 352–367. https://doi.org/10.1107/S0907444912001308

Aitken, C.E., Lorsch, J.R., 2012. A mechanistic overview of translation initiation in eukaryotes. Nat. Struct. Mol. Biol. 19, 568–576. https://doi.org/10.1038/nsmb.2303

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anger, A., Armache, J.-P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D., Beckmann, R., 2013. Structures of the human and Drosophila 80S ribosome. Nature 497, 80–5. https://doi.org/10.1038/nature12104

Apirion, D., Miczak, A., 1993. RNA processing in prokaryotic cells. BioEssays 15, 113–120. https://doi.org/10.1002/bies.950150207

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M., Depledge, D.P., Fischer, S., Gajria, B., Gao, X., Gardner, M.J., Gingle, A., Grant, G., Harb, O.S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., Kissinger, J.C., Kraemer, E., Li, W., Logan, F.J., Miller, J.A., Mitra, S., Myler, P.J., Nayak, V., Pennington, C., Phan, I., Pinney, D.F., Ramasamy, G., Rogers, M.B., Roos, D.S., Ross, C., Sivam, D., Smith, D.F., Srinivasamoorthy, G., Stoeckert, C.J., Subramanian, S., Thibodeau, R., Tivey, A., Treatman, C., Velarde, G., Wang, H., 2010. TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. 38, D457–D462. https://doi.org/10.1093/nar/gkp851

Bai, X.-C., Fernandez, I.S., McMullan, G., Scheres, S.H.W., 2013. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. eLife 2, e00461. https://doi.org/10.7554/eLife.00461

Baker, N., de Koning, H.P., Mäser, P., Horn, D., 2013. Drug resistance in African trypanosomiasis: the melarsoprol and pentamidine story. Trends Parasitol. 29. https://doi.org/10.1016/j.pt.2012.12.005

Ban, N., 2000. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 A Resolution. Science 289, 905–920. https://doi.org/10.1126/science.289.5481.905

Banerjee, A.K., 1980. 5'-terminal cap structure in eucaryotic messenger ribonucleic acids. Microbiol. Rev. 44, 175–205. https://doi.org/10.1128/MMBR.44.2.175-205.1980

Barrett, M.P., Croft, S.L., 2012. Management of trypanosomiasis and leishmaniasis. Br. Med. Bull. 104, 175–196. https://doi.org/10.1093/bmb/lds031

Bell, S.P., Dutta, A., 2002. DNA Replication in Eukaryotic Cells. Annu. Rev. Biochem. 71, 333–374. https://doi.org/10.1146/annurev.biochem.71.110601.135425

Ben-Shem, A., Jenner, L., Yusupova, G., Yusupov, M., 2010. Crystal structure of the eukaryotic ribosome. Science 330, 1203–1209. https://doi.org/10.1126/science.1194294

Berg, J.M., Tymoczko, J.L., Stryer, L., Stryer, L., 2002. Biochemistry, 5th ed. ed. W.H. Freeman, New York.

Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.H., Srinivasan, A.R., Schneider, B., 1992. The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. Biophys. J. 63, 751–759. https://doi.org/10.1016/S0006-3495(92)81649-1

Borst, P., Grivell, L.A., 1971. Mitochondrial ribosomes. FEBS Lett. 13, 73–88. https://doi.org/10.1016/0014-5793(71)80204-1

Brun, R., Blum, J., Chappuis, F., Burri, C., 2010. Human African trypanosomiasis. Lancet Lond. Engl. 375, 148–159. https://doi.org/10.1016/S0140-6736(09)60829-1

Buckner, F.S., Wilson, A.J., White, T.C., Van Voorhis, W.C., 1998. Induction of resistance to azole drugs in Trypanosoma cruzi. Antimicrob. Agents Chemother. 42, 3245–3250.
Burley, S.K., Berman, H.M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., Christie, C., Dalenberg, K., Duarte, J.M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D.S., Green, R.K., Guranović, V., Guzenko, D., Hudson, B.P., Kalro, T., Liang, Y., Lowe, R., Namkoong, H., Peisach, E., Periskova, I., Prlić, A., Randle, C., Rose, A., Rose, P., Sala, R., Sekharan, M., Shao, C., Tan, L., Tao, Y.-P., Valasatava, Y., Voigt, M., Westbrook, J., Woo, J., Yang, H., Young, J., Zhuravleva, M., Zardecki, C., 2019. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Res. 47, D464–D474. https://doi.org/10.1093/nar/gky1004
Campbell, D.A., Kubo, K., Clark, C.G., Boothroyd, J.C., 1987. Precise identification of cleavage sites involved in the unusual processing of trypanosome ribosomal RNA. J. Mol. Biol. 196, 113–124. https://doi.org/10.1016/0022-2836(87)90514-6
CDC - DPDx - American Trypanosomiasis [WWW Document], 2019. URL https://www.cdc.gov/dpdx/trypanosomiasisamerican/index.html (accessed 2.3.20).
CDC - Global Health - Neglected Tropical Diseases [WWW Document], 2020. URL https://www.cdc.gov/globalhealth/ntd/index.html (accessed 2.3.20).
Chan, C.W., Badong, D., Rajan, R., Mondragón, A., 2020. Crystal structures of an unmodified bacterial tRNA reveal intrinsic structural flexibility and plasticity as general properties of unbound tRNAs. RNA N. Y. N 26, 278–289. https://doi.org/10.1261/rna.073478.119
Chappuis, F., Sundar, S., Hailu, A., Ghalib, H., Rijal, S., Peeling, R.W., Alvar, J., Boelaert, M., 2007. Visceral leishmaniasis: what are the needs for diagnosis, treatment and control? Nat. Rev. Microbiol. 5, 873–882. https://doi.org/10.1038/nrmicro1748
Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 66, 12–21. https://doi.org/10.1107/S0907444909042073
Cheng, Y., 2018. Single-particle cryo-EM—How did it get here and where will it go. Science 361, 876–880. https://doi.org/10.1126/science.aat4346
Cheng, Y., Grigorieff, N., Penczek, P.A., Walz, T., 2015. A Primer to Single-Particle Cryo-Electron Microscopy. Cell 161, 438–449. https://doi.org/10.1016/j.cell.2015.03.050
Chou, F.-C., Echols, N., Terwilliger, T.C., Das, R., 2016. RNA Structure Refinement Using the ERRASER-Phenix Pipeline. Methods Mol. Biol. Clifton NJ 1320, 269–282. https://doi.org/10.1007/978-1-4939-2763-0_17
Clayton, C., 2016. Gene expression in Kinetoplastids. Curr. Opin. Microbiol., Host-microbe interactions: parasites/fungi/viruses 32, 46–51. https://doi.org/10.1016/j.mib.2016.04.018
Cornford, E.M., Bocash, W.D., Braun, L.D., Crane, P.D., Oldendorf, W.H., MacInnis, A.J., 1979. Rapid distribution of tryptophol (3-indole ethanol) to the brain and other tissues. J. Clin. Invest. 63, 1241–1248.
Crick, F.H., 1958. On protein synthesis. Symp. Soc. Exp. Biol. 12, 138–163.
Crowther, R.A., Amos, L.A., Finch, J.T., De Rosier, D.J., Klug, A., 1970. Three Dimensional Reconstructions of Spherical Viruses by Fourier Synthesis from Electron Micrographs. Nature 226, 421–425. https://doi.org/10.1038/226421a0
Curry, S., 2015. Structural Biology: A Century-long Journey into an Unseen World. Interdiscip. Sci. Rev. 40, 308–328. https://doi.org/10.1179/0308018815Z.000000000120
Dalgarno, L., Shine, J., 1973. Conserved Terminal Sequence in 18S rRNA May Represent Terminator Anticodons. Nature. New Biol. 245, 261–262. https://doi.org/10.1038/newbio245261a0
De Rosier, D.J., Klug, A., 1968. Reconstruction of Three Dimensional Structures from Electron Micrographs. Nature 217, 130–134. https://doi.org/10.1038/217130a0
Dever, T.E., Green, R., 2012. The Elongation, Termination, and Recycling Phases of Translation in

Eukaryotes. Cold Spring Harb. Perspect. Biol. 4. https://doi.org/10.1101/cshperspect.a013706
Docampo, R., de Souza, W., Miranda, K., Rohloff, P., Moreno, S.N.J., 2005. Acidocalcisomes ? conserved from bacteria to man. Nat. Rev. Microbiol. 3, 251–261. https://doi.org/10.1038/nrmicro1097
Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowall, A.W., Schultz, P., 1988. Cryo-electron microscopy of vitrified specimens. BioPhys 129–228.
Dubochet, J., Chang, J.-J., Freeman, R., Lepault, J., McDowall, A.W., 1982. Frozen aqueous suspensions. Ultramicroscopy 10, 55–61. https://doi.org/10.1016/0304-3991(82)90187-5
Dubochet, J., Stahlberg, H., 2001. Electron Cryomicroscopy, in: John Wiley & Sons, Ltd (Ed.), Encyclopedia of Life Sciences. John Wiley & Sons, Ltd, Chichester, UK, p. a0002999. https://doi.org/10.1038/npg.els.0002999
Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of *Coot*. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501. https://doi.org/10.1107/S0907444910007493
Ezra, F.S., Lee, C.-H., Kondo, N.S., Danyluk, S.S., Sarma, R.H., 1977. Conformational properties of purine-pyrimidine and pyrimidine-purine dinucleoside monophosphates. Biochemistry 16, 1977–1987. https://doi.org/10.1021/bi00628a035
Fernandez-Leiro, R., Scheres, S.H.W., 2016. Unravelling biological macromolecules with cryo-electron microscopy. Nature 537, 339–346. https://doi.org/10.1038/nature19948
Field, M.C., Horn, D., Fairlamb, A.H., Ferguson, M.A.J., Gray, D.W., Read, K.D., De Rycker, M., Torrie, L.S., Wyatt, P.G., Wyllie, S., Gilbert, I.H., 2017. Anti-trypanosomatid drug discovery: an ongoing challenge and a continuing need. Nat. Rev. Microbiol. 15, 217–231. https://doi.org/10.1038/nrmicro.2016.193
Finkelstein, A.V., Razin, S.V., Spirin, A.S., 2018. Intersubunit Mobility of the Ribosome. Mol. Biol. 52, 799–811. https://doi.org/10.1134/S0026893318060080
Franco, J.R., Cecchi, G., Priotto, G., Paone, M., Diarra, A., Grout, L., Mattioli, R.C., Argaw, D., 2017. Monitoring the elimination of human African trypanosomiasis: Update to 2014. PLoS Negl. Trop. Dis. 11, e0005585. https://doi.org/10.1371/journal.pntd.0005585
Franco, J.R., Simarro, P.P., Diarra, A., Ruiz-Postigo, J.A., Jannin, J.G., 2014. The journey towards elimination of gambiense human African trypanosomiasis: not far, nor easy. Parasitology 141, 748–760. https://doi.org/10.1017/S0031182013002102
Frank, J., 2002. Single-Particle Imaging of Macromolecules by Cryo-Electron Microscopy. Annu. Rev. Biophys. Biomol. Struct. 31, 303–319. https://doi.org/10.1146/annurev.biophys.31.082901.134202
Frank, J., 1975. Averaging of low exposure electron micrographs of non-periodic objects. Ultramicroscopy 1, 159–162. https://doi.org/10.1016/s0304-3991(75)80020-9
Fraser, C.S., Doudna, J.A., 2007. Structural and mechanistic insights into hepatitis C viral translation initiation. Nat. Rev. Microbiol. 5, 29–38. https://doi.org/10.1038/nrmicro1558
Ganapathiraju, M.K., Klein-Seetharaman, J., Balakrishnan, N., Reddy, R., 2004. Characterization of protein secondary structure. IEEE Signal Process. Mag. 21, 78–87. https://doi.org/10.1109/MSP.2004.1296545
Garcia, S., Ramos, C.O., Senra, J.F.V., Vilas-Boas, F., Rodrigues, M.M., Campos-de-Carvalho, A.C., Ribeiro-Dos-Santos, R., Soares, M.B.P., 2005. Treatment with benznidazole during the chronic phase of experimental Chagas' disease decreases cardiac alterations. Antimicrob. Agents Chemother. 49, 1521–1528. https://doi.org/10.1128/AAC.49.4.1521-1528.2005
Genick, U.K., Borgstahl, G.E.O., Ng, K., Ren, Z., Pradervand, C., Burke, P.M., Šrajer, V., Teng, T.-Y., Schildkamp, W., McRee, D.E., Moffat, K., Getzoff, E.D., 1997. Structure of a Protein Photocycle Intermediate by Millisecond Time-Resolved Crystallography. Science 275, 1471–1475. https://doi.org/10.1126/science.275.5305.1471
Georges, A. des, Dhote, V., Kuhn, L., Hellen, C.U.T., Pestova, T.V., Frank, J., Hashem, Y., 2015. Structure of mammalian eIF3 in the context of the 43S preinitiation complex. Nature 525,

491–495. https://doi.org/10.1038/nature14891
Gilbert, W., 1978. Why genes in pieces? Nature 271, 501–501. https://doi.org/10.1038/271501a0
Gilmartin, G.M., 2005. Eukaryotic mRNA 3′ processing: a common means to different ends. Genes Dev. 19, 2517–2521. https://doi.org/10.1101/gad.1378105
Green and, R., Noller, H.F., 1997. RIBOSOMES AND TRANSLATION. Annu. Rev. Biochem. 66, 679–716. https://doi.org/10.1146/annurev.biochem.66.1.679
Harris, E.H., Boynton, J.E., Gillham, N.W., 1994. Chloroplast ribosomes and protein synthesis. Microbiol. Rev. 58, 700–754.
Hashem, Y., des Georges, A., Fu, J., Buss, S.N., Jossinet, F., Jobe, A., Zhang, Q., Liao, H.Y., Grassucci, R.A., Bajaj, C., Westhof, E., Madison-Antenucci, S., Frank, J., 2013a. High-resolution cryo-electron microscopy structure of the Trypanosoma brucei ribosome. Nature 494, 385–389. https://doi.org/10.1038/nature11872
Hashem, Y., des Georges, A., Dhote, V., Langlois, R., Liao, H.Y., Grassucci, R.A., Hellen, C.U.T., Pestova, T.V., Frank, J., 2013b. Structure of the Mammalian Ribosomal 43S Preinitiation Complex Bound to the Scanning Factor DHX29. Cell 153, 1108–1119. https://doi.org/10.1016/j.cell.2013.04.036
Hauptman, H.A., 1990. History of X-ray crystallography. Struct. Chem. 1, 617–620. https://doi.org/10.1007/BF00674136
Hinnebusch, A.G., 2006. eIF3: a versatile scaffold for translation initiation complexes. Trends Biochem. Sci. 31, 553–562. https://doi.org/10.1016/j.tibs.2006.08.005
Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M., Schuster, P., 1994. Fast folding and comparison of RNA secondary structures. Monatshefte F�r Chem. Chem. Mon. 125, 167–188. https://doi.org/10.1007/BF00818163
Holland, J.A., Hansen, M.R., Du, Z., Hoffman, D.W., 1999. An examination of coaxial stacking of helical stems in a pseudoknot motif: the gene 32 messenger RNA pseudoknot of bacteriophage T2. RNA N. Y. N 5, 257–271. https://doi.org/10.1017/s1355838299981360
Hong, W., Zeng, J., Xie, J., 2014. Antibiotic drugs targeting bacterial RNAs. Acta Pharm. Sin. B, Special Issue: Research Progress of Anti-infective Drugs 4, 258–265. https://doi.org/10.1016/j.apsb.2014.06.012
Honigberg, B.M., 1963. A contribution to systematics of the non-pigmented flagellates [WWW Document]. Proceedings. URL https://geoscience.net/research/037/582/037582270.php (accessed 2.3.20).
Jackson, R.J., Hellen, C.U.T., Pestova, T.V., 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. 11, 113–127. https://doi.org/10.1038/nrm2838
Jost, D., Everaers, R., 2010. Prediction of RNA multiloop and pseudoknot conformations from a lattice-based, coarse-grain tertiary structure model. J. Chem. Phys. 132, 095101. https://doi.org/10.1063/1.3330906
Kaledhonkar, S., Fu, Z., Caban, K., Li, W., Chen, B., Sun, M., Gonzalez, R.L., Frank, J., 2019. Late steps in bacterial translation initiation visualized using time-resolved cryo-EM. Nature 570, 400–404. https://doi.org/10.1038/s41586-019-1249-5
Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858. https://doi.org/10.1038/nprot.2015.053
Kendrew, J., Bodo, G., Dintzis, H., Parrish, R., Wyckoff, H., Phillips, D., 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature 181, 662--666.
Kolupaeva, V.G., Unbehaun, A., Lomakin, I.B., Hellen, C.U.T., Pestova, T.V., 2005. Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association. RNA N. Y. N 11, 470–486. https://doi.org/10.1261/rna.7215305
Korostelev, A., Trakhanov, S., Laurberg, M., Noller, H.F., 2006. Crystal Structure of a 70S Ribosome-tRNA Complex Reveals Functional Interactions and Rearrangements. Cell 126,

1065–1077. https://doi.org/10.1016/j.cell.2006.08.032
Kozak, M., 1999. Initiation of translation in prokaryotes and eukaryotes. Gene 234, 187–208. https://doi.org/10.1016/s0378-1119(99)00210-3
Leontis, N.B., Westhof, E., 2001. Geometric nomenclature and classification of RNA base pairs. RNA 7, 499–512.
Lescoute, A., Westhof, E., 2006. Topology of three-way junctions in folded RNAs. RNA 12, 83–93. https://doi.org/10.1261/rna.2208106
Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., Cheng, Y., 2013. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat. Methods 10, 584–590. https://doi.org/10.1038/nmeth.2472
Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A.J., Moriarty, N.W., Oeffner, R.D., Poon, B.K., Prisant, M.G., Read, R.J., Richardson, J.S., Richardson, D.C., Sammito, M.D., Sobolev, O.V., Stockwell, D.H., Terwilliger, T.C., Urzhumtsev, A.G., Videau, L.L., Williams, C.J., Adams, P.D., 2019. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. Acta Crystallogr. Sect. Struct. Biol. 75, 861–877. https://doi.org/10.1107/S2059798319011471
Liu, Z., Gutierrez-Vargas, C., Wei, J., Grassucci, R.A., Ramesh, M., Espina, N., Sun, M., Tutuncuoglu, B., Madison-Antenucci, S., Woolford, J.L., Tong, L., Frank, J., 2016. Structure and assembly model for the *Trypanosoma cruzi* 60S ribosomal subunit. Proc. Natl. Acad. Sci. 113, 12174–12179. https://doi.org/10.1073/pnas.1614594113
Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S.Y., Alvarado, M., Anderson, H.R., Anderson, L.M., Andrews, K.G., Atkinson, C., Baddour, L.M., Barker-Collo, S., Bartels, D.H., Bell, M.L., Benjamin, E.J., Bennett, D., Bhalla, K., Bikbov, B., Bin Abdulhak, A., Birbeck, G., Blyth, F., Bolliger, I., Boufous, S., Bucello, C., Burch, M., Burney, P., Carapetis, J., Chen, H., Chou, D., Chugh, S.S., Coffeng, L.E., Colan, S.D., Colquhoun, S., Colson, K.E., Condon, J., Connor, M.D., Cooper, L.T., Corriere, M., Cortinovis, M., de Vaccaro, K.C., Couser, W., Cowie, B.C., Criqui, M.H., Cross, M., Dabhadkar, K.C., Dahodwala, N., De Leo, D., Degenhardt, L., Delossantos, A., Denenberg, J., Des Jarlais, D.C., Dharmaratne, S.D., Dorsey, E.R., Driscoll, T., Duber, H., Ebel, B., Erwin, P.J., Espindola, P., Ezzati, M., Feigin, V., Flaxman, A.D., Forouzanfar, M.H., Fowkes, F.G.R., Franklin, R., Fransen, M., Freeman, M.K., Gabriel, S.E., Gakidou, E., Gaspari, F., Gillum, R.F., Gonzalez-Medina, D., Halasa, Y.A., Haring, D., Harrison, J.E., Havmoeller, R., Hay, R.J., Hoen, B., Hotez, P.J., Hoy, D., Jacobsen, K.H., James, S.L., Jasrasaria, R., Jayaraman, S., Johns, N., Karthikeyan, G., Kassebaum, N., Keren, A., Khoo, J.-P., Knowlton, L.M., Kobusingye, O., Koranteng, A., Krishnamurthi, R., Lipnick, M., Lipshultz, S.E., Ohno, S.L., Mabweijano, J., MacIntyre, M.F., Mallinger, L., March, L., Marks, G.B., Marks, R., Matsumori, A., Matzopoulos, R., Mayosi, B.M., McAnulty, J.H., McDermott, M.M., McGrath, J., Mensah, G.A., Merriman, T.R., Michaud, C., Miller, M., Miller, T.R., Mock, C., Mocumbi, A.O., Mokdad, A.A., Moran, A., Mulholland, K., Nair, M.N., Naldi, L., Narayan, K.M.V., Nasseri, K., Norman, P., O'Donnell, M., Omer, S.B., Ortblad, K., Osborne, R., Ozgediz, D., Pahari, B., Pandian, J.D., Rivero, A.P., Padilla, R.P., Perez-Ruiz, F., Perico, N., Phillips, D., Pierce, K., Pope, C.A., Porrini, E., Pourmalek, F., Raju, M., Ranganathan, D., Rehm, J.T., Rein, D.B., Remuzzi, G., Rivara, F.P., Roberts, T., De León, F.R., Rosenfeld, L.C., Rushton, L., Sacco, R.L., Salomon, J.A., Sampson, U., Sanman, E., Schwebel, D.C., Segui-Gomez, M., Shepard, D.S., Singh, D., Singleton, J., Sliwa, K., Smith, E., Steer, A., Taylor, J.A., Thomas, B., Tleyjeh, I.M., Towbin, J.A., Truelsen, T., Undurraga, E.A., Venketasubramanian, N., Vijayakumar, L., Vos, T., Wagner, G.R., Wang, M., Wang, W., Watt, K., Weinstock, M.A., Weintraub, R., Wilkinson, J.D., Woolf, A.D., Wulf, S., Yeh, P.-H., Yip, P., Zabetian, A., Zheng, Z.-J., Lopez, A.D., Murray, C.J.L., AlMazroa, M.A., Memish, Z.A., 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a

systematic analysis for the Global Burden of Disease Study 2010. Lancet Lond. Engl. 380, 2095–2128. https://doi.org/10.1016/S0140-6736(12)61728-0
Lučić, V., Rigort, A., Baumeister, W., 2013. Cryo-electron tomography: The challenge of doing structural biology in situ. J. Cell Biol. 202, 407–419. https://doi.org/10.1083/jcb.201304193
Malhotra, S., Träger, S., Dal Peraro, M., Topf, M., 2019. Modelling structures in cryo-EM maps. Curr. Opin. Struct. Biol., Cryo electron microscopy ● Biophysical and computational methods ● Biophysical and computational methods - Part B 58, 105–114. https://doi.org/10.1016/j.sbi.2019.05.024
Malvy, D., Chappuis, F., 2011. Sleeping sickness. Clin. Microbiol. Infect. 17, 986–995. https://doi.org/10.1111/j.1469-0691.2011.03536.x
Marion, D., 2013. An introduction to biological NMR spectroscopy. Mol. Cell. Proteomics MCP 12, 3006–3025. https://doi.org/10.1074/mcp.O113.030239
Masters, B.R., 2009. History of the Electron Microscope in Cell Biology, in: John Wiley & Sons, Ltd (Ed.), Encyclopedia of Life Sciences. John Wiley & Sons, Ltd, Chichester, UK, p. a0021539. https://doi.org/10.1002/9780470015902.a0021539
Matera, A.G., Wang, Z., 2014. A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol. 15, 108–121. https://doi.org/10.1038/nrm3742
McMullan, G., Faruqi, A.R., Henderson, R., 2016. Direct Electron Detectors. Methods Enzymol. 579, 1–17. https://doi.org/10.1016/bs.mie.2016.05.056
Merrick, W.C., 2004. Cap-dependent and cap-independent translation in eukaryotic systems. Gene 332, 1–11. https://doi.org/10.1016/j.gene.2004.02.051
Nič, M., Jirát, J., Košata, B., Jenkins, A., McNaught, A. (Eds.), 2009. tertiary structure, in: IUPAC Compendium of Chemical Terminology. IUPAC, Research Triagle Park, NC. https://doi.org/10.1351/goldbook.T06282
Nick Pace, C., Martin Scholtz, J., 1998. A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. Biophys. J. 75, 422–427. https://doi.org/10.1016/S0006-3495(98)77529-0
Nitta, R., Imasaki, T., Nitta, E., 2018. Recent progress in structural biology: lessons from our research history. Microscopy 67, 187–195. https://doi.org/10.1093/jmicro/dfy022
Palade, G.E., 1955. A SMALL PARTICULATE COMPONENT OF THE CYTOPLASM. J. Biophys. Biochem. Cytol. 1, 59–68.
Patikoglou, G., Burley, S.K., 1997. Eukaryotic Transcription Factor-Dna Complexes. Annu. Rev. Biophys. Biomol. Struct. 26, 289–325. https://doi.org/10.1146/annurev.biophys.26.1.289
Perutz, M.F., Weisz, O., 1947. Crystal Structure of Human Carboxyhæmoglobin. Nature 160, 786–787. https://doi.org/10.1038/160786a0
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera?A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084
Pollitt, L.C., MacGregor, P., Matthews, K., Reece, S.E., 2011. Malaria and trypanosome transmission: different parasites, same rules? Trends Parasitol. 27, 197–203. https://doi.org/10.1016/j.pt.2011.01.004
Ramakrishnan, V., 2002. Ribosome Structure and the Mechanism of Translation. Cell 108, 557–572. https://doi.org/10.1016/S0092-8674(02)00619-0
Ramanathan, A., Robb, G.B., Chan, S.-H., 2016a. mRNA capping: biological functions and applications. Nucleic Acids Res. 44, 7511–7526. https://doi.org/10.1093/nar/gkw551
Ramanathan, A., Robb, G.B., Chan, S.-H., 2016b. mRNA capping: biological functions and applications. Nucleic Acids Res. 44, 7511–7526. https://doi.org/10.1093/nar/gkw551
Rassi, A., Rassi, A., Marin-Neto, J.A., 2010. Chagas disease. Lancet Lond. Engl. 375, 1388–1402. https://doi.org/10.1016/S0140-6736(10)60061-X
Rezende, A.M., Assis, L.A., Nunes, E.C., da Costa Lima, T.D., Marchini, F.K., Freire, E.R., Reis, C.R., 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. https://doi.org/10.1186/1471-2164-15-1175
Richardson, J.S., 1981. The Anatomy and Taxonomy of Protein Structure, in: Advances in Protein Chemistry. Elsevier, pp. 167–339. https://doi.org/10.1016/S0065-3233(08)60520-3
Rodnina, M.V., 2018. Translation in Prokaryotes. Cold Spring Harb. Perspect. Biol. 10, a032664. https://doi.org/10.1101/cshperspect.a032664
Rogers, G.W., Komar, A.A., Merrick, W.C., 2002. eIF4A: The godfather of the DEAD box helicases, in: Progress in Nucleic Acid Research and Molecular Biology. Elsevier, pp. 307–331. https://doi.org/10.1016/S0079-6603(02)72073-4
Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., Yonath, A., 2000. Structure of Functionally Activated Small Ribosomal Subunit at 3.3 Å Resolution. Cell 102, 615–623. https://doi.org/10.1016/S0092-8674(00)00084-2
Schmeing, T.M., Ramakrishnan, V., 2009. What recent ribosome structures have revealed about the mechanism of translation. Nature 461, 1234–1242. https://doi.org/10.1038/nature08403
Shatsky, I.N., Terenin, I.M., Smirnova, V.V., Andreev, D.E., 2018. Cap-Independent Translation: What's in a Name? Trends Biochem. Sci. 43, 882–895. https://doi.org/10.1016/j.tibs.2018.04.011
Simarro, P., Franco, J., Diarra, A., Jannin, J., 2014. Epidemiology of human African trypanosomiasis. Clin. Epidemiol. 257. https://doi.org/10.2147/CLEP.S39728
Simossis, V.A., Heringa, J., 2004. Optimally segmented consensus secondary structure prediction. Bioinformatics.
Sirri, V., Urcuqui-Inchima, S., Roussel, P., Hernandez-Verdun, D., 2008. Nucleolus: the fascinating nuclear body. Histochem. Cell Biol. 129, 13–31. https://doi.org/10.1007/s00418-007-0359-6
Sonenberg, N., Gingras, A.-C., 1998. The mRNA 5′ cap-binding protein eIF4E and control of cell growth. Curr. Opin. Cell Biol. 10, 268–275. https://doi.org/10.1016/S0955-0674(98)80150-6
Srivastava, S., Verschoor, A., Radermacher, M., Grassucci, R., Frank, J., 1995. Three-Dimensional Reconstruction of Mammalian 40 S Ribosomal Subunit Embedded in Ice. J. Mol. Biol. 245, 461–6. https://doi.org/10.1006/jmbi.1994.0037
Steitz, T.A., 2008. A structural understanding of the dynamic ribosome machine. Nat. Rev. Mol. Cell Biol. 9, 242–253. https://doi.org/10.1038/nrm2352
Tanaka, Y., Fujii, S., Hiroaki, H., Sakata, T., Tanaka, T., Uesugi, S., Tomita, K., Kyogoku, Y., 1999. A'-form RNA double helix in the single crystal structure of r(UGAGCUUCGGCUC). Nucleic Acids Res. 27, 949–955. https://doi.org/10.1093/nar/27.4.949
Taylor, K.A., Glaeser, R.M., 2008. Retrospective on the early development of cryoelectron microscopy of macromolecules and a prospective on opportunities for the future. J. Struct. Biol. 163, 214–223. https://doi.org/10.1016/j.jsb.2008.06.004
Taylor, K.A., Glaeser, R.M., 1974. Electron diffraction of frozen, hydrated protein crystals. Science 186, 1036–1037. https://doi.org/10.1126/science.186.4168.1036
Teixeira, A.R., Monteiro, P.S., Rebelo, J.M., Argañaraz, E.R., Vieira, D., Lauria-Pires, L., Nascimento, R., Vexenat, C.A., Silva, A.R., Ault, S.K., Costa, J.M., 2001. Emerging Chagas disease: trophic network and cycle of transmission of Trypanosoma cruzi from palm trees in the Amazon. Emerg. Infect. Dis. 7, 100–112.
Trabuco, L.G., Villa, E., Mitra, K., Frank, J., Schulten, K., 2008. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Struct. Lond. Engl. 1993 16, 673–683. https://doi.org/10.1016/j.str.2008.03.005
Unwin, P.N.T., Henderson, R., 1975. Molecular structure determination by electron microscopy of unstained crystalline specimens. J. Mol. Biol. 94, 425–440. https://doi.org/10.1016/0022-2836(75)90212-0
Valásek, L.S., 2012. 'Ribozoomin'--translation initiation from the perspective of the ribosome-bound eukaryotic initiation factors (eIFs). Curr. Protein Pept. Sci. 13, 305–330. https://doi.org/10.2174/138920312801619385
Valášek, L.S., Zeman, J., Wagner, S., Beznosková, P., Pavlíková, Z., Mohammad, M.P., Hronová,

V., Herrmannová, A., Hashem, Y., 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. https://doi.org/10.1093/nar/gkx805

Vargas-Parada, L., 2010. Kinetoplastids and Their Networks of Interlocked DNA. Nat. Educ. 3(9):63.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303. https://doi.org/10.1093/nar/gky427

Will, S., Joshi, T., Hofacker, I.L., Stadler, P.F., Backofen, R., 2012. LocARNA-P: Accurate boundary prediction and improved detection of structural RNAs. RNA 18, 900–914. https://doi.org/10.1261/rna.029041.111

Wilson, D.N., 2014. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. Nat. Rev. Microbiol. 12, 35–48. https://doi.org/10.1038/nrmicro3155

Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., Ramakrishnan, V., 2000. Structure of the 30S ribosomal subunit. Nature 407, 327–339. https://doi.org/10.1038/35030006

World Health Organization, 2012. Research priorities for Chagas disease, human African trypanosomiasis and leishmaniasis. World Health Organ. Tech. Rep. Ser. v–xii, 1–100.

Wu, S., Skolnick, J., Zhang, Y., 2007. Ab initio modeling of small proteins by iterative TASSER simulations. BMC Biol. 5, 17. https://doi.org/10.1186/1741-7007-5-17

Wüthrich, K., 2001. The way to NMR structures of proteins. Nat. Struct. Biol. 8, 923–925. https://doi.org/10.1038/nsb1101-923

Yonath, A., Mussig, J., Tesche, B., Lorenz, S., Erdmann, V.A., Wittman, H.G., 1980. Crystallization of the large ribosomal subunits from Bacillus stearothermophilus. Biochem Int 1, 428–435.

Yusupov, M.M., 2001. Crystal Structure of the Ribosome at 5.5 A Resolution. Science 292, 883–896. https://doi.org/10.1126/science.1060089

Yusupova, G., Yusupov, M., 2017. Crystal structure of eukaryotic ribosome and its complexes with inhibitors. Philos. Trans. R. Soc. B Biol. Sci. 372. https://doi.org/10.1098/rstb.2016.0184

Zerba, E.N., 1999. SUSCEPTIBILITY AND RESISTANCE TO INSECTICIDES OF CHAGAS DISEASE VECTORS 6.

Zhang, H., Endrizzi, J.A., Shu, Y., Haque, F., Sauter, C., Shlyakhtenko, L.S., Lyubchenko, Y., Guo, P., Chi, Y.-I., 2013. Crystal structure of 3WJ core revealing divalent ion-promoted thermostability and assembly of the Phi29 hexameric motor pRNA. RNA N. Y. N 19, 1226–1237. https://doi.org/10.1261/rna.037077.112

Zhao, J., Hyman, L., Moore, C., 1999. Formation of mRNA 3′ Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in mRNA Synthesis. Microbiol. Mol. Biol. Rev. 63, 405–445.

Université de Strasbourg

Anthony BOCHLER

Structural differences and specificities of translation initiation complexes between trypanosomes and their hosts

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

ABSTRACT

Kinetoplastids is a group of flagellated protozoans including parasites causing worldwide spread disease, threatening millions of people. Because they are eukaryotes, like their mammalian hosts, their molecular complexes of translation are highly conserved with their hosts. This conservation massively slowdown the development of new drugs targeting ribosomes which is a well-known therapeutic target. Indeed, actual drugs are often toxic and induce negative secondary effects on the host. Despite the conservation, their protein biosynthesis pathway contains specificities. The structural studies by cryo-EM of their translation complexes could lead to the development of safer antiparasitic drugs by discovering new potential therapeutic targets specific to kinetoplastids. During our projects, we discovered a novel kinetoplastid-specific ribosomal protein (KSRP). Moreover, we discovered a new helicase named DDX60-like that might play an important role during translation initiation. We also depicted the potential role of expansion segments ES6s and ES7s as a docking platform for eIF3 as well as proposing the first kinetoplastids eIF3 model from *T.cruzi.* All these freshly modeled structures and discoveries are or will be the starting point of several biochemical studies to know more about the functions of these factors.

RÉSUMÉ

Les kinétoplastides sont un groupe de protozoaires flagellés dont quelques espèces parasites exclusives causent des maladies menaçant plusieurs millions de personnes à travers le monde. Etant des espèces eucaryotes parasitant d'autres espèces eucaryotes comme les mammifères, il existe une forte conservation de leurs complexes moléculaires de la traduction. Cela freine énormément l'avancée du développement des traitements antibiotiques ciblant ces complexes car les antibiotiques vont également cibler ceux de l'hôte, entrainant des effets secondaires néfastes. Malgré cette conservation, leur mécanisme de biosynthèse des protéines comporte des spécificités uniques comparé à leurs hôtes. L'étude au niveau structurale de leurs complexes d'initiation de la traduction par cryo-microscopie électronique pourrait permettre de mettre en avant ces spécificités qui pourront être utilisées pour des études plus poussées telles que le développement d'antibiotique ciblant spécifiquement les parasites en évitant les effets secondaires néfastes des traitements actuels. Durant nos travaux, nous avons découvert une nouvelle protéine ribosomale spécifique aux kinétoplastides (KSRP) ainsi qu'une hélicase nommée DDX60-like ayant un rôle dans l'initiation de la traduction. Nous avons également mis en avant le potentiel rôle des segments d'expansion ES6s et ES7s comme plateforme d'accommodation pour le facteur d'initiation 3 (eIF3) tout en proposant pour la première fois un modèle pour eIF3 chez *T.cruzi*. Toutes ces nouvelles structures représentent le point de départ à de nouveaux projets pour connaître leurs fonctions grâce à des études biochimiques.