HAL
open science

# Structural study of mRNA translation in kinetoplastids by Cryo-electron microscopy

Jailson Fernando Brito Querido

## To cite this version:

Jailson Fernando Brito Querido. Structural study of mRNA translation in kinetoplastids by Cryo-electron microscopy. Genomics [q-bio.GN]. Université de Strasbourg, 2017. English. NNT : 2017STRAJ108 . tel-02918027

**HAL Id: tel-02918027**

**https://theses.hal.science/tel-02918027v1**

Submitted on 20 Aug 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE STRASBOURG

# UNIVERSITÉ DE STRASBOURG

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

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

## THÈSE

présentée par:

**Jailson Fernando Brito Querido**

soutenue le : 11 Décembre 2017

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 study of mRNA translation in kinetoplastids by Cryo-electron microscopy

## Etude structurale de la traduction de ARNm chez les kinétoplastides par cryomicroscopie électronique

**Thèse dirigée par :**

| | |
|---|---|
| **M. Yaser Hashem** | Chargé de recherche, Université de Strasbourg |

**Rappoteurs :**

| | |
|---|---|
| **Mme. Christine CLAYTON** | Professeur, Université de Heidelberg |
| **M. Axel INNIS** | Chargé de recherche, Université de Bordeaux |

**Autres membres du jury :**

| | |
|---|---|
| **M. Eric WESTHOF** | Professeur, Université de Strasbourg |
| **Mme. Magali FRUGIER** | Directrice de recherche, Université de Strasbourg |
| **M. Philippe BASTIN** | Directeur de recherche, Institut Pasteur |

**To: Diogo and Rita**

One day someone told me that to do science, you need "more than science". Each of us need to find what "more than science" means. My "more than science" is my source of inspiration, is what makes me want to be better every single day. How gratefully one should be if you feel that someone is interrupting her own path and dreams just to allow you to reach yours?

*Se um dia alguém*
*Perguntar por mim*
*Diz que vivi*
*Para te amar*

*Antes de ti*
*Só existi*
*Cansado e sem nada p'ra dar*

By Luísa Sobral

*"Os bons e os maus resultados dos nossos ditos e obras vão-se distribuindo, supõe-se que de uma maneira bastante uniforme e equilibrada, por todos os dias do futuro, incluindo aqueles, infindáveis, em que já cá não estaremos para poder comprová-lo, para congratularmo-nos ou para pedir perdão, aliás, há quem diga que é isto a imortalidade de que tanto se fala."*

José Saramago (1922 – 2010)
Ensaio sobre a cegueira (1995, p. 84)

## ACKNOWLEDGEMENT

This thesis described the work I have been doing during the last three years. Furthermore, it also described the knowledge I have been building since 2009, under the supervision of several people.

First, I would like to deeply acknowledge Yaser for the opportunity, for his support and guidance. It was an honor for me to have been his first PhD student.

I would like to thank every single people in our lab, starting from Angelita who introduced me to the ribosome purification and for all the discussions we use to have. Eder Mancera, who worked with me during the characterization of KSRP. Quentin for all the guidance. Anthony who has been introducing me to the RNA and protein modelling. Margarita for all the discussion and scientific exchange.

I would like to thank former members of our lab, including Claire and Adeline for all the support they have given me.

I would like to thank the MS/MS platform, including Johana, Lauriane and Philippe. Without their commitment and effort this work would not be possible.

To Stefano for the gradient fractionator machine and for all scientific discussion we used to have.

Franck Martin and Claude Sauter for allowed me to use their instruments for the purification and characterization of the complexes.

Danièle Werling, Léandra Hernandez and Pascale Romby for the incredible support they gave me over the last 3 years.

My deeply acknowledgement to Eric Westhof for his support and guidance.

All the jury members, including Axel Innis, Christine Clayton, Eric Westhof, Magali Frugier and Philippe Bastin for have accepted to take part in the analysis of this manuscript.

My deeply acknowledgement to Diego Guerin and Marcelo Silva. They made me believe that I could do science. To Aline for the parasites.

Last but not the least, my special thanks to my family and friends for their support.

# RESUMÉ


**Etude structurale de la traduction de l'ARNm dans les kinétoplastides par cryo-microscopie électronique**

**Initiation à la traduction chez les eucaryotes**

Comme chez les procaryotes, la traduction chez les eucaryotes comprend plusieurs étapes, telles que l'initiation, l'élongation, la terminaison et le recyclage.

L'initiation de la traduction chez les eucaryotes est un processus dynamique, cyclique et complexe, régulé pour plus de 20 protéines différentes (revu dans Aitken et Lorsch, 2012). Ce processus peut être subdivisé en deux grands groupes, dépendants de la coiffe (cap) et indépendants de la coiffe.

L'initiation dépendant du cap est le processus plus commun par lequel l'ARNm est traduit chez les eucaryotes, c'est pourquoi on l'appelle aussi le processus d'initiation canonique de la traduction. La première étape principale de l'initiation de la traduction est l'assemblage du complexe de pré-initiation 43S (43S PIC). Ce processus commence avec la formation d'un complexe ternaire (eIF2-TC), composé d'un ARNt initiateur (Met-ARNt$_i^{Met}$) et d'un GTP (Guanosine tri-phosphates) lié au facteur d'initiation eucaryote 2 (eIF2). Une fois assemblé, eIF2-TC est recruté dans la sous-unité ribosomique 40S. Le recrutement d'eIF2-TC est également renforcé par d'autres facteurs, notamment eIF1, eIF1A, eIF3 et eIF5, qui aboutissent ensemble à la formation du 43S PIC (Pestova et Kolupaeva, 2002, de Georges et al., 2015, Hashem et al. 2013b, Hussain et al., 2014a, Llacer et al., 2015a). ABCE1 (Rli1 dans la levure), un facteur bien caractérisé de recyclage ribosomique, a également été récemment observé impliqué dans l'assemblage du 43S PIC (Dong et al., 2004a, Mancera-Martínez et al., 2017, Heuer et al., 2017).

L'étape suivante de l'initiation de la traduction est l'assemblage du complexe d'initiation 48S (48S IC). Une fois assemblé, le 43S PIC est recruté dans l'ARNm, ce qui permet la reconnaissance base par base du messager jusqu'au codon d'initiation AUG, un processus mieux connu sous le nom de « scanning ». Le scanning inclut le mouvement du messager dans le complexe de préinitiation 43S dans un sens unique allant de l'extrémité 3' vers l'extrémité 5', et ce faisant permettre l'appariement de bases entre le codon dans l'ARNm et l'anticodon contenu dans sa tige-boucle (ASL) du Met-ARNt$_i^{Met}$. Il a été montré que le 43S PIC est capable

de se lier à l'ARNm par sa région 5' non traduite (5'UTR). Les 5'UTRs peuvent être non structurées et dans ce cas le recrutement du 43S PIC se fait directement t(Pestova et Kolupaeva, 2002), cependant les ARNm naturels possèdent bien souvent différents degrés de structure secondaire à leurs 5'UTRs, incluant des boucles et des hélices (Kozak, 1999). Par conséquent, le recrutement du 43S PIC à l'ARNm nécessite l'activité d'un complexe composé de facteurs multiples appelé le complexe de liaison à la coiffe, ou communément en Anglais « cap-binding complex (CBC) », formé par eIF4F, eIF4B et eIF4A.

eIF4F à son tour est constitué de la protéine de liaison à la coiffe (eIF4E), d'une deuxième copie de l'hélicase à ARN de type DEAD (eIF4A) et de la protéine d'échafaudage eIF4G (revue dans Jackson et al., 2010). L'activité hélicase de eIF4A est renforcée par l'activité de eIF4B qui se lie à l'ARNm simple-brin et empêche son repliement en structures secondaires (Marintchev et al., 2009). Le complexe multifactoriel eIF4F/4B/4A est capable de dérouler l'ARNm 5' UTR et de favoriser sa fixation sur le 43S PIC via l'interaction eIF4G-eIF3. Néanmoins, les ARNm ayant un degré élevé de structures secondaires nécessitent d'autres hélicases à boîte DEAD en plus de eIF4A, comme la protéine DEAH boîte DHX29 (Pisareva et al., 2008; Parsyan et al., 2009).

Après le scanning, les facteurs d'initiation sont libérés, ce qui permet l'association du 60S et la formation du complexe d'initiation 80S.

**Perspectives structurales et fonctionnelles du ribosome des kinetoplastides**

Le kinétoplastide est un groupe de protozoaires flagellés ayant une importance médicale et économique. Ils présentent un élargissement d'une région sur la mitochondrie appelée kinétoplaste, d'où leur nom. De plus, leurs ribosomes présentent des différences significatives comparés aux ribosomes de leurs hôtes mammaliens.

Le ribosome eucaryotique est composé de deux sous-unités, une 40S/SSU et une 60S/LSU. La petite sous-unité contient l'ARNr 18S, tandis que la grande sous-unité contient les ARNr 28S, 5.8S et 5S (revue dans Mandal, 1984). Les 18S, 5.8S et 28S sont codés dans une seule unité de transcription, connue comme l'unité de transcription principale. Le gène de l'ARNr 5S n'est pas lié à l'unité de transcription principale. En effet, le gène 5S est transcrit par l'ARN polymérase III, tandis que le reste est transcrit par l'ARN pol I.

L'analyse de l'ARNr des kinetoplastides a démontré le caractère unique de leurs ribosomes. Contrairement à leurs hôtes mammifères, l'homologue kinétoplastidien de l'ARNr 28S eucaryotique présente de multiples sites de clivage (Campbell et al., 1987). En effet, l'ARNr de la LSU des trypanosomes est composée de LSU-α, LSU-β, de quatre petits ARNr (srRNA1,

srRNA2, srRNA3 et srRNA4), d'ARNr 5S et 5.8S (Campbell et al., 1987, Hashem et al., 2013a, Liu et al., 2016a). De plus, il a été démontré qu'ils possèdent un nombre important de sites de modification de l'ARNr (Eliaz et al., 2015).

Les avancées technologiques récentes de la cryo-microscopie électronique (cryo-ME) a conduit à l'obtention de la structure du ribosome des kinétoplastides à une résolution aussi élevée que 2,5 Å. La structure à haute résolution de ces ribosomes a mis en évidence la présence des larges segments d'expansion (SEs), ainsi que l'extension des protéines ribosomiques (protéines r) (Hashem et al., 2013a, Liu et al., 2016a, Zhang et al., 2016a).

Le rôle de leurs plus grandes SEs a été discuté dans plusieurs publications (Hashem et al., 2013a, Zhang et al., 2016). Ils ont été associés à l'initiation de la traduction et à la stabilité du ribosome 80S. En effet, la structure cryo-ME du ribosome 80S de *T. brucei* a révélé qu'ils sont impliqués dans la formation de quatre ponts additionnels antre les deux sous-unités ribosomiques, qui semblent être spécifiques aux kinétoplastides (Hashem et al., 2013a).

Un autre aspect important des kinétoplastides dans leur 40S est l'emplacement de leurs grands segments d'expansion $SE6^S$ et $SE7^S$ près de la sortie du canal de l'ARNm. Il a été proposé qu'un tel emplacement puisse leur permettre de jouer un rôle dans l'initiation de la traduction en interagissant avec eIF3 (Hashem et al., 2013a).

En résumé, les 40S des kinétoplastides possèdent certains éléments spécifiques qui suggèrent que l'initiation de la traduction dans ces organismes peut avoir plusieurs divergences par rapport à celle observée chez leur hôte mammifère.

**La biologie des kinétoplastides et les maladies transmises par des vecteurs humains**

Depuis plus d'un siècle, les kinétoplastides attirent l'attention de la communauté scientifique et médicale en raison de leur importance en matière de santé publique. Néanmoins, l'Organisation mondiale de la santé (OMS) a toujours classé toutes les maladies humaines causées par les kinétoplastides dans un groupe de maladies tropicales négligées.

*Trypanosoma cruzi* (*T. cruzi*), *Trypanosoma brucei spp*. et *Leishmania spp*. sont des kinétoplastides responsables de plusieurs maladies à transmission vectorielle, qui menacent ensemble plus de 400 millions de personnes dans le monde (Organisation mondiale de la Santé, 2012).

Le mécanisme d'expression des gènes chez les kinétoplastides est nettement différent de leurs hôtes mammifères. En outre, compte tenu de leur déficience de la régulation de la transcription, toutes les étapes post-transcriptionnelles, incluant la traduction, ont une importance accrue pour la régulation de l'expression génique et l'homéostasie cellulaire.

Néanmoins, on sait peu de choses sur l'initiation de la traduction dans les kinétoplastides. Les connaissances actuelles soulignent la faible identité des eIFs des kinétoplastides par rapport à leurs hôtes mammifères (Meleppattu et al., 2015, Rezende et al., 2014, Li et al., 2017), ce qui met en évidence la traduction en tant que domaine prioritaire à étudier dans ces organismes. Par conséquent, l'objectif principal de cette thèse est de comprendre l'architecture du complexe d'initiation de la traduction dans les kinétoplastides, ce qui peut nous permettre de comprendre certains aspects spécifiques de ces parasites quant à la régulation de la traduction de l'ARNm. Ces eucaryotes précoces ont une biologie et une séquence génomique similaires (El-Sayed et al., 2005b). En outre, ils présentent certaines caractéristiques uniques liées au mécanisme d'expression génique. Ces caractéristiques communes représentent une opportunité de traiter ces maladies en utilisant une nouvelle famille de médicaments qui pourraient être plus efficaces et bien moins toxiques que les peu d'alternatives actuelles (Khare et al., 2016).

**L'étude structurale des complexes 40S de Trypanosoma cruzi révèle l'existence d'une voie de régulation de l'initiation de la traduction spécifique aux kinétoplastides**

Bien que de nombreuses études aient mis en évidence la divergence évolutive de la traduction de l'ARNm dans les kinétoplastides, on sait peu de choses sur ses bases structurales. Nous présentons ici des reconstructions de cryo-ME de complexes 40S natifs et à l'arrêt, purifiés à partir de Trypanosoma cruzi, le parasite kinétoplastidique responsable de la maladie de Chagas.

Nos structures montrent que le facteur de recyclage des ribosomes ABCE1 interagit avec le site conservé de liaison aux GTPases ribosomiques sur la face interne de la sous-unités 40S, et adopte une conformation qui diffère de celle observée pendant le recyclage ribosomique. Plus important encore, nous découvrons un facteur non caractérisé à la plate-forme des complexes 40S, appelé provisoirement ηF. De plus, notre étude a dévoilé une structure de dimères 40S, éventuellement spécifiques aux kinétoplastidés. Notamment, ηF se chevauche les contacts impliqués dans la formation des dimères de 40S, suggérant un rôle dans la prévention de la dimérisation 40S. Enfin, nos structures montrent que ηF inhibe le formation des complexe eIF2-ternaires, indiquant que ηF doit être libéré avant l'assemblage du complexe de (pré)initiation. Notre travail représente un premier aperçu structural des aspects spécifiques aux kinétoplastide de la régulation de la traduction.

**La structure par cryo-ME d'une nouvelle protéine ribosomique spécifique du kinétoplastide 40S**

Il existe un besoin critique de cibles plus spécifiques pour le développement de molécules thérapeutiques anti-kinétoplastidiques, plus sûres qui peuvent remplacer les médicaments actuels rares et hautement cytotoxiques. Le ribosome kinétoplastidique représente une cible thérapeutique potentielle en raison de sa divergence structurale par rapport à son homologue humain. Cependant, plusieurs caractéristiques ribosomiques spécifiques aux kinétoplastidiens restent non caractérisées. Nous présentons ici la structure par cryo-ME à résolution quasi-atomique d'une nouvelle protéine ribosomique kinétoplastidique spécifique (KSRP) liée à la petite unité ribosomique 40S. KSRP est une protéine essentielle située à la face du solvant de la sous-unité 40S, où elle se lie et stabilise les domaines spécifiques aux kinetoplastides faisant partie de l'ARNr, suggérant son rôle dans l'intégrité de la structure du ribosome. KSRP interagit également avec la protéine-r eS6 à une région qui est seulement conservée dans les kinétoplastides. L'environnement ribosomique spécifique des kinétoplastidés autour de KSRP constitue une cible prometteuse pour la conception de médicaments anti-kinétoplastidiens plus sûrs et plus efficaces.

**La structure par Cryo-ME du complexe natif d'initiation de la traduction chez Trypanosoma cruzi**

Les kinétoplastides possèdent des caractéristiques inhabituelles, telles que la présence d'une coiffe conservée d'épissage dans tous les ARNm matures, appelée en Anglais « Spliced-Leader ». De plus, les segments d'expansions (SE) ribosomiques de l'ARN 40S, tels que $SE3^S$, $SE6^S$, $SE7^S$ et $SE9^S$ sont plus grands que ceux de leurs équivalents hôtes mammifères. Ici, nous avons purifié à partir de lysats cellulaires de T. cruzi des complexes natifs de l'initiation de la traduction que nous avons ensuite analysés par cryo-ME.

La structure des complexes natifs de l'initiation révèle certains aspects spécifiques de l'initiation de la traduction, un tel modèle d'accommodation du noyau structural d'eIF3 dans le 48S, grâce à l'interaction d'eIF3a, eIF3c et eIF3e avec $SE6^S$ et $SE7^S$. Notre structure révèle également que eIF3c est pris en étau par les hélices A et B du $SE7^S$ ($SE7^S$-hA et $SE7^S$-hB). En outre, $SE9^S$ se lie et stabilise la sous-unité d d'eIF3. Nos analyses structurale et protéomique révèlent que le la protéine DDX60 chez les kinétoplastids peut jouer un rôle important dans l'initiation de la traduction en déroulant les ARNm coiffés avec le Spliced-Leader. Les particularités révélées ici dans cette première structure de complexe d'initiation du parasite

protozoaire ouvre de nouvelles perspectives pour le développement d'agents thérapeutiques contre les kinétoplastides ciblant l'étape de la traduction, et plus particulièrement son initiation.

### Conclusions

La caractérisation de la sous-unité native du 40S nous a permis de découvrir une nouvelle protéine ribosomique spécifique des kinétoplastides, appelée ici KSRP. KSRP se lie au pied gauche du 40S où il interagit avec SE3$^{S}$ et SE6$^{S}$. En outre, KSRP interagit également avec l'extension spécifique kinétoplastidique de la protéine ribosomique eS6. Le site de liaison des deux domaines de motifs de reconnaissance d'ARN (RRM) de KSRP suggère un rôle en tant que chaperonne de l'ARNr ainsi qu'un échafaudage pour la liaison d'autres protéines régulatrices. La découverte d'une nouvelle protéine ribosomqie essentielle à la viabilité des kinétoplastides a introduit une nouvelle cible potentielle pour le développement d'agents thérapeutiques anti-kinétoplastidiens plus sûrs et plus spécifiques qui bloquent la traduction de l'ARNm chez ce pathogène.

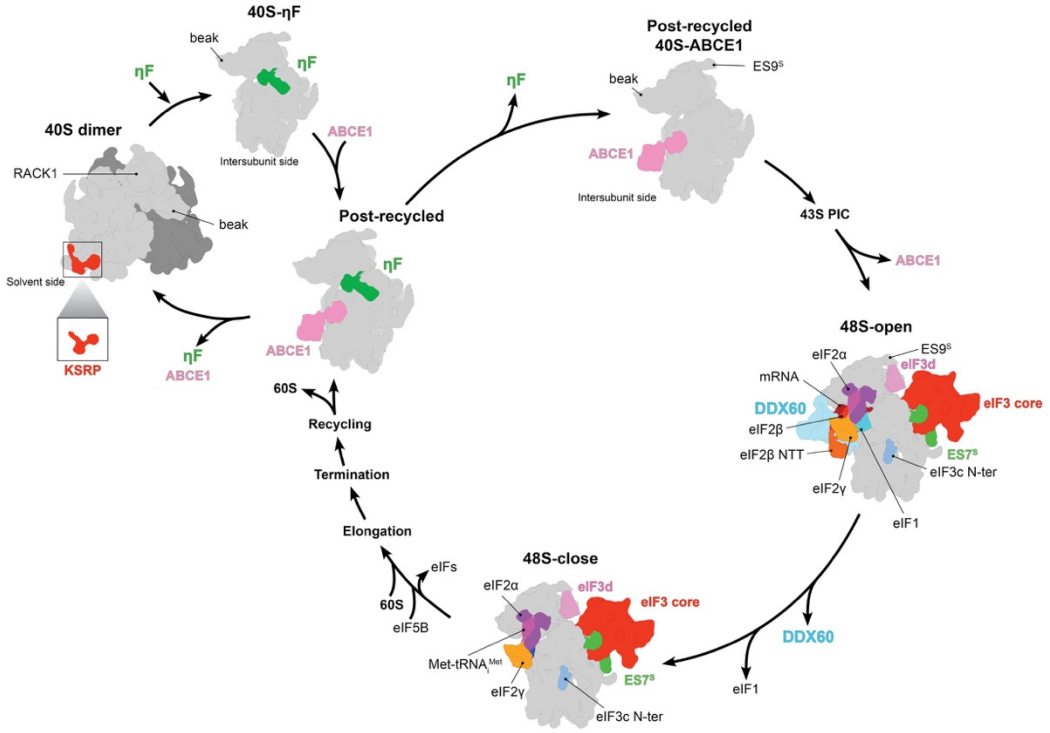


**Figure I: Modèle schématique de la traduction de l'ARNm dans les kinétoplastides.** La protéine-r KSRP se lie au pied gauche de 40S. Nous proposons que le facteur spécifique des kinétoplastides, ηF, empêche la dimérisation des 40S, ce qui rend les sous-unités 40S disponibles pour la traduction. Cependant, ηF doit être libéré car sa liaison empêche le recrutement de l'ARNm et de l'eIF2-TC. ABCE1 et ηF agissent comme des facteurs d'anti-association ribosomiqueq critiques. Après l'assemblage du 43S, ABCE1 doit être libéré, car le domaine de cluster de fer se heurte au site de fixation de DDX60. Enfin, DDX60 lui-même doit également être libéré car sa présence empêche la formation de la conformation fermée du complexe 48S.

Nos reconstructions par cryo-ME des complexes 40S natifs et liés à ABCE1 à partir de lysats de T. cruzi nous permettent d'identifier un nouveau facteur lié à la plateforme de 40S proche du site E, établissant des interactions protéine/ARNr avec les hélices 24 et 23 du 18S. Ce facteur, appelé ici ηF, se connecte également au canal de l'ARNm. Le site de liaison de ηF se heurte à eIF2-TC ainsi qu'à l'ARNm. Par conséquent, ηF empêche l'assemblage de 43S PIC (figure I). Sur la base de cela, nous proposons ici que ηF est un facteur de régulation de la traduction spécifique aux kinétoplastides qui doit être libéré avant l'assemblage du complexe d'initiation. Bien que nos données structurales aient suggéré un rôle pour ηF dans le contrôle de la traduction, nous ne disposons pas de données expliquant le mécanisme moléculaire par lequel ηF est libéré du ribosome. Par conséquent, plus d'études biochimiques visant à mieux caractériser ηF ainsi que son profil d'expression au cours de différents stades du cycle de vie du parasite seront obligatoires pour élucider son mécanisme dans le contrôle de la traduction.

De plus, nous avons montré ici que la liaison d'ABCE1 au 40S est entièrement compatible avec l'assemblage du PIC 43S. En effet, il est en conformité avec notre récente étude chez les mammifères (Mancera-Martínez et al., 2017), où nous avons présenté la structure d'ABCE1 dans le contexte d'un complexe d'initiation de stade tardif chez les mammifères. Cette structure nous a permis de conclure que ABCE1 agit comme un facteur d'anti-association empêchant l'assemblage prématuré de la sous-unité large 60S (figure I). Une autre étude récente a également démontré que la conformation du domaine de cluster de fer d'ABCE1 dans le contexte du complexe de post-recyclage empêche la formation du pont ribosomique formé entre uL14 et h44 (Heuer et al., 2017), entre les deux sous-unités.

Enfin, la structure du complexe de pré-initiation 48S quasi complet nous permet d'identifier le schéma de liaison de la majorité des eIFs (figure I), qui révèle l'architecture du noyau structural d'eIF3 dans le contexte du complexe d'initiation kinétoplastidien. Les segments d'expansion de taille inhabituellement grande dans les kinétoplastides jouent un rôle important dans l'initiation de la traduction en agissant comme un socle pour accueillir le noyau structural d'eIF3 et en liant et stabilisant la sous-unité périphérique d d'eIF3. De plus, nos analyses structurale et protéomique révèlent un rôle insoupçonné pour DDX60 dans l'initiation de la traduction, probablement en liant et en déroulant l'ARNm coiffée du Spliced-Leader. Le site de liaison de DDX60 au 43S PIC n'est pas compatible avec la liaison de ABCE1. Par conséquent, contrairement aux mammifères où il a été constaté que ABCE1 est présent durant l'initiation même à un stade tardif du processus (Mancera-Martínez et al., 2017), dans les kinétoplastides ABCE1 doit être relâché pour permettre la liaison de DDX60 (Figure I). Bien que notre structure actuelle du 48S IC présente un grand nombre d'aspects de traduction

spécifiques aux kinétoplastides, une résolution améliorée combinée à des approches biochimiques, telle le profilage des ribosomes sont encore nécessaires pour une meilleure interprétation de nos reconstructions.

En conclusion, nos analyses structurale et biochimique mettent en évidence la singularité de la machinerie traductionnelle ainsi que le mécanisme d'initiation de la traduction chez les pathogènes kinétoplastides humains.

## LIST OF FIGURES

**Figure 1:** Structure and some conserved features of eukaryotic ribosome. ..... 19
**Figure 2:** Structure 40-eIF1A-eIF1 in context of yeast preinitiation complex. ..... 22
**Figure 3:** Structure of eIF2-TC in context of initiation complex..... 24
**Figure 4:** Cryo-EM structure of mammalian eIF3 octamer core. ..... 26
**Figure 5:** Structure of mammalian eIF3 in context of 43S pre-initiation complex..... 27
**Figure 6:** Schematic representation of structural domains of eIF4G. ..... 28
**Figure 7:** Structure of DHX29 in context of mammalian 43S pre-initiation complex. ..... 30
**Figure 8:** Schematic representation of the last events in eukaryotic translation initiation. .... 31
**Figure 9:** General organization of the ribosomal main transcription unit..... 32
**Figure 10:** Cryo-EM reconstruction of *T. brucei* ribosome, highlighting some kinetoplastid-specific elements that do not exist in other ribosomes. ..... 33
**Figure 11:** The assembly model for kinetoplastids 60S rRNAs. ..... 34
**Figure 12:** Structure *T. brucei* 40S subunit..... 36
**Figure 13:** Schematic representation of RNA pol II transcription and mRNA maturation process..... 37
**Figure 14:** Chemical structure of kinetoplastids cap-4. ..... 38
**Figure 15:** Schematic representation of subcellular structure and main cellular forms of kinetoplastids. ..... 40
**Figure 16:** Life cycle of *T. cruzi*, from insect to mammalian host. ..... 42
**Figure 17:** Life cycle of *T. brucei*, highlighting the different stage development..... 43
**Figure 18:** The life cycle of Leishmania, from insect vector to human host. ..... 45
**Figure 19:** Schematic model of mRNA translation in kinetoplastids. ..... 138

# CONTENTS

**ACKNOWLEDGEMENT** ..... **5**

**RESUMÉ** ..... **7**

**LIST OF FIGURES** ..... **15**

**INTRODUCTION** ..... **18**

**Structural and functional insights on the eukaryotic ribosome** ..... **18**

**Translation initiation in eukaryotes** ..... **21**

The role of eIF1 and eIF1A during translation initiation ..... 22

eIF2-Ternary complex and the codon-anticodon base pairing ..... 23

Structure of eIF3 and its interactions with 40S subunit ..... 25

The role of RNA helicases during translation initiation ..... 28

**80S initiation complex formation** ..... **30**

**Structure of the kinetoplastids ribosome** ..... **31**

Kinetoplastid unusual 60S puzzle ..... 33

Structure and function of the 40S expansion segments ..... 35

**The unusual RNA polymerase II from kinetoplastids** ..... **36**

**Kinetoplastids biology and human vector borne diseases** ..... **39**

*Trypanosoma cruzi (T. cruzi)* ..... 41

*Trypanosoma brucei spp. (T. brucei spp.)* ..... 43

*Leishmania spp.* ..... 44

**RESULTS** ..... **47**

**Article 1: Structural study of 40S complexes from *Trypanosoma cruzi* reveals the existence of a kinetoplastid-specific translation initiation regulation pathway** ..... **47**

**Article 2: The cryo-EM structure of a novel 40S kinetoplastid-specific ribosomal protein** ..... **85**

**Article 3: Cryo-EM structure of native translation initiation complex from human pathogen *Trypanosoma cruzi*** ..... **107**

**CONCLUSION AND PERSPECTIVES** ..... **138**

**KEY RECOURSES TABLE** ..... **141**

**REFERENCE** ..... **142**

**ABSTRACT** ..... **161**

## LIST OF ABBREVIATIONS

| Abbreviation | Term |
|---|---|
| eIF | Eukaryotic initiation factor |
| ORF | Open reading frame |
| UTR | Untranslated region |
| ABCE1 | ATP-binding cassette subfamily E member 1 |
| Rli1 | RNase L inhibitor 1 |
| PABP | Poly-A binding protein |
| IRES | Internal ribosome entry site |
| tRNA | Transfer RNA |
| TC | Ternary complex |
| rRNA | Ribosomal RNA |
| r-protein | Ribosomal protein |
| Met-$tRNA_i^{Met}$ | Methionine initiator transfer RNA |
| GTP | Guanosine-5'-triphosphate |
| GMP-PNP | Guanosine 5′-[β,γ-imido]triphosphate |
| PIC | Preinitiation complex |
| SSU | Ribosomal small subunit |
| LSU | Ribosomal large subunit |
| ES | Ribosomal expansion segment |
| Pol | RNA Polymerase |
| eRF1 | eukaryotic release factor 1 |
| KSD | Kinetoplastid-specific domain |
| PAP | Poly (A) polymerase |
| KSRP | Kinetoplastid-specific ribosomal protein |
| PKR | Protein kinase RNA |

# INTRODUCTION

mRNA translation is the cornerstone of the central dogma of molecular biology. The genetic information encoded in DNA is transcribed into mRNA and the latter is translated, forming proteins. Like in bacteria, translation in eukaryotes includes several stages, such as initiation, elongation, termination and recycling. All these steps of mRNA translation are assisted by different proteins/factors termed according to the stage of translation that they assist. Translation initiation in eukaryotes is more complex than that in bacteria. While in bacteria the small ribosomal subunit (30S or 40S in eukaryotes) is recruited to the mRNA through an interaction between the Shine-Dalgarno sequence and the complementary sequence in 16S rRNA (anti-Shine-Dalgarno sequence), in eukaryotes this process is more complex and requires a scanning mechanism (reviewed in Jackson et al., 2010).

Although bacteria and eukaryotes present a deep divergence for the initiation step of translation, they share several common aspects for elongation. For instance, this process in eukaryotes is assisted by eukaryotic elongation factor 1 (eEF1) and eEF2 that are homologs of bacterial EF-Tu and EF-G, respectively.

Similarly to the initiation step of translation, the termination and recycling are divergent between bacteria and eukaryotes. For instance, ATP-binding cassette sub-family E member 1 (ABCE1) is a key player for the recycling process in eukaryotic translation, however this protein is not present in bacteria. All this process of mRNA translation occurs in one of the largest and most complex molecular machines, the ribosome.

## Structural and functional insights on the eukaryotic ribosome

The ribosome is a large macromolecular complex composed of both rRNAs and proteins (r-proteins). The discovery of this ribonucleoprotein dates back to 1955, when George E. Palade (Palade, 1955) described it as a "component of the ground substance of the cytoplasm which is particulate in nature and small in size". Because it is essential for any living organism, the architecture and functions of ribosome have been thoroughly studied for decades.

Early structural studies of bacteria and archaea ribosomes using X-ray crystallography had allowed us to have the first structural and functional insights of the ribosome (Ban et al., 2000; Wimberly et al., 2000; Schluenzen et al., 2000). However, eukaryotic ribosome presents several differences when compared to that from bacteria or archaea. The first structural attempt to reveal the architecture of eukaryotic ribosome was performed by cro-EM, which had allowed the identification of some structural differences as well as the universally conserved core

(Spahn et al., 2001). However, the discovery of the full architecture of the eukaryotic ribosome was possible only with recent advances of cryo-EM together with new approaches from X-ray crystallography (Armache et al., 2010; Ben-Shem et al., 2011; Hashem et al., 2013a; Anger et al., 2013; Khatter et al., 2015).

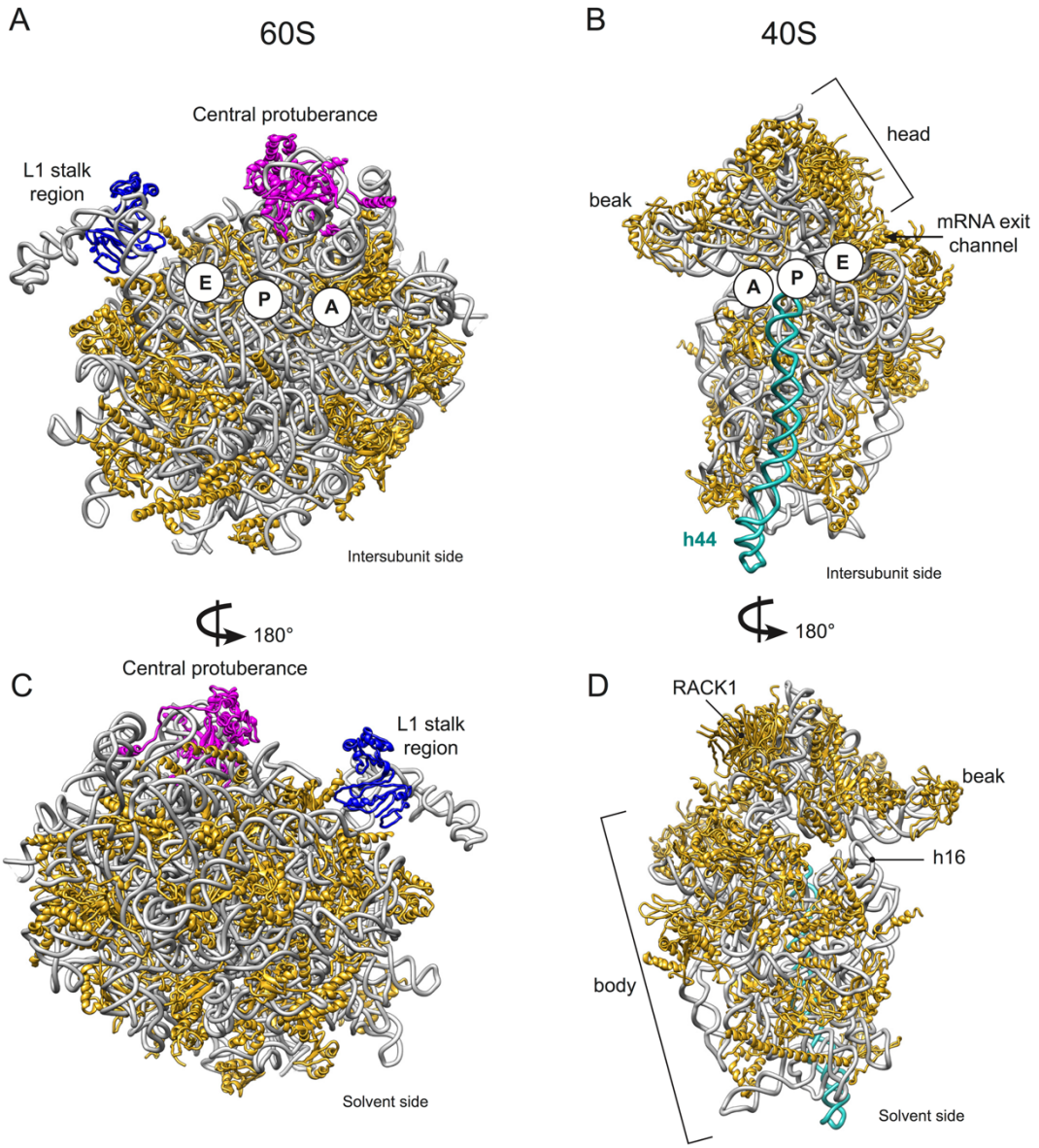


**Figure 1: Structure and some conserved features of eukaryotic ribosome.** The 60S and 40S subunits (PDB: 4UG0 from Khatter et al., 2015) viewed from different orientations. The r-proteins are colored in goldd and the rRNAs are colored in gray. In both subunits the three binding sites of tRNA are indicated, A, P and E-site. The central protuberance proteins are colored in magenta, while the L1 stalk region is colored in blue.

The eukaryotic ribosome is composed of 40S (small subunit) and 60S (large subunit) (Figure 1). The two subunits can reversibly associate through the inter-subunit bridges to form the 80S ribosome (Ben-Shem et al., 2011). 40S is composed by the 18S rRNA and by 33 r-proteins, while the 60S is built of 47 r-proteins and rRNA 5S, 5.8S and 28S (reviewed in Klinge et al., 2012).

In addition to more ribosomal proteins, one of the most remarkable differences between eukaryote and bacterial ribosome is the presence of large rRNA insertions named expansion segments (ESs) in eukaryotes (Spahn et al., 2001; Ben-Shem et al., 2011; Hashem et al., 2013a; Anger et al., 2013; Khatter et al., 2015). The 18S contains several ESs that interlink to form often species-specific features. For instance, $ES6^S$ (the largest ES from 18S) can form five sub-structure that interact with $ES3^S$ to form a eukaryotic-specific structure at the left foot of 40S (Armache et al., 2010).

The r-proteins are also another remarkably difference between eukaryotic and bacterial ribosome. In addition to eukaryotic-specific r-proteins from both 40S and 60S, eukaryotic ribosome also present some r-proteins extensions (reviewed in Klinge et al., 2012).

As a protein factory, the ribosome needs to read the instructions contained in the mRNA and translate it into protein sequence. For that, the ribosomes possess conserved binding sites for aminoacyl-tRNA, termed aminoacyl-site (A-site), peptidyl-site (P-site) and exit-site (E-site) (Figure 1A-B). The structure of the bacterial large ribosomal subunit revealed the ribozyme activity of rRNA and allowed the identification of the conserved peptidyl transferase center (PCA) (Nissen et al., 2000). The aminoacyl-tRNA that are ready to be incorporated into the growing polypeptide chain binds at the A-site, while the P and E-sites are occupied by deacylated tRNAs prior their dissociation from the ribosome. However, during the translation initiation in eukaryotes the initiatior tRNA must be accommodated into the P-site (discussed later).

In addition to the active core, eukaryotic ribosomes also share many structural features with bacteria, such as the central protuberance and the L1-stalk (Figures 1 A and C) (reviewed in Melnikov et al., 2012). The L1-stalk complex (formed by uL1 and rRNA) adopts different conformations during the translation cycle, and allows it to coordinate the position of E-site tRNA and promotes its release from the ribosome.

Although they conserve numerous structural and functional features, translation initiation is markedly different between eukaryotes and bacteria, in part the importance of translation initiation during the evolution.

**Translation initiation in eukaryotes**

Translation initiation in eukaryotes is a dynamic, cyclical and complex process, regulated by more than 20 different proteins (reviewed in Aitken and Lorsch, 2012). This process can be subdivided in two large types, cap-dependent and cap-independent.

The cap-dependent initiation is the most common process through which eukaryotic mRNA is translated, therefore is also known as canonical translation initiation. The first main step of translation initiation is the assembly of the 43S pre-initiation complex (43S PIC). This process starts with the formation of a ternary complex (eIF2-TC), composed by initiator tRNA (Met-$tRNA_i^{Met}$) and eukaryotic initiation factor 2 (eIF2) bound to GTP. Once assembled, eIF2-TC is recruited to the to the 40S ribosomal subunit. The recruitment of eIF2-TC is also enhanced by other factors, including eIF1, eIF1A, eIF3 and eIF5, which together lead to the formation of 43S PIC (Pestova and Kolupaeva, 2002; des Georges et al., 2015; Hashem et al., 2013b; Hussain et al., 2014a; Llácer et al., 2015). Interestingly, ABCE1 (Rli1 in yeast), a well-characteized recycling factor, was also shown to take part in 43S PIC assembly (Dong et al., 2004; Mancera-Martínez et al., 2017; Heuer et al., 2017).

The next step of translation initiation is the formation of the 48S initiation complex (48S IC). Once assembled, the 43S PIC is recruited to the mRNA, yielding the 48S IC. After the attachment of the mRNA to the 43S PIC, the scanning for the AGU start codon takes place. The scanning process include the movement of 43S PIC downstream the 5′ cap, which allows the base-pairing between the codon within 5′ mRNA to the anticodon within the Met-$tRNA_i^{Met}$ anticodon stem loop (ASL).

It was found that 43S PIC is able to bind the mRNA through their 5' unstructured regions (5′UTR) (Pestova and Kolupaeva, 2002). However, natural mRNAs possess different degree of secondary structures at their 5'UTR, including loops and helices (Kozak, 1999). The recruitment of the 43S PIC to the mRNA requires the activity of the multifactor cap-binding complex (eIF4F).

eIF4F is constituted by the cap binding protein (eIF4E), a DEAD-box RNA helicase (eIF4A), and the scaffold protein eIF4G (reviewed in Jackson et al., 2010). The helicase activity of eIF4A is enhanced by the activity of eIF4B (homologous of eIF4H) which binds to the single-stranded mRNA and prevent the re-annealing (Marintchev et al., 2009). The multifactorial complex eIF4F/4B is able to unwind the 5′ UTR mRNA, and promote it attachment to the 43S PIC likely via eIF4G-eIF3 interaction. Nevertheless, mRNAs with highly structured 5'UTR

require an auxiliary DEAD or DEAH box helicases, such as DHX29 (Pisareva et al., 2008; Parsyan et al., 2009).

After scanning, the initiation factors are released progressively, allowing the binding of the 60S subunit and the formation of 80S initiation complex.

Although cap-dependent is the general mechanism of translation initiation in eukaryotes, about 10% of cellular mRNA are able to be translated in a cap-independent mechanism (Stoneley and Willis, 2004; Johannes et al., 1999). These mRNAs contain an Internal Ribosome Entry Site (IRES) at the 5′ UTR that directly recruit ribosome for translation initiation.

### The role of eIF1 and eIF1A during translation initiation

eIF1 and eIF1A are essential proteins that bind to 40S subunit (Figure 2A) with high affinity (Lomakin et al., 2003; Battiste et al., 2000; Pestova et al., 1998). eIF1A is a monomeric protein that binds near the A-site (Figure 2B), while eIF1 binds near the P-site (Yu et al., 2009; Weisser et al., 2013; des Georges et al., 2015).

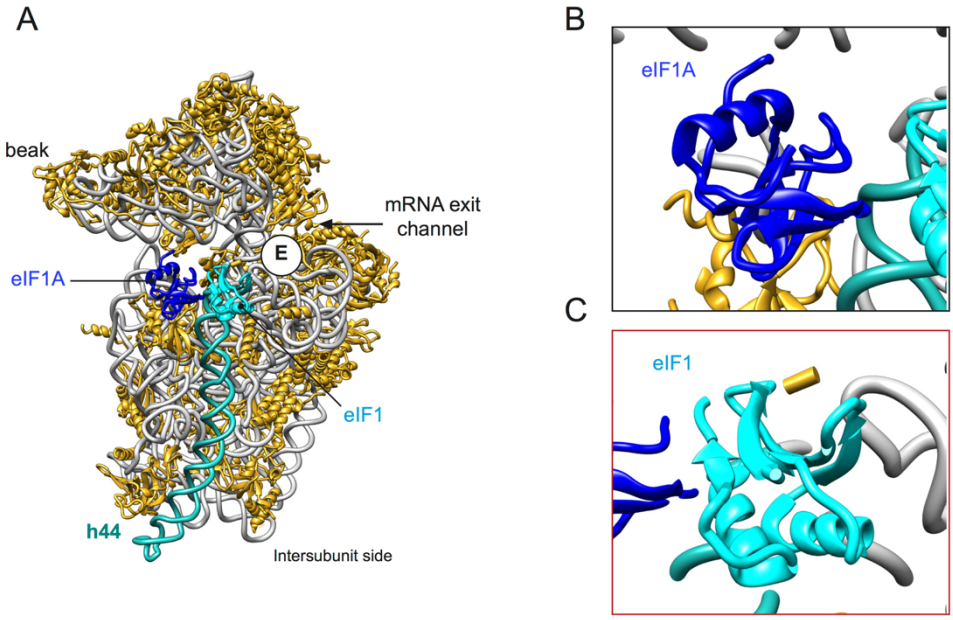


**Figure 2: Structure 40-eIF1A-eIF1 in context of yeast preinitiation complex. A,** Atomic model of yeast eIF1 and eIF1A bound on the ribosome (PDB: 3JAQ from Llácer et al., 2015), highlighting their position on the top of h44. **B**, Close-up view of the A-site, highlighting the eIF1A. **C**, Close-up view of the P-site, highlighting the eIF1.

X-ray crystallography and cryo-EM structures of eIF1A and eIF1 in context of 40S have revealed their OB barrel folds (globular). eIF1 does not interacts with the head of 40S subunit, instead it interacts only with rRNA helices 24, 44 and 45 located at the body of 40S. On the

other hand, eIF1A interacts with r-protein eS31 located on the head of 40S. Furthermore, eIF1A also interacts with elements on the body of 40S, such as r-proteins uS12 and eS30, and with the rRNA helices 18 and 44 (Weisser et al., 2013). Although their close proximity in the context of initiation complex, eIF1 and eIF1A do not interact. Indeed, the only common contact point they possess is the h44 (Weisser et al., 2013).

The scanning process is only possible if the mRNA entry channel latch of 40S subunit is open. In this context, the first attempt to reveal the architecture of these two factors in context of 40S subunit have suggested that they bind synergistically to 40S and that induce an upward movement of the 40S head, which open the mRNA latch (Passmore et al., 2007). However, structural studies that proceed that first attempted have found that this activity is performed synergistically with other initiation factors (Llácer et al., 2015; Weisser et al., 2013).

eIF1A is a homolog of bacterial IF1, however eIF1A possesses additional unstructured N- and C-terminal tail (CTT) (Battiste et al., 2000). Indeed, the widely established function for eIF1 and eIF1A is assured in part through the tails that stabilize the open conformation. For instance, the CTT of eIF1A is located in the P-site, where it prevents the full accommodation of the tRNA, while its NTT enhance the start codon recognition (Saini et al., 2010).

eIF1 also play another important role in translation by ensuring the fidelity of start codon selection. It prevents premature hydrolysis of eIF2-bound GTP (reviewed in Jackson et al., 2010). Nevertheless, due to its location near to the P-site, its release is important for the formation of stable codon-anticodon base pairing and the close conformation (Maag et al., 2005; Unbehaun et al., 2004; Lomakin et al., 2003).

**eIF2-Ternary complex and the codon-anticodon base pairing**

A key step in the mRNA translation is the base paring between the anticodon of a tRNA to the codon within the mRNA (reviewed in Jackson et al., 2010). Cells contain 20 different amino acids attached to their corresponding tRNAs, in a process mediated by a group of enzymes called aminoacyl tRNA synthetases. The first tRNA which participates in translation is the Met-$tRNA_i^{Met}$. The accommodation of the Met-$tRNA_i^{Met}$ into the P-site is guaranteed by eIF2 (Figure 3). eIF2 is a heterotrimeric protein composed by eIF2α, eIF2β and eIF2γ subunits (Lloyd et al., 1980; Schmitt et al., 2010). The eIF2γ subunit is the central core of eIF2 and it contains the main domain responsible for the binding of GTP and Met-$tRNA_i^{Met}$ (Kimball, 1999; Naranda et al., 1995), while eIF2α is the global regulator of mRNA translation. The functional role of eIF2β is still poorly understood, however it has been suggested to be involved the the

stabilization of Met-tRNA$_i^{Met}$ (Naveau et al., 2010; Llácer et al., 2015). Furthermore, eIF2β is involved in several interactions with other initiation factors, such as eIF1, eIF1A and eIF5, as well as with the anticodon stem loop (reviewed in Aitken and Lorsch, 2012; Hinnebusch, 2017). These interactions are important for the fidelity of the initiation process.

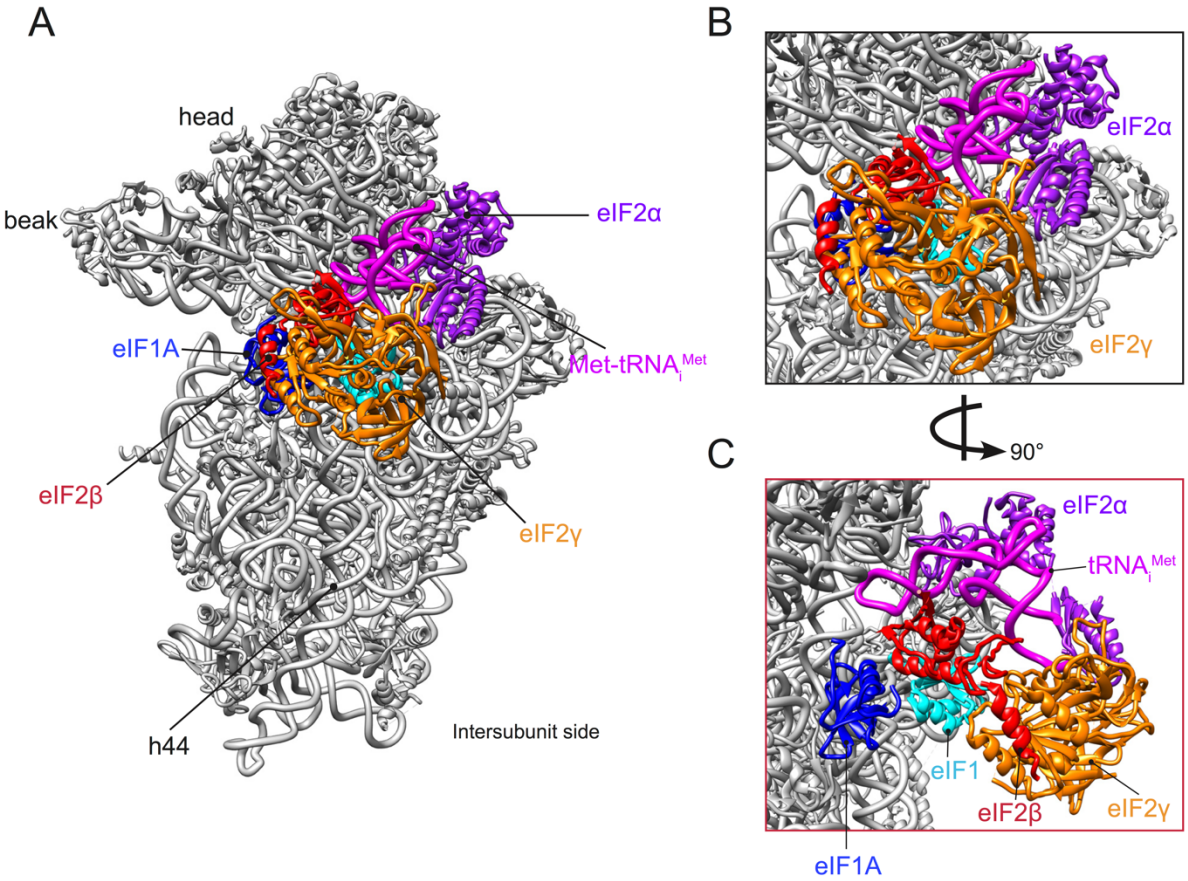


**Figure 3: Structure of eIF2-TC in context of initiation complex. A,** Structure of yeast initiation complex highlighting the architecture of eIF2 in context of pre-initiation complex (PDB: 3JAQ from Llácer et al., 2015). **B-C**, close-up views highlighting the arrangement of the eIF2-TC during the Met-tRNA$_i^{Met}$ accommodation into the P-site.

Fidelity of initiation is a discriminatory process that prevents partial base pairing of triplets of the 5′ UTR mRNA with the anticodon. During canonical translation initiation the codon is the first AUG in the correct frame GCC(A/G)CC**AUG**G, and normally it is flanked by a purine (A or G) at -3 position and a G at the +4 position (relative to the A of the AUG codon, as described by Kozak) (Kozak, 1986; Kozak, 1987; Kozak, 1991). It was found that the -3 and +4 positions play a role on start codon selection by interacting with eIF2α and h44 of 18S

rRNA, which stabilizes the conformational changes that take place after the codon-anticodon base pairing (Pisarev et al., 2006).

After the start codon recognition, the eIF2 is released from the 48S IC. The start codon recognition lead to the release of eIF1, and consequent eIF2γ-bound GTP hydrolysis. The GTPase activity of eIF2γ is activated by the GTPase activating protein, eIF5 (Paulin et al., 2001). The carboxyl terminal domain (CTD) of eIF5 binds to eIF2β (Luna et al., 2012), which allows the eIF5 N-terminal domain to induces the GTP hydrolysis performed by eIF2γ, and consequent eIF2 release together with eIF5 (Unbehaun et al., 2004; Paulin et al., 2001; Saini et al., 2014).

The affinity of eIF2 to the Met-$tRNA_i^{Met}$ is the base of the main mechanism of translation initiation regulation among eukaryotes. The intrinsic affinity of eIF2 to GDP is higher than to GTP (Erickson and Hannig, 1996). Therefore, the guanine nucleotide exchange is catalyzed by another factor, eIF2B (guanine nucleotide-exchange factor). In order to promote the nucleotide exchanging, the C-terminal segment of eIF2Bε interacts directly with the G domain of eIF2γ and with lysine-rich segments of eIF2β (Gomez and Pavitt, 2000; Gomez et al., 2002). Nevertheless, eIF2α also interacts with eIF2B through the regulatory sub-complex (eIF2Bα, eIF2Bβ and eIF2Bδ), and this interaction is highly reinforced under eIF2α phosphorylation conditions (Krishnamoorthy et al., 2001). It has been proposed that the phosphorylation of serine 51 within eIF2α lead to formation of a tight bound between eIF2α and eIF2B regulatory sub-complex, which affects the interaction between eIF2γ and eIF2Bε and lead to the formation of inactive eIF2–eIF2B complex (reviewed in Holcik and Sonenberg, 2005). Translational control plays a key role on gene expression regulation and cell homeostasis. Its effects on cell is faster than the transcriptional control, therefore it is important for the viability of the cells under certain environmental conditions, such as stress or starvation.

In summary, the eIF2 plays a key role during the initiation by recruiting the Met-$tRNA_i^{Met}$ and accommodated it into the P-site, on the other hand its regulation also guaranteed cells viability under stress condition.

**Structure of eIF3 and its interactions with 40S subunit**

eIF3 is the largest (~800 kDa) and the most complex eukaryotic translation initiation factor. It participates in nearly every step of translation initiation. Indeed, it plays a key role during the recruitment of both eIF2-TC and mRNA. Furthermore, it is also important for the fidelity of

the start codon selection (reviewed in Valášek et al., 2017). Finally, eIF3 is also important in preventing the premature association of 60S subunit (Kolupaeva et al., 2005).

In mammals, this multimeric protein complex is composed of 13 subunits, termed eIF3a to eIF3m. In other eukaryotic organisms such as *Saccharomyces cerevisae* or kinetoplastids, eIF3 is composed by 6 and 12 orthologous of mammalian eIF3, respectively (Phan et al., 1998; Rezende et al., 2014).

The subunits eIF3a, c, e, k, l and m possess a PCI domain (for **P**roteasome-**C**OP9 signalosome-e**I**F3), characterized by harboring several helical repeats. On the other hand, the subunits eIF3f and eIF3h harbour a MPN domain (for **M**pr1-**P**ad1 **N**-terminal), which consist of a β-barrel surrounded by α helices and additional β strands. That suggest that together they belong to a large multiprotein complex (Sun et al., 2011; des Georges et al., 2015). Accordingly, structure of mammalian eIF3 in context of 43S PIC confirmed that these 8 subunits form the eIF3 structural core (Figure 4) (des Georges et al., 2015). The remaining subunits (eIF3b, d, g, i and j) are known as peripheral subunits.

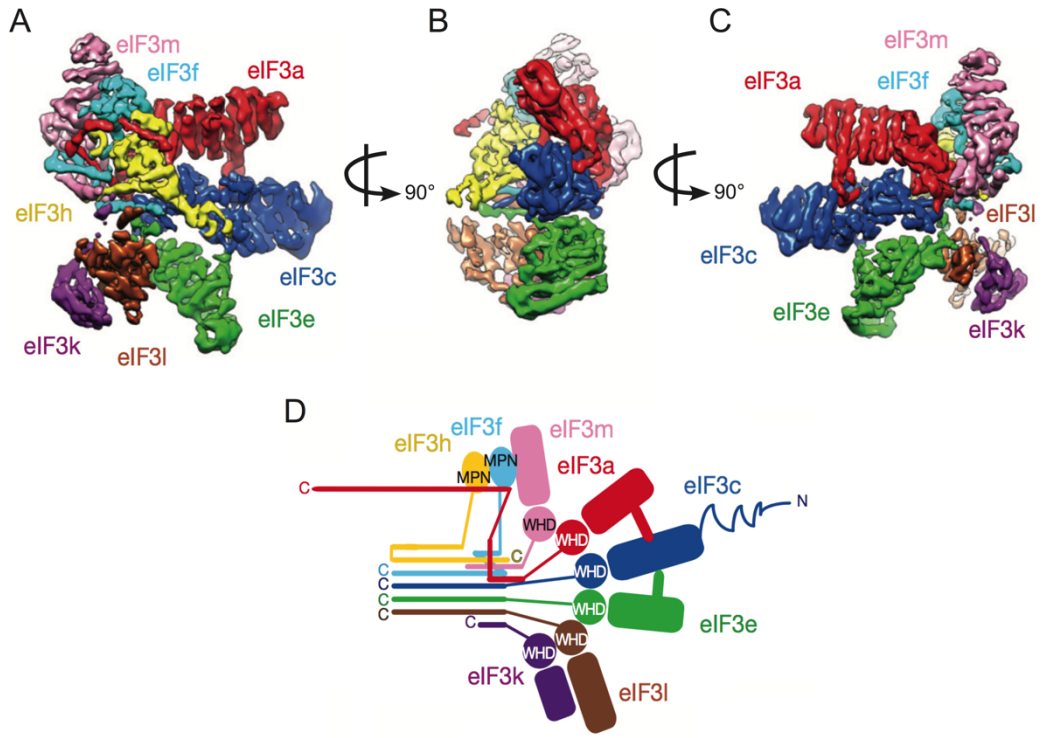


**Figure 4: Cryo-EM structure of mammalian eIF3 octamer core. A–C,** Segmented Cryo-EM map of eIF3 core viewed from different orientations. Each colour represents one of the subunit that composed the eIF3 octamer core. **D**, Schematic two-dimensional representation of the three-dimensional structure of eIF3 core. Zig-zagged line on eIF3c indicates possibly unstructured N-terminal tail. (Adapted from des Georges et al., 2015).

Functionally, is the sub-complex formed by the conserved subunits eIF3a and eIF3b is of pivotal importance. The interaction between these two subunits occurs through the N-terminal RNA recognition motif (RMM) of eIF3b with the CTD of eIF3a (Valásek et al., 2001; Khoshnevis et al., 2012). In yeast it has been shown that the CTD of eIF3b is responsible for the recruitment of the eIF3g and eIF3i (Asano et al., 1998). Accordingly, it has been shown that in mammals the eIF3b also form a subcomplex with eIF3g and eIF3i, that bind at the solvent side of 40S subunit in close vicinity to h16. Indeed, this subcomplex also interacts with the CTD of eIF3a (Figure 5) (Hashem et al., 2013b; des Georges et al., 2015).

According the present model, the first step of eIF3 formation is the binding of eIF3c, eIF3f, eIF3h and eIF3m to the eIF3a, already bound to the eIF3 b-i-g. The following step is the binding of eIF3e and eIF3d. The last event for the assembly of eIF3 is the binding of the subunits eIF3k and eIF3l (Wagner et al., 2016); Smith et al., 2016). This model is also in agreement with the structure of mammalian eIF3, which also indicated that eIF3k and eIF3l are loosely attached to the peripheral region of the octamer structural core (des Georges et al., 2015).

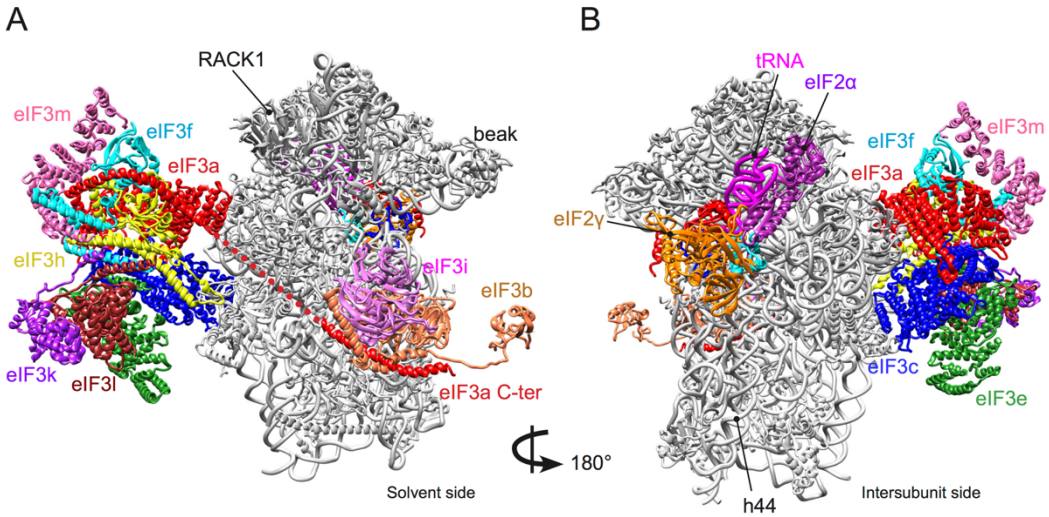


**Figure 5: Structure of mammalian eIF3 in context of 43S pre-initiation complex. A-B,** Atomic model of mammalian 43S PIC (PDB: 5A5T and 5A5U from des Georges et al., 2015), highlighting the distribution of eIF3 around 40S subunit (40S PDB: 5K0Y from Simonetti et al., 2016).

Cryo-EM structure of mammalian 43S PIC mapped the entire network of eIF3 octamer core with 40S subunit (Figure 5), excluding the CTT and NTT of eIF3a and eIF3c, respecitvely. In mammals, eIF3 core interactions with ribosome occur mainly through the subunits eIF3a and eIF3c. eIF3a interacts with eS1 and slightly with eS26. Furthermore, it also interacts with rRNA

through the apical loop of ES7$^{S}$. On the other hand, eIF3c interacts with r-proteins eS27 and slightly with uS15. Finally, eIF3c also interact with rRNAs through the apical loop of ES7$^{S}$ and with h22 (des Georges et al., 2015).

In yeast the interaction between eIF3 and the 40S subunit also occurs though the subunits eIF3a and eIF3c. Nevertheless, in yeast these two factors form a core that is rotated towards the solvent side, when compared to that one observed in mammals (Erzberger et al., 2014; Aylett et al., 2015; Valášek et al., 2003). Indeed, as mentioned above, there are some difference between yeast and mammals eIF3. For instance, yeast possess only six subunits of eIF3, including eIF3a, eIF3b, eIF3c, eIF3g, eIF3i and the non-consensual subunit, eIF3j. In yeast it binds bellow the beak of the 40S, bridging the r-proteins eS30 and uS12 (Aylett et al., 2015). That position in mammals can also be occupied by DHX29 (des Georges et al., 2015). Indeed, there is an ongoing controversy about the real nature of mammalian eIF3j, wondering whether eIF3j is a true subunit of eIF3 or it is an eIF3-associated factor (Wagner et al., 2016; Beznosková et al., 2013). Therefore, in theory the interaction of eIF3 with 40S can be markedly different among species, such as in kinetoplastids, where the eIF3 presents low conservation when compared to mammals.

### The role of RNA helicases during translation initiation

The attachment of the mRNA to the 43S PIC requires the activity of a multifactor complex, including the cap-binding complex (eIF4F) (Pestova and Kolupaeva, 2002).

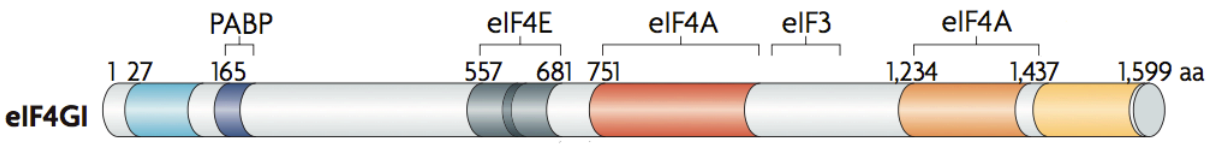


**Figure 6: Schematic representation of structural domains of eIF4G.** eIF4G is a scaffold protein that interacts with several others proteins during the scanning process. It contains binding domains of PABP, eIF4A, eIF4B, eIF4E, and eIF3. (Adapted from Jackson et al., 2010).

eIF4F is a multifactor complex constituted by the 5′ cap binding protein (eIF4E), a RNA helicase (eIF4A), and the scaffold protein (eIF4G). The eIF4F complex is able to unwind the 5′ UTR mRNA, and promote its attachment to the 43S, likely via eIF4G-eIF3 interaction (Figure 6). Nevertheless, in addition to eIF4A (also known as DDX2), several RNA helicases from the

DEAD and DEAH box family are also shown to be linked to translation. These two families of helicase harbor characteristic motifs Asp-Glu-Ala-Asp (DEAD) and Asp-Glu-Ala-His (DEAH), that distinguish them from other helicases (Linder et al., 1989). At least seven RNA helicase from these families are shown to be linked to translation, including eIF4A, DHX29, Ded1, Vasa, RNA helicase A (also known as DHX9), RCK (also known DDX6) and DDX25 (reviewed in Parsyan et al., 2011).

Although canonical helicases processively unwind RNA, the DEAD and DEAH box helicases that are linked to translation initiation seems to present poor processivity (reviewed in Parsyan et al., 2011). Furthermore, these ATP dependent RNA helicase do not use the energy from ATP hydrolysis to unwind the mRNA, instead the ATP is only important for the active close conformation and for the release of the enzyme from its RNA substrate (Liu et al., 2008; Chen et al., 2008). Indeed, several structural studies have addressed the structure of the helicase core (reviewed in Linder and Jankowsky, 2011). They have found that the core harbours the binding site for RNA and ATP. While RNA binds on the surface of the RecA (for recombinase A) domains, ATP bound in a pocket between the two RecA domains.

eIF4A and DHX29 are the two helicase widely studied in context of translation initiation. As mentioned above, eIF4A is an integrated protein within eIF4F complex. The main role of this protein during the initiation in through the unwinding the secondary structure at the 5′ UTR mRNA (Rozen et al., 1990). At least three isoforms of eIF4A have been identified in eukaryotes, eIF4A I to III. Nevertheless, only eIF4AI and eIF4AII are known to be functionally related and assist mRNA translation (Nielsen and Trachsel, 1988). In spite of the tremendous advances in structural studies thanks to cryo-EM, eIF4A structure in the context if the initiation complex has remained elusive. Functionally, eIF4A is a non-processive helicase, able to unwind short duplex. Indeed, it has been shown that its activity decreases with the increasing of the stability of the RNA duplex (Rogers et al., 1999). Therefore, the translation of mRNA with long and high structured 5’ UTR require the activity of other helicases such as Ded1 or DHX29 (Pisareva et al., 2008; Parsyan et al., 2009).

DHX29 is a RNA helicase member of the DEAH helicase family. This helicase is required for the translation of mRNA with highly structure 5’UTR (Pisareva et al., 2008) and only exists in higher eukaryotes such as mammals. It binds to the 40S subunit at the solvent site near the mRNA channel entry (Figure 7A). Indeed, structural studies found that it contacts the rRNA helix 16 and perhaps also the r-protein eS9 at the frontal face of 40S (Figure 7B). Furthermore, it also interacts directly with eIF3i and eIF3b (Hashem et al., 2013b, des Georges et al., 2015).

The binding of DHX29 at the mRNA channel entry allows it to be in a privileged position to unwind mRNA prior its entrance into the channel. Nevertheless, DHX29 is also a non-processive helicase, and its activity is stimulated by eIF3 rather than mRNA (Pisareva and Pisarev, 2016), and it has been shown that this helicase presents week contacts with mRNA in context of 48S IC. Therefore, it has been proposed that in addition to mRNA unwinding, DHX29 may play different roles during the scanning. Consistently, it was found that DHX29 NTD interacts with eIF1A OB domain and that it stimulate the recognition of the AUG codon (Pisareva and Pisarev, 2016).

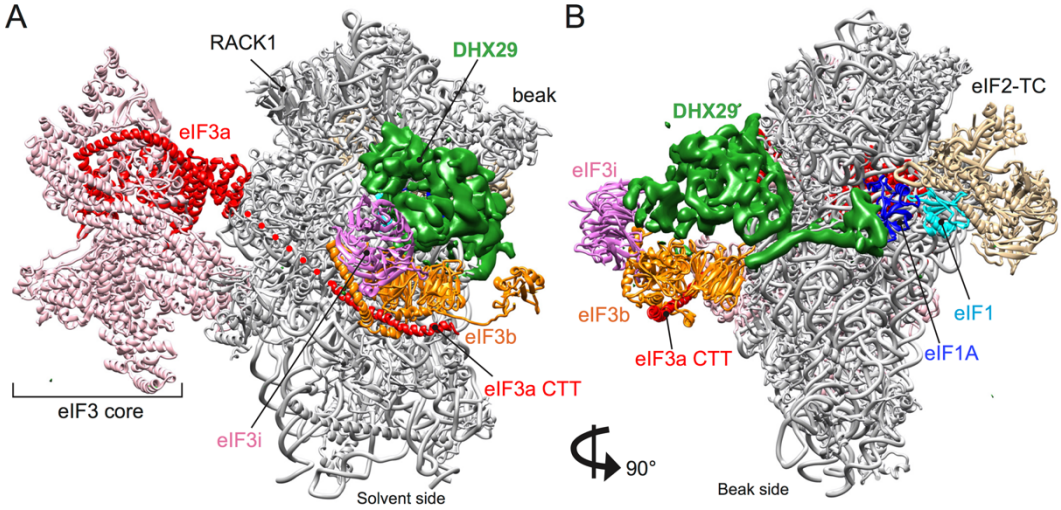


**Figure 7: Structure of DHX29 in context of mammalian 43S pre-initiation complex. A-B,** Structure of mammalian 43S PIC, viewed from different side, highlighting the binding site of DHX29 (eIF3 PDB: 5A5T and 5A5U; EMD-3057 from des Georges et al., 2015) (40S PDB: 5K0Y from Simonetti et al., 2016).

Put together, eIF4A and DHX29 seems to be the most important helicases during translation initiation. Nevertheless, structural data is still needed to complete our understanding of their entire mechanisms during translation initiation.

### 80S initiation complex formation

The last step of translation initiation is the 60S subunit joining. It has been shown that 60S can bind to 40S and form inactive 80S (Kolupaeva et al., 2005). Nevertheless, ABCE1 and some eIFs such as eIF3 and eIF6 can prevent the pre-mature association of 60S, which allows

the proper assembly of 48S initiation complex before the subunit joining. The subunit joining is mediate by eIF5B (Pestova et al., 2000), the eukaryotic GTPase that is homologous to the prokaryotic IF2 (Choi et al., 1998). The release of eIF5B is one of the last events that take place during the translation initiation process.

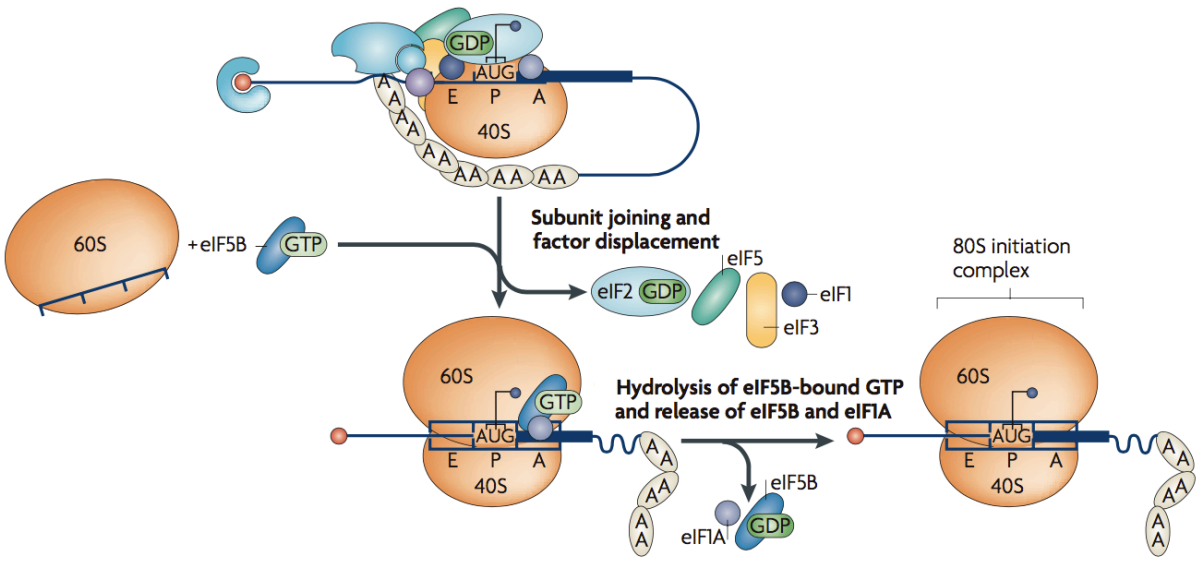


**Figure 8: Schematic representation of the last events in eukaryotic translation initiation.** After the start codon recognition and hydrolysis of eIF2-bound GTP, the initiation factors are displaced, remaining only eIF1A. The binding of eIF5B allows the 60S recruitment, forming 80S initiation complex. The eIF1A is also displaced together with eIF5B after the hydrolysis of eIF5B-bound GTP. (Adapted from Jackson et al., 2010).

The current model (Figure 8) suggests that after the displacement of the C-terminal tail of eIF1A from the P-site, it can interacts with eIF5B-CTD (Marintchev et al., 2003) and trigged the hydrolysis of eIF5B-bound GTP. The hydrolysis of GTP induce a conformational change on eIF5B and it release with eIF1A (reviewed in Jackson et al., 2010). The release of the factors allows the formation of active 80S initiation complex, able to accommodate the aminoacyl-tRNA into the A-site and proceed to the elongation steep of mRNA translation.

### Structure of the kinetoplastids ribosome

Kinetoplastid is a group of flagellated protozoans. Kinetoplastids possess some very unusual features in their ribosomes. Indeed, it was found that their ribosomes present significant structural differences when compared to those of their mammalian hosts (Hashem et al., 2013a).

As mentioned earlier, eukaryotic ribosome is composed by 40S/SSU and 60S/LSU subunits. The small subunit contains the 18S rRNA, while the large subunits contains 28S, 5.8S and 5S rRNAs (reviewed in Mandal, 1984). The 18S, 5.8S and 28S are encoded in a single transcription unit, known as the main transcription unit. The 5S rRNA gene is not linked to the main transcription unit. Indeed, the 5S gene is transcribed by RNA polymerase III, while the rest are transcribed by RNA pol I.

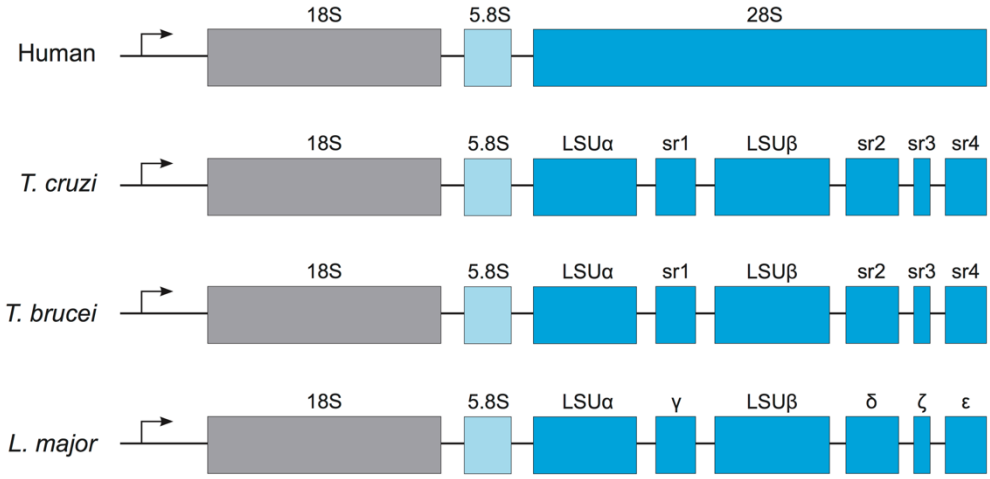


**Figure 9: General organization of the ribosomal main transcription unit.** Schematic representation of rDNA kinetoplastids and their human host counterpart.

The first evidence of the uniqueness of kinetoplastids ribosome came from the analysis of their rRNA. In contrast to their mammalian hosts, the kinetoplastidian homologous of 28S rRNA have multiple cleavage site (Campbell et al., 1987). Trypanosoma LSU rRNA is composed by LSU-α, LSU-β, four small rRNAs (srRNA1, srRNA2, srRNA3 and srRNA4), 5S and 5.8S rRNA (Figure 9) (Campbell et al., 1987; Hashem et al., 2013a; Liu et al., 2016). Furthermore, it has been shown that they possess an extensive number of rRNA modification sites (Eliaz et al., 2015).

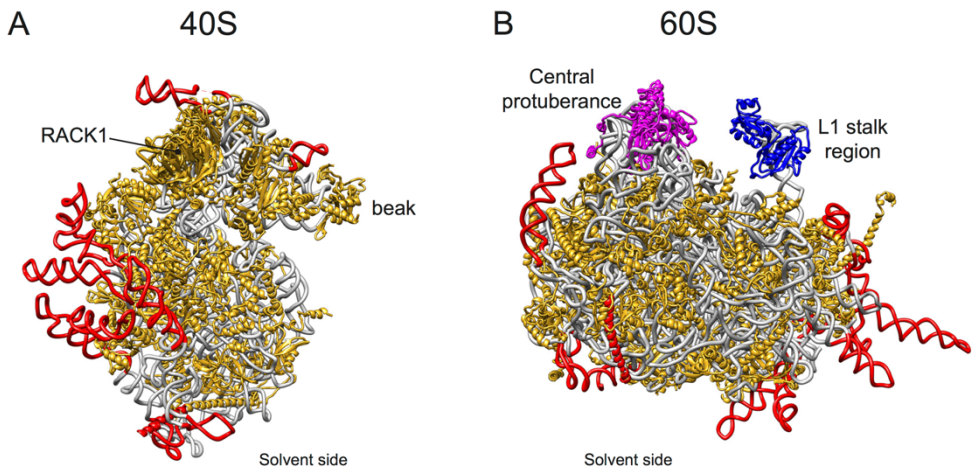


**Figure 10: Atomic model of the *T. brucei* ribosome, highlighting some kinetoplastid-specific elements that do not exist in other ribosomes.** The kinetoplastid-specific elements are colored in red. (PDB: 4V8M from Hashem et al., 2013a).

Recent advance of cryo-EM has led to the obtaining of kinetoplastids ribosome structure at a resolution as high as 2.5 Å. High resolution structure of kinetoplastids ribosomes have highlighted the presence of large expansion segments, as well as r-proteins extension (Figure 10) (Hashem et al., 2013; Liu et al., 2016; Zhang et al., 2016).

**Kinetoplastid unusual 60S puzzle**

The kinetoplastidian LSU-α is analogous to domains I and II of the yeast 25S rRNA, while LSU-β is analogous to domains IV and V. The srRNA1 and the srRNA2-4 are analogous to yeast 25S rRNA domains III and IV, respectively (Liu et al., 2016). Although these analogies to yeast, it is known that kinetoplastids LSU rRNA is larger than that in yeast (Hashem et al., 2013a).

Another remarkably difference of kinetoplastids 60S compared to other eukaryotes relies in their deficiency of the r-protein eL41 (Shalev-Benami et al., 2016a). eL41 is an important r-protein, responsible for the only eukaryote-specific intersubunit bridge locate at the center of the ribosome (Ben-Shem et al., 2011). Instead, they possess some uniqueness protein extensions that give them additional rRNA–r-protein interactions absent in any other known eukaryote ribosomes (Hashem et al., 2013a; Liu et al. 2016; Shalev-Benami et al., 2016a; Zhang et al., 2016a).

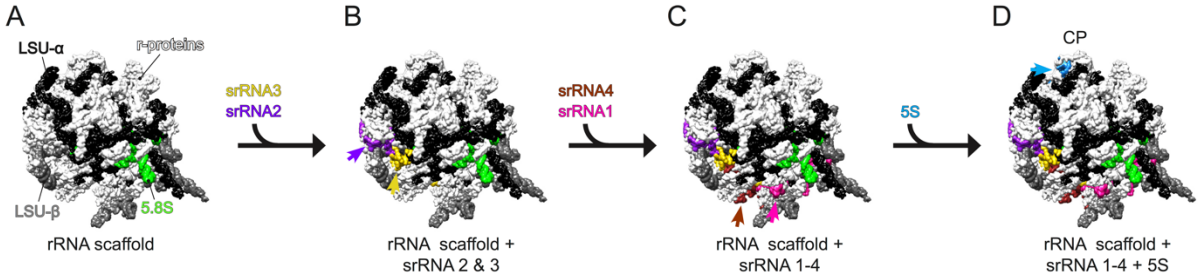


**Figure 11: The assembly model for kinetoplastids 60S rRNAs.** The rRNA LSU-α (black), LSU-β (gray) and 5.8S (green) form the scaffold which allow the assembly of the srRNA2 (purple) and srRNA3 (hellow), followed by the assembly of srRNA1 (pink) and srRNA4 (brown). The last piece of rRNA to be assembled is 5S (sky-blue), which is placed at the central protuberance (CP), forming the full assembled 60S. The r-proteins are colored in white (PDB: 4V8M from Hashem et al., 2013a).

As mentioned above, kinetoplastids LSU rRNA is an eight pieces' puzzle, and its proper assembly has been an enigma. Nevertheless, recent cryo-EM structure of *T. cruzi* 60S at atomic resolution (2.5 Å) allowed a detailed analysis of the rRNA–rRNA and rRNA–r-protein interactions involved in the stabilization of this "biological puzzle" (Liu et al., 2016). The comparative analysis of the kinetoplastidian 60S with that from yeast allowed them to propose a model for the assembly of the 60S (Figure 11). In the present model the 5.8S rRNA together with LSU-α and LSU-β form the core that acts as a scaffold for the assembly of the other LSU rRNAs. 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. Together, these two interactions form and stabilize the core of the scaffold. It is worth mentioning that both groups of interactions were observed also in yeast, but in *T. cruzi* LSU is via $ES3^{L}$ within 5.8S rRNA, which point 5.8S rRNA as central player for 60S rRNAs assembly (Liu et al., 2016).

The formation of the scaffold allows the assembly of srRNA2 and srRNA3. Indeed, it has been shown that srRNA2 has five contact sites with LSU-β and form several contacts with LSU-α, which support that assembly model (Shalev-Benami et al., 2016; Hashem et al., 2013a; Zhang et al., 2016; Liu et al., 2016).

It is well known that kinetoplastids present some ribosomal protein extensions (Hashem et al., 2013a; Zhang et al., 2016; Liu et al., 2016). The structure of 60S indicated that an extension in the globular domain of eL33 provides a binding site for both srRNA3 and LSU-α (Liu et al., 2016), that explain the assembly of srRNA3 through the scaffold. Finally, srRNA1 and srRNA4, the two last srRNAs to be assembled. The srRNA2 play a key role in the assembly

of these two srRNAs, because it provides a bind surface that allow them to be anchored on the scaffold (Liu et al., 2016).

The last piece of the puzzle seems to be 5S rRNA. It was found that 5S rRNA can bind the ribosomal proteins uL5 and uL18, forming the 5S ribonucleoprotein complex (RNP), which is important for the 60S biogenesis (Leidig et al., 2014). In yeast it has been proposed that RNP can acts as chaperone-like to guide the flanking 25S rRNA helices (Leidig et al., 2014). Remarkably, 5S ribonucleoprotein complex seems to be extremely conserved among eukaryotes, because it is also present in kinetoplastids (Michaeli and Agabian, 1990). Therefore, likely it can play a similar role in kinetoplastids.

The role of 5S ribonucleoprotein complex in ribosome biogenesis in kinetoplastids has been associated with the activity of a kinetoplastid-specific 60S-biogenisis factor, P34/P37 (Umaer et al., 2014). Nevertheless, recent study (Oliveira et al., 2016) reopened the discussion about the role of P34/P37 in 60S biogenesis and in translation in general, since P34/P37 has been found to co-purify with monosomes and polysomes (Oliveira et al., 2016). Several other biochemical studies reported the same observation (Ayub et al., 2009; Klein et al., 2015; Hellman et al., 2007; Umaer et al., 2014), which could have already indicated that P34/P37 is a r-protein rather than a 60S biogenesis factor.

Put together, kinetoplastids 60S presents some features markedly different from that observed in their mammalian host counterpart. Cryo-EM has decisively contributed in the understanding of their architecture and their assembly mechanism. Nevertheless, the 60S biogenesis in this organism still represents a big gap, highlighting the need a structure of a pre-mature kinetoplastids 60S, which can clarify the hypothetical role of P34/37 in this process.

**Structure and function of the 40S expansion segments**

The uniqueness of kinetoplastids ribosome extends also to 40S subunit. Cryo-EM structure of *T. brucei* and *L. donovani* 80S ribosome (Hashem et al., 2013a; Zhang et al., 2016) revealed their unusual large expansion segments, include that of 40S, such as $ES3^S$, $ES6^S$, $ES7^S$, $ES9^S$ and $ES10^S$.

The role of their larger ESs were discussed in several publications (Hashem et al., 2013a; Zhang et al., 2016). They have been associated with translation initiation and with the stability of 80S ribosome. Indeed, cryo-EM structure of *T. brucei* 80S ribosome revealed that

they are involved in the formation of four additional inter-subunit bridges that seems to be kinetoplastids-specific (Hashem et al., 2013a).

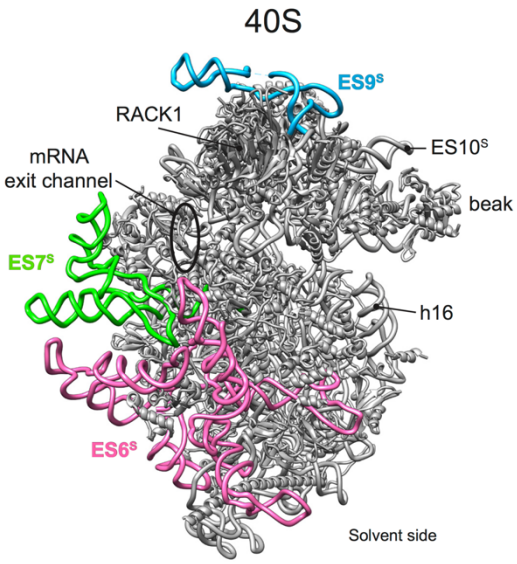


**Figure 12: Structure *T. brucei* 40S subunit.** The solvent side of the *T. brucei* 40S (PDB: 4V8M from Hashem et al., 2013a) highlighting expansion segments of 40S.

Another important aspect of kinetoplastids 40S is the location of their large expansion segments ES6$^{S}$ and ES7$^{S}$ near to the mRNA exit channel (Figure 12). It has been proposed that such location can allow them to play a role in translation initiation by interacting with eIF3 (Hashem et al., 2013a). Although no data from Structural Biology or Biochemistry corroborate such observation, this hypothesis is plausible since according to our current knowledge of the structure of eIF3 (Hashem et al., 2013a; des Georges et al., 2015) it is reasonable to assume that eIF3 interacts with both expansion segments during the assembly of the 43S pre-initiation complex.

In summary, kinetoplastids 40S possesses some specific elements which suggest that translation initiation in these organisms may have several divergences when compared to that observe in their mammalian host.

### The unusual RNA polymerase II from kinetoplastids

Kinetoplastids have a complex life cycle, including stages in invertebrate, vertebrate or plant hosts. All these developmental stages are orchestrated by an unusual gene expression mechanism, which include polycistronic transcription of the protein encoding genes and mRNA maturation through the trans-splicing (Jensen et al., 2009; Agabian, 1990; Sutton and

Boothroyd, 1986; Clayton, 2016). Furthermore, they possess a highly divergent mechanism for gene expression regulation relative to other eukaryotes (reviewed in Clayton and Shapira, 2007).

The gene expression starts with the transcription, in which the RNA pol enzymes copy a segment of DNA, yielding different types of RNA. Unlike prokaryotes that perform the transcription by using a single type of RNA pol, in eukaryotes the transcription is performed by at least three types of RNA pol. RNA pol I transcribes rRNA precursor, RNA pol II transcribes pre-mRNA and regulatory RNAs, and finally RNA pol III transcribes small non-coding RNAs such as transfer RNAs (tRNA).

Distinctively to what happens in most eukaryotes, kinetoplastids RNA pol I can also transcribe a subset of mRNA. This unexpected role has been observed in *T. brucei*, in which RNA pol I can transcribe variant surface glycoproteins (VSGs) and the procyclins encoding genes (Günzl et al., 2003).

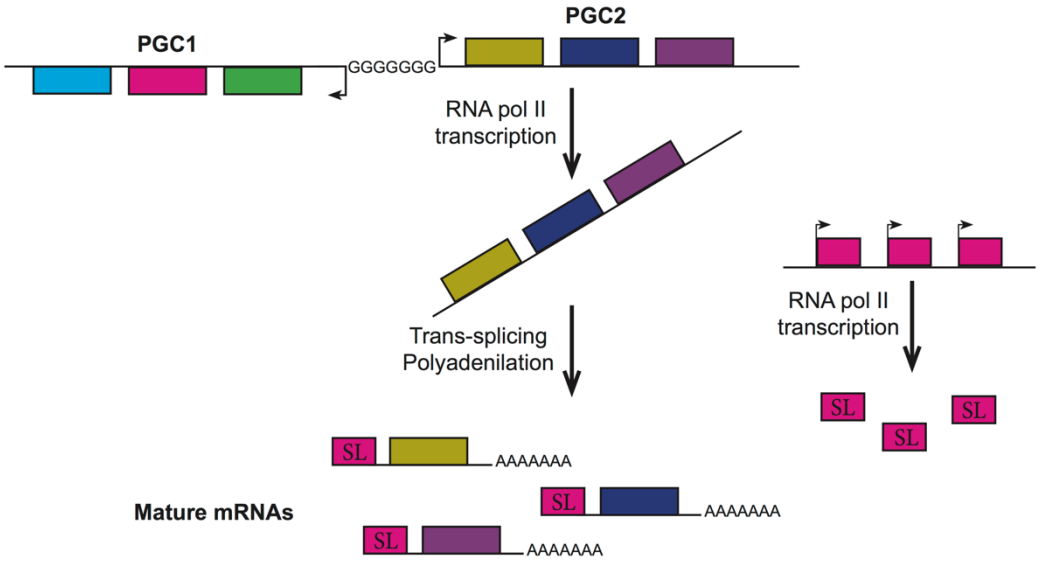


**Figure 13: Schematic representation of RNA pol II transcription and mRNA maturation process.** RNA pol II starts the transcription of the of the polycistronic gene cluster (PGCs) at the strand-switch region (SSR), yielding a polycistronic pre-mRNA. The primary transcript is processed by trans-splicing at the 5′ end and polyadenylation at the 3′ end, which lead to the formation of a monocistronic mature mRNA.

The nuclear genome of kinetoplastids are organized into large polycistronic gene clusters (PGCs) (Ivens et al., 2005; El-Sayed et al., 2005b; Berriman et al., 2005). These clusters can contain up to hundreds open reading frames arranged sequential on the same strand of DNA (Figure 13) (Berriman et al., 2005; El-Sayed et al., 2005a; Ivens et al., 2005; Johnson et al.,

1987). Therefore, the RNA pol II transcription in kinetoplastids is polycistronic. Furthermore, it has been shown that kinetoplastids RNA pol II starts the transcription at the strand-switch region (SSR) containing a bidirectional promoter (Martínez-Calvillo et al., 2003). In addition to the transcription of protein encoding gene, kinetoplastids RNA pol II also play another important role in gene expression, by transcribing the spliced leader RNA encoding genes (Gilinger and Bellofatto, 2001).

Although eukaryotic transcription initiation by RNA pol II is well-known (reviewed in Sainsbury et al., 2015), this process is still poorly understood in kinetoplastids. Indeed, only few general transcription factors have been identified in kinetoplastids, and some of them present very low degree of similarity to other eukaryotes (Martínez-Calvillo et al., 2010). Therefore, is difficult to present a full picture of transcription initiation in kinetoplastids.

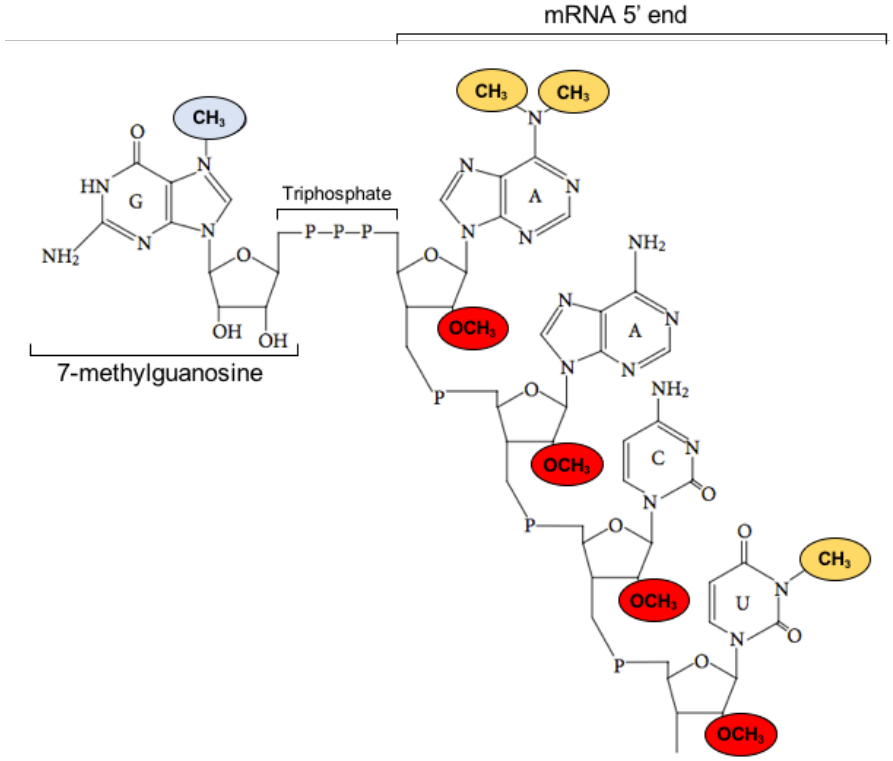


**Figure 14: Chemical structure of kinetoplastids cap-4.** Splice leader confer to all mature mRNA an unusual hyper-methylated cap-4 structure. cap-4 has 2′-O ribose methylations on the first four nucleotides and two base methylations on the first adenosine and fourth uridine (Adapted from Bangs et al., 1992).

Unlike prokaryotes where the nascent RNAs are ready to be translated, in eukaryotes the first transcript undergoes to a maturation process. The maturation process includes capping, polyadenylation and splicing. Capping is a co-transcriptional process by which a 7-methylguanosine cap is added to the 5′ end of the pre-mRNA. Polyadenylation is also a co-transcriptional process, by which the primary transcript is cleaved downstream from the

AAUAAA conserved sequence, and that allows the poly (A) polymerase (PAP) to add an adenine tail at the 3′ end. The third maturation process is splicing, a mechanism by which the introns are excised by spliceosome. Remarkably, kinetoplastids present an unusual low number of introns. Therefore, instead of *cis*-splicing, kinetoplastids mRNA maturation is essentially by the trans-splicing (Jäger et al., 2007; Parsons et al., 1984; Sutton and Boothroyd, 1986).

Trans-splicing is a process through which a highly conserved 39 nucleotides (Kramer and Carrington, 2011; Parsons et al., 1984; Preußer et al., 2012; Sutton and Boothroyd, 1986) miniexon, also known as the spliced-leader (transcribed by RNA pol II), is transpliced at the 5′ end of the mRNA (Figure 13). The spliced-leader confers to all mature mRNA a 5′ hyper-methylated cap-4 structure (Figure 14).

Transcriptional control is one of the key steps of gene expression regulation in eukaryotes. Although the RNA pol II promoter has been identified for this group of ancestral eukaryote, clues about its regulation mechanisms are lacking (reviewed in Clayton and Shapira, 2007). The general accepted model for gene expression regulation in kinetoplastids alleges that the regulation is essential at the post-transcriptional level. The post-transcriptional regulation introduces an additional importance to the translational control mechanism in the general control of gene expression and cell homeostasis. Several mechanisms have been indicated to regulate the gene expression in kinetoplastids, including trans-splicing, polyadenylation, mRNA degradation (both in nucleus and cytosol), mRNA export from the nucleus, formation of cytoplasmic RNP granules, and translation (reviewed in Clayton and Shapira, 2007). Nevertheless, little is known about translation control in these organisms.

**Kinetoplastids biology and human vector borne diseases**

For more than one century kinetoplastids have been attracting the attention of the scientific and medical community due to their relevance to public health. Nevertheless, the World Health Organization (WHO) to date, is still considering all the human diseases caused by kinetoplastids as a group of neglected tropical diseases.

*Trypanosoma cruzi* (*T. cruzi*), *T. brucei spp.* and *Leishmania spp.* are kinetoplastids responsible for several human vector-borne diseases, that together threaten more than 400 million people world-wide (World Health Organization, 2012).

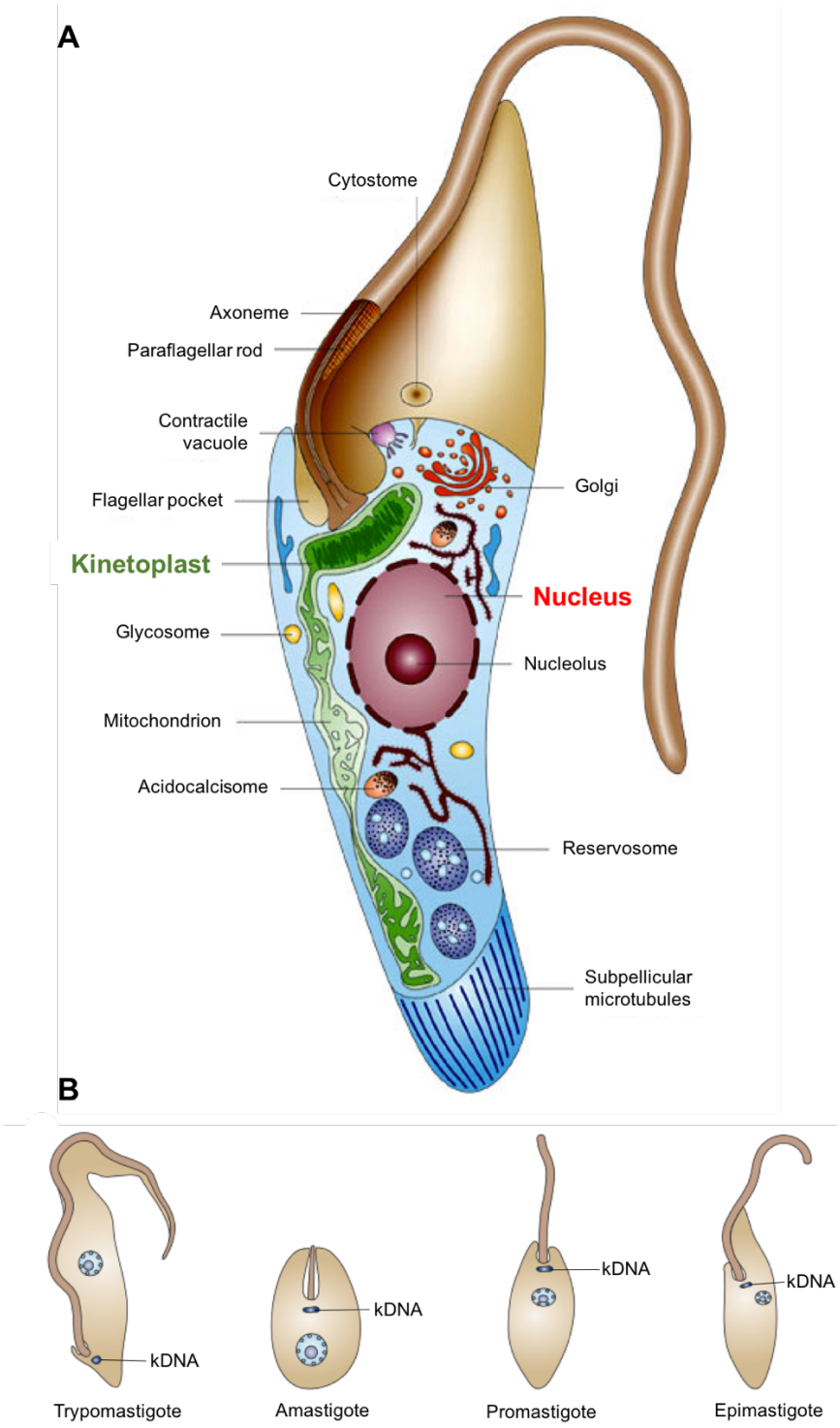


**Figure 15: Schematic representation of subcellular structure and main cellular forms of kinetoplastids. A,** Schematic representation of longitudinal section of *T. cruzi* epimastigote form. **B,** The main cellular forms of kinetoplastids based on cell shape, position of nucleus, and kinetoplast. (Adapted from Docampo et al., 2005).

As early-diverged from other eukaryotes, kinetoplastids share some uniqueness features, such as a single mitochondrion with an enlarged region known as kinetoplast (Figure 15A). Kinetoplast is located close to the basal body in an anterior or posterior position relative to nucleus –depending on life cycle stage (Figure 15B)– and it contains the unusual mitochondrial DNA structure known as kinetoplast DNA (kDNA).

kDNA is a mitochondrial DNA arranged in a very unusual fashion, which contains a network of catenated rings (Riou and Delain, 1969; Laurent and Steinert, 1970). It represents about 30% of the total cell DNA, and contains two different types of DNA, minicircles and maxicircles.

Maxicircles is the homologous of mitochondrial DNA in eukaryotes, and like in other eukaryotes it is responsible for encoding the rRNA and mRNAs for some mitochondrial proteins, include thus involved in oxidative phosphorylation process (Simpson, 1987). On the other hand, the minicircles encode for some small mitochondrial transcripts (guide RNAs) involved in RNA editing process (Sturm and Simpson, 1990).

Put together (*Leishmania spp., T. cruzi* and *T. brucei spp.*), about 20 million people worldwide are infected with kinetoplastids. In spite of their similarity, they are transmitted by different insect vectors and their life cycle diverge significantly from each others.

***Trypanosoma cruzi (T. cruzi)***

*T. cruzi* is the etiological agent of Chagas disease (also known as American trypanosomiasis), and it is transmitted to humans and to more than 150 species of domestic and wild animals. The transmission is mainly by the contact with feces of the blood-sucking vector, triatomine. In addition to the vectorial transmission, *T. cruzi* also can be transmitted to humans by non-vectorial mechanisms, such as from mother to infant (vertical), blood transfusion, organ transplantation, or by ingestion of contaminated food or drink (reviewed in Rassi et al., 2010).

The life cycle of *T. cruzi* (Figure 16) presents multiple developmental stages in both the insect vector and mammalian host. During its blood meal, triatomine ingests trypomastigotes from infected host. Once in the midgut of the insect, the trypomastigotes differentiate into the replicative form, epimastigotes. The epimastigotes differentiate into infective metacyclic trypomastigotes, readying them for mammalian host infection.

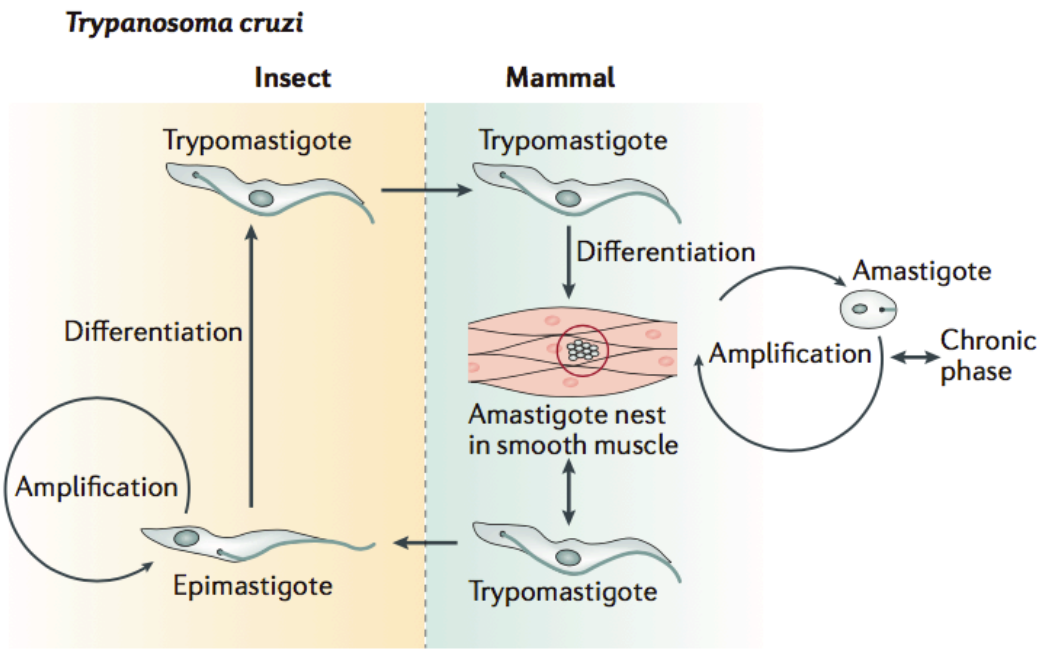


**Figure 16: Life cycle of *T. cruzi*, from insect to mammalian host.** *T. cruzi* has a complex life cycle, with a multiple developmental stages. Epimastigote is present only during the insect stage, while amastigote is the intracellular form, which is present only during the mammalian stage of the life cycle. Trypomastigote can be found both in mammalian host and in insect vector. (Adapted from Field et al., 2017).

The trypomastigotes are excreted in the feces of triatomine. The contaminated feces carrying parasites that can then enter through a bite wound or a mucous membrane of the host, where they can invade several types of cells. *T. cruzi* is an intracellular parasite, so once in the cytoplasm, the trypomastigotes differentiate into the intracellular replicative form, amastigotes. After the intracellular replication, the amastigotes differentiate into trypomastigotes, followed by the rupture of the cells, which released them into the bloodstream. The released parasite can invade new cells or infect an insect vector during its blood meal.

Chagas disease used to be confined only to Latin America countries – where vector-borne transmission occurs– but today this disease is becoming challenging also in non-endemic countries, such as USA or several European countries (Basile et al., 2011; Coura and Viñas, 2010).

According to the WHO, Chagas disease affects about 7 million people worldwide, however there is no preventives vaccines and there are only two drugs recommended for the treatment. Moreover, the first-line therapy, benznidazole (not currently approved by the Food and Drug Administration, USA), can cause some adverse effects in up to 30% of patients, which to lead the discontinuation of the treatment (Bern, 2011). Therefore, almost 50 years after the

introduction of benznidazole, there are a critical need for new treatments (Field et al., 2017; Clayton, 2010; Khare et al., 2016).

***Trypanosoma brucei spp. (T. brucei spp.)***

*T. brucei spp*. is the etiological agent of sleeping sickness (also known as human African trypanosomiasis). It is transmitted by the insect vector tsetse flies (*Glossina* genus). According to WHO, sleeping sickness is endemic in 36 sub-Saharan countries, where it threatens more than 50 million people (World Health Organization, 2012). Unlike Chagas disease, which is caused by a single species of *T. cruzi*, sleep sickness can be caused by at least two different species of parasites, *T. b. gambiense* and *T. b. rhodesiense*. The most common form of sleeping sickness is caused by *T. b. gambiense,* which is normally associated with the chronic infection. It is present in 24 countries in west and central Africa, and represents about 97% of the total cases of sleep sickness reported between 2000 and 2014 (Franco et al., 2017).

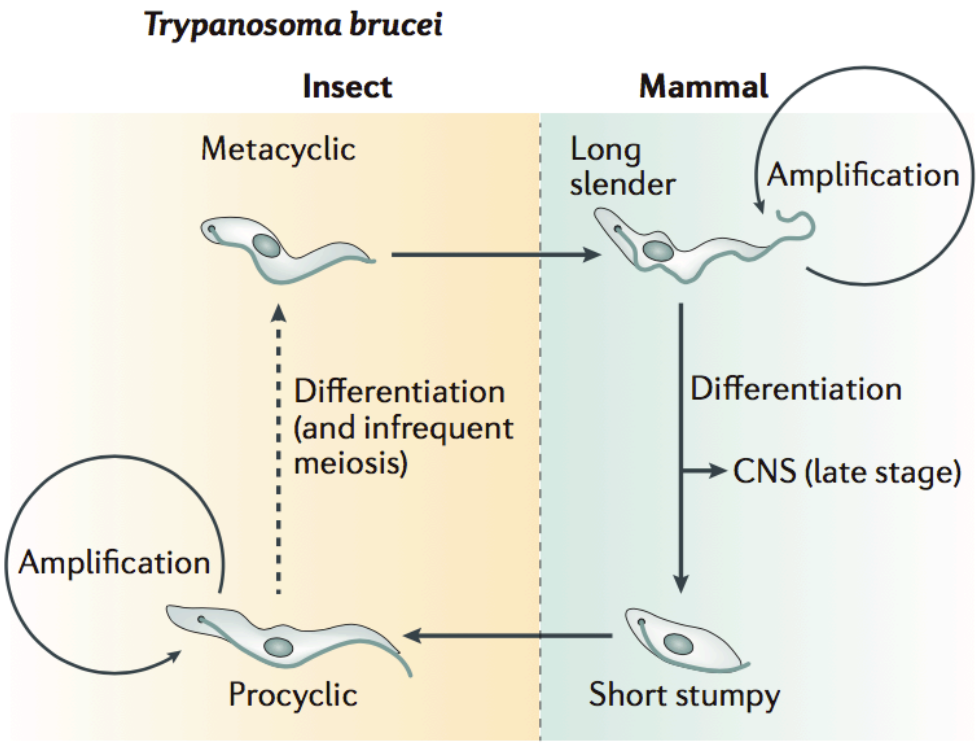


**Figure 17: Life cycle of *T. brucei*, highlighting the different stage development.** *T. brucei* can be transmitted to humans and others mammals by more than 20 species and subspecies of tsetse. (Adapted from Field et al., 2017).

The life cycle of *T. brucei* is complex and has multiple stages (Figure 17). During its blood meal a tsetse fly can injects the metacyclic trypomastigotes into the mammalian host. Once into the host, metacyclic trypomastigotes transform into bloodstream trypomastigotes, which can spread to different part of the body, including the central nervous system (CNS). *T. brucei* is an extracellular parasite and the trypomastigotes can multiply by binary fission in different body fluids, such as blood, lymph and spinal fluid. The blood stream long slender form can differentiate into short stump form, readying them for the life cycle into the insect vector (MacGregor et al., 2011). Once into the midgut of the insect, the short stump can differentiate into procyclic trypomastigotes and multiply by binary fission. The procyclic trypomastigotes can leave the midgut and transform into procyclic epimastigotes, which migrate to the salivary gland. Finally, after reach the salivary gland the epimastigotes multiply and transform into metacyclic trypomastigotes, able to start a new cycle of infection.

Like Chagas disease, the treatment for sleeping sickness still a challenge (Field et al., 2017). For example, despite the fact that sleeping sickness is treatable, for many years the main medication used for the late-stage of this disease was melarsoprol – a toxic compound derived from arsenic – that is responsible for the death of 3-10% of patients under treatment (Willyard, 2011).

***Leishmania spp.***

*Leishmania spp*. is the etiological agent of leishmaniasis, a vector-borne disease caused by more than 20 different species of *Leishmania*. Leishmaniasis is transmitted to humans through biting of about 30 different species of phlebotomine sandflies (Chappuis et al., 2007). According to WHO leishmaniasis is present in 98 countries or territories, threatening about 350 million people (World Health Organization, 2012).

Several species of *Leishmania* are responsible for different clinical form of leishmaniasis, including cutaneous, mucocutaneous, visceral and dermal leishmaniasis. Cutaneous leishmaniasis is the most common form of this disease and is caused by different species of *Leishmania*, include *L. major*, *L. tropica*, *L. donovani*, *L. braziliensis*, *L. mexicana*, or *L. aethiopica*. While mucocutaneous leishmaniasis is caused by *L. braziliensis* and *L. guyanensis*. Finally, visceral leishmaniasis, also known as kala-azar, which is known to be the most severe form of leishmaniasis. Visceral leishmaniasis is a systemic disease caused by the *Leishmania donovani* complex (*L. donovani* and *L. infantum*) (Lukes et al., 2007).

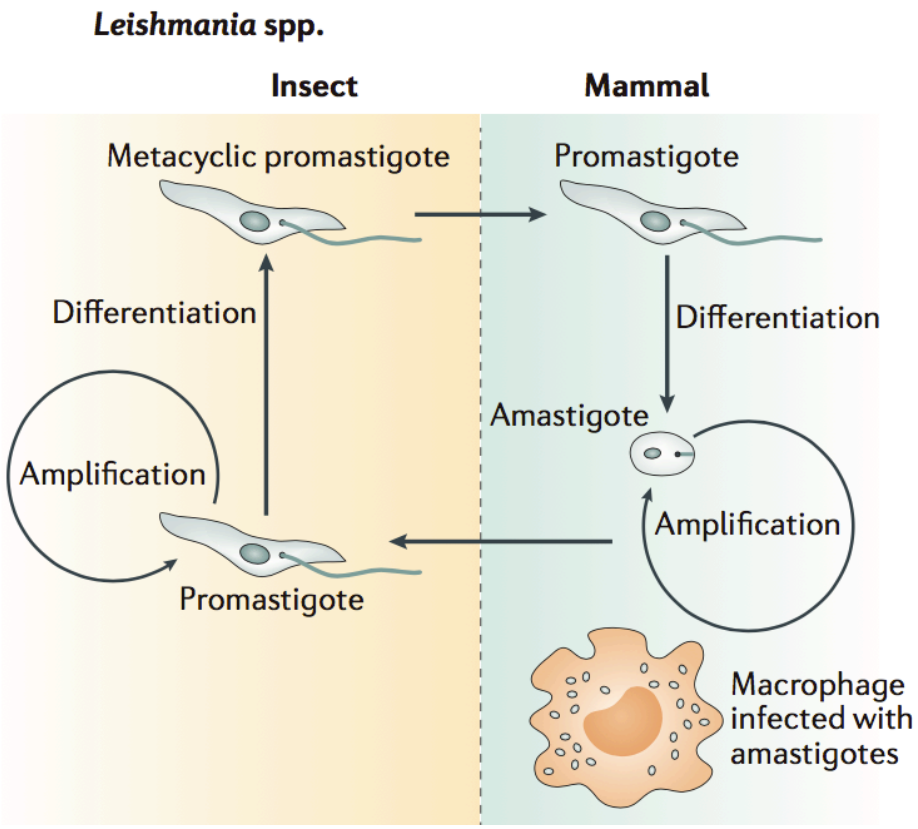


**Figure 18: The life cycle of *Leishmania*, from insect vector to human host.** Like any other kinetoplastids, *Leishmania* also present different forms during developmental stage. (Adapted from Field et al., 2017).

*Leishmania* is inoculated into mammalian hosts by female sandflies during a blood meal. (Figure 18). The injected promastigotes are phagocyted by several types of cells, include macrophages, and that allow them to transform into intracellular amastigotes. The life cycle of the parasite proceeds with the ingestion of infected macrophages by a sandfly. Once the amastigotes reach the midgut, they transform into promastigotes, which divide and migrate towards proboscis, readying them for a new cycle of infection.

Considering the deficiency of transcription regulation in kinetoplastids, the entire post-transcriptional step, including translation, has an additional importance for the gene expression regulation and cell homeostasis. Nevertheless, little is known about the translation initiation in kinetoplastids. The current knowledge point out the low identity of kinetoplastids eIFs compared to their mammalian host (Meleppattu et al., 2015; Rezende et al., 2014; Li et al., 2017), which highlights translation as a privileged area to be studied in these organisms. Therefore, the main aim of this PhD thesis is to understand the architecture of the translation initiation complexes in kinetoplastids, which can allow us to understand some of the kinetoplastid-specific aspect of mRNA translation. Kinetoplastids share a similar biology and

genomic sequence (El-Sayed et al., 2005b). These commons features may represent an opportunity to treat these diseases using a single effective and well tolerated drug (Khare et al., 2016).

# RESULTS

## Article 1: Structural study of 40S complexes from *Trypanosoma cruzi* reveals the existence of a kinetoplastid-specific translation initiation regulation pathway

### In Brief

The gene expression regulation in kinetoplastids is essentially at the post-transcriptional level. That suggest an additional importance to the translation control process. Querido et al., discovered an uncharacterized factor (termed provisionally ηF) at the platform of the 40S complexes. The binding site of ηF is incompatible with the recruitment of eIF2-ternary complex, as well as with the mRNA attachment to the 43S PIC. Our findings suggest ηF as a kinetoplastid-specific translation regulation factor. Furthermore, our structure revels that the binding of ABCE1 to the 40S is fully compatible with the formation of 43S PIC. Put together, here we present the first structural insights on kinetoplastids-specific aspects of translation regulation.

### Highlights

- Structure of 40S complex from *Trypanosoma cruzi* reveals a novel uncharacterized factor (ηF).
- ηF binds at the 40S platform and threads into the mRNA channel.
- The binding of ηF prevents the assembly of the initiation complex as well as the dimerization of 40S.
- The binding of ABCE1 to the 40S is fully compatible with the assembly of the initiation complex formation

**Structural study of 40S complexes from *Trypanosoma cruzi* reveals the existence of a kinetoplastid-specific translation initiation regulation pathway**

Jailson Brito Querido[1], Eder Mancera-Martínez[1], Angelita Simonetti[1], Marcelo Sousa Silva[2,3], Johana Chicher[4], and Yaser Hashem[1,*]

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

[2]Department of Clinical and Toxicological Analysis, Faculty of Pharmacy – Universidade Federal do Rio Grande do Norte, Natal, Brazil

[3]Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical – Universidade Nova de Lisboa, Lisbon, Portugal

[4]Université de Strasbourg, CNRS, Platforme Protéomique Strasbourg-Esplanade, FRC 1589, Strasbourg, France

*Correspondence should be sent to: y.hashem@ibmc- cnrs.unistra.fr (Y.H.)

**Summary**

Although many studies have highlighted the evolutionary divergence of mRNA translation in kinetoplastids, little is known about its structural basis. Here we present cryo-EM reconstructions of native and stalled 40S complexes purified from *Trypanosoma cruzi*, the kinetoplastid parasite causing Chagas disease. Our structures show that the ribosome-recycling factor ABCE1 interacts with the GTPase-binding site on the 40S intersubunit-face, and adopts a conformation that differs from that observed during ribosomal-recycling. Most importanly, we discover an uncharacterized factor at the platform of the 40S complexes, termed provisionally ηF. In addition, our study unveiled a structure of 40S dimers, possibly kinetoplastid-specific. Notably, ηF overlaps with key 40S dimers contacts, suggesting a role in preventing their formation. Finally, our structures show that ηF clashes with the eIF2-ternary complex, indicating that ηF must be released before the assembly of the (pre)initiation complex. Our work represents the first structural insights of kinetoplastids-specific aspects of translation regulation.

Keywords: ABCE1, translation initiation, Trypanosoma cruzi, ribosome, cryo-EM

## Introduction

The advances over the last years in determining the structures and molecular mechanisms of the eukaryotic protein factories have revealed differences among species. As early diverging eukaryotes, kinetoplastids, unicellular eukaryotic pathogens responsible for a wide range of human, animal and plant diseases, display several unusual features within their translational machineries. Recent cryo-electron microscopy (cryo-EM) reconstructions of the kinetoplastidian ribosomes have highlighted some structural differences when compared to other eukaryotes, including the size and arrangement of the eukaryotic-specific ribosomal proteins and the species-specific ribosomal RNA expansion segments (Hashem et al., 2013a; Liu et al., 2016; Shalev-Benami et al., 2016b; Zhang et al., 2016b). In addition, although orthologs of most of the canonical translation initiation factors are encoded by kinetoplastidian genomes (Bindereif, 2012; Ivens et al., 2005), several genetic and biochemical reports revealed that they substantially diverge when compared to their eukaryotic counterparts on the basis of amino sequence similarity and functional activity (Bindereif, 2012; Freire et al., 2011, 2014; Meleppattu et al., 2015; Rezende et al., 2014; Yoffe et al., 2004, 2009; Zinoviev and Shapira, 2012; Zinoviev et al., 2011). Moreover, all mRNAs from kinetoplastids harbor a 39-nucleotide conserved spliced leader (SL) that is trans-spliced onto their 5' ends (Perry et al., 1987) and confer them a chemically hypermodified 5'-cap structure known as cap-4 (Bangs et al., 1992). Both the 5'-SL and the cap-4 are essential for mRNA translation in kinetoplastids (Zamudio et al., 2009; Zeiner et al., 2003). Despite certain progress in determining the molecular basis of kinetoplastidian translation, deeper structural studies are needed to better understand the molecular rationale of their species-specific ribosomal features in orchestrating protein synthesis, a key step towards developing safer anti-kinetoplastid therapies.

Examples of uncharacterized fundamental ribosomal processes in kinetoplastids are numerous, such as ribosomal recycling and translation initiation. Ribosome recycling in eukaryotes is an essential process that triggers splitting of post-terminating ribosome complexes (post-TCs) into 40S and 60S subunits for readying them for a new cycle of translation initiation (Pisarev et al., 2010). Upon stop-codon recognition, the release factor 1 (eRF1) (Nakamura and Ito, 2003) catalyzes hydrolysis of the ester bond that links the nascent polypeptide chain to P-site tRNA, stimulated by the GTPase activity of eRF3, which couples stop-codon recognition and P-site-bound peptidyl-tRNA hydrolysis to ensure rapid and efficient peptide release (Jacobson, 2005; Salas-Marco and Bedwell, 2004). In eukaryotes, besides the known termination factors, the highly conserved and essential ATP-binding cassette (ABC)-type NTPase Rli1 (yeast)/ABCE1 (mammals) also plays a key role in eRF1-associated post-TC

recycling (Khoshnevis et al., 2010; Pisarev et al., 2010). After peptide release, eRF1 remains bound to post-TCs and, together with ABCE1, splits them into free 60S subunits and tRNA- and mRNA-associated 40S subunits (Pisarev et al., 2010; Shoemaker and Green, 2011).

It has been previously shown, that ATP-binding cassette proteins bind and hydrolyze (at similar rates) any natural nucleotide triphosphates (Fendley et al., 2016). The binding of nucleotide triphosphates triggers the fully closed conformation, in which nucleotide binding domains (NBDs) dimerize (Fendley et al., 2016). Notably, comparative sequence analyses have shown that this protein is one of the most conserved in evolution, especially their NBDs, which share 90% amino acid identity across all eukaryotes (Gabaldón and Huynen; Kerr, 2004). Because RNase L is only found in mammals, the evolutionary conserved role of ABCE1 must be different from its inhibition of RNase L (Gabaldón and Huynen; Kerr, 2004).

So far, the current model of the ABCE1/Rli1 function in translation is attributed to its binding to eRF1- associated post-TCs to promote their splitting (Pisarev et al., 2010). However, several reports have provided solid evidence that ABCE1/Rli1 associates with translation initiation factors, as well as with the small ribosomal subunit (eukaryotic 40S and archaeal 30S) (Andersen and Leevers, 2007; Chen et al., 2006; Dong et al., 2004; Heuer et al., 2017; Kispal et al., 2005; Pisarev et al., 2010; Yarunin et al., 2005; Zhao et al., 2004) pointing to its underexplored role in translation initiation. For example, the *Drosophila melanogaster* ABCE1 ortholog (called Pixie) binds 40S in an ATP-dependent manner, as well as eIF3 and eIF5, and it is required for normal levels of translation initiation (Andersen and Leevers, 2007). Moreover the essential ABCE1 yeast ortholog Rli1 co-purifies with eIF3, eIF2 and eIF5 and it associates with 40S *in vivo* (Dong et al., 2004; Yarunin et al., 2005). In addition, binding analyses strongly suggest that Rli1 binds directly to eIF3a, eIF3j and eIF5 (Gavin et al., 2002; Kispal et al., 2005; Yarunin et al., 2005). Furthermore Rli1 was shown to be required for proper 43S preinitiation complex assembly (Dong et al., 2004). Additional data indicate that the vertebrate and *Caenorhabditis elegans* ABCE1 orthologues also play a role in translation initiation (Chen et al., 2006; Roy et al., 2005; Zhao et al., 2004). Consistent with these previous reports, it was observed that the AMP-PNP-bound form of ABCE1 efficiently binds 43S complexes (Pisarev et al., 2010). However, the molecular rearrangement induced by NTP-hydrolysis on ABCE1, its molecular mechanism and its function in translation initiation in eukaryotes remain unclear. In kinetoplastids more specifically, none of the above-mentioned processes has been characterized biochemically or structurally.

Here, we present a 5.9Å cryo-EM structure of ABCE1 bound to stalled 40S in near-native conditions from cell extracts (through GMP-PNP) from *Trypanosoma cruzi*, the

kinetoplastidian pathogen that causes Chagas disease. This complex shows that ABCE1 interacts with 40S subunits under a conformation that differs substantially from that displayed in the context of the recycling ribosomal complexes (Becker et al., 2012). These particles along with the non-stalled 40S complexes (native 40S subunits solved at an average resolution of 6.75 Å), include one additional uncharacterized factor that we termed provisionally as η (ETA in Greek) factor (ηF), which binds the intersubunit side of 40S platform and extends near the P-site. A second complex at low resolution represents the first structure of 40S dimers from a kinetoplastid organism (the presence of 40S dimers was reported in early studies in mammals (Goumans et al., 1980; Henshaw et al., 1973; Peterson et al., 1979). This work provides the first structural insights into kinetoplastid-specific aspects of translation regulation and highlights relevant mechanistic differences during an early stage of initiation when compared to other eukaryotes.

## Results

### ABCE1-bound 40S complex purification from *Trypanosoma cruzi*

Near-native ABCE1-containing 40S complexes were isolated *ex vivo* from a lysate of *Trypanosoma cruzi* epimastigotes treated with non-hydrolysable GTP analogue GMP-PNP (Figures 1A and Supplementary S1). Whole *T. cruzi* cell extracts treated or not with GMP-PNP (yielding stalled and unstalled native 40S complexes, respectively) were fractionated by sucrose density gradients. Pelleted 80S and 40S complexes were characterized by SDS-PAGE and non-denaturing agarose gel electrophoresis for checking protein patterns and RNA quality, respectively (Figures 1A and Supplementary S1; see Methods).

In the absence of GMP-PNP, ABCE1 was undetectable in 80S or 40S (Figure 1A and Supplementary Table 1). In contrast, a strong enrichment of ABCE1 in the 40S-related fractions was observed only when ABCE1 is stalled by the addition of GMP-PNP (Figure 1A and Supplementary Table 1). The 40S binding of ABCE1 in the presence of GMP-PNP was confirmed by gel filtration chromatography of high molecular weight complexes pelleted from 40S-related peak fractions after sucrose gradient fractionation and subsequent liquid and in-gel MS/MS analyses (Figures 1A, Supplementary S1 and Table S1).

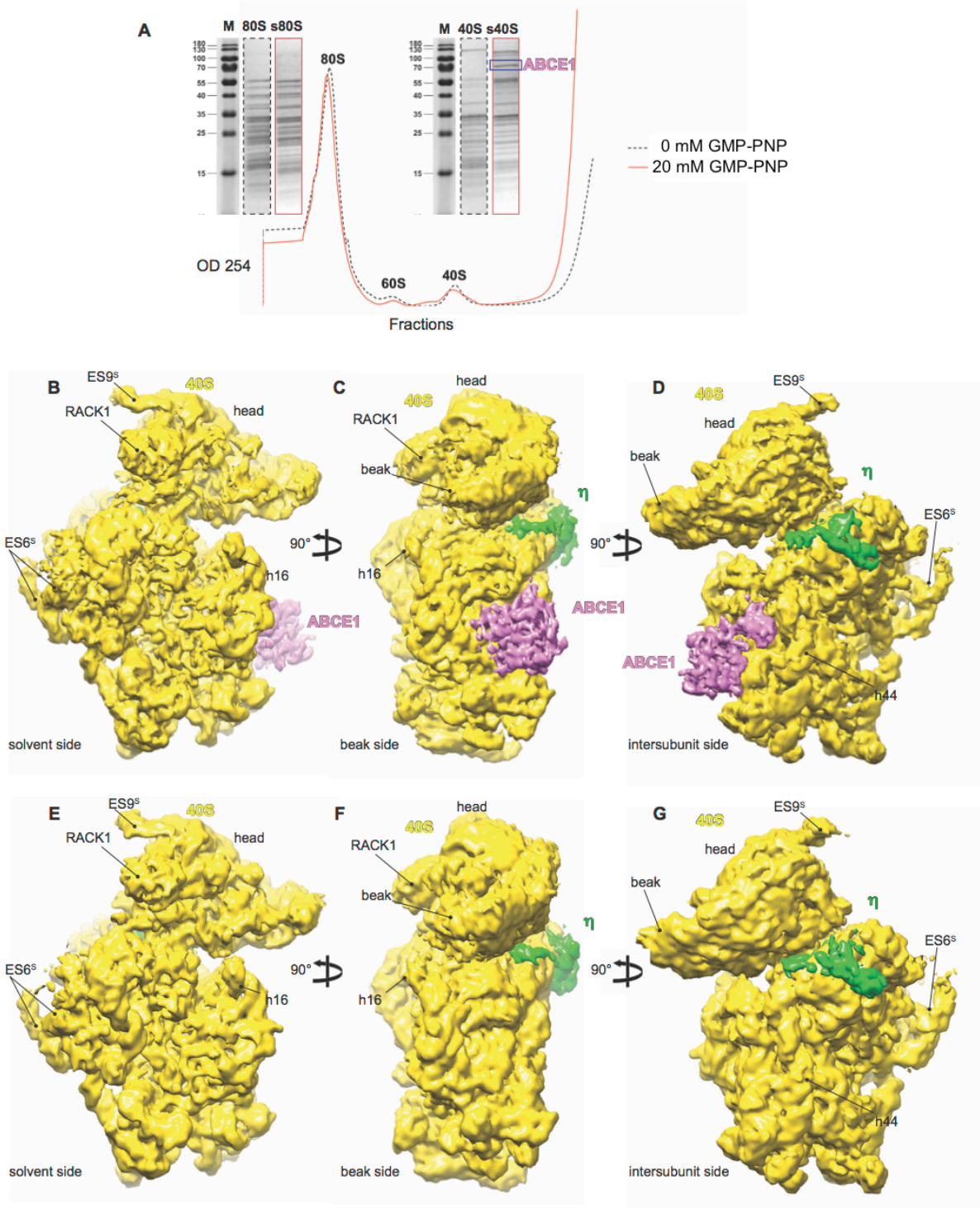


**Figure 1: SDS-PAGE/MS analyses and cryo-EM reconstructions of the native and stalled 40S complexes from T. cruzi.** (A) Absorbance profile at 254 nm from 10-30% sucrose gradient, displaying peaks for native and stalled 40S, 60S and 80S complexes. Red curve indicates stalled complexes and dashed curve indicates native complexes. Stalling was accomplished by adding 20 mM of GMP-PNP to *T. cruzi* lysate. SDS-PAGE analysis of the stalled 40S (s40S) fractions revealed the presence of some additional bands when compared with native 40S fractions. By mass spectrometry analysis we identified ABCE1 on the additional band of higher intensity, see Supplementary Table 1. (B-D) Cryo-EM structure of the *T. cruzi* ABCE1-bound 40S complex viewed from different orientations. Density segments correspond to 40S (yellow), ABCE1

(pink) and ηF (green). (E-G) Cryo-EM structure of the *T. cruzi* native 40S complex viewed from different orientations. Density segments correspond to 40S (yellow) and ηF (green). All the displayed maps are low-pass filtered according to the average resolution of each reconstruction.

GMP-PNP has been widely used as a tool in many biochemical and structural studies as it blocks translation initiation, inducing the accumulation of 48S initiation complexes stalled at the AUG initiation codon (Anthony and Merrick, 1992; Hershey and Monro, 1966). Indeed, trapping eIF2 by GMP-PNP synchronizes initiation complexes at a late stage of translation initiation, just before its dissociation and subsequent 60S joining (Simonetti et al., 2016). Consistent with our previous report in mammals (Simonetti et al., 2016), we found that adding GMP-PNP to cell extracts promotes an enrichment of eIF3 in the pelleted 40S-related fractions. However, the spectral counts for other 48S IC-related factors, remain low when compared to the levels observed in mammals (Simonetti et al., 2016) (Supplementary Table 1; see ref. (Lundgren et al., 2010) for details on semi-quantitative MS/MS analysis). These results show that stalling with GMP-PNP does not yield substantial amounts of 43S/48S initiation complexes and strongly suggest the existence of other uncharacterized mechanisms preceding the complete 43S/48S IC assembly in kinetoplastids. Finally, our data suggest that ABCE1 may associate to 40S subunits to play an NTP-hydrolysis dependent role during an early-stage of the preinitiation complex formation, even before the association of the eIF2-tRNAiMet-GTP ternary complex (eIF2-TC) (Andersen and Leevers, 2007; Dong et al., 2004), and provide an additional argument in favor of the underexplored role of ABCE1 in translation initiation.

**Model of ABCE1-40S complexes**

Using single particle cryo-EM analysis (see Methods and Supplementary Figure S2) on the pelleted 40S-related fractions isolated from GMP-PNP treated lysates, we obtained a major class comprising 40S ribosomal complexes (class III, 37% of the total data set), yielding a cryo-EM reconstruction with an average resolution of 5.9Å (Supplementary Figures S2A and B). Interestingly, a fraction of 40S dimers (~2%) was also sorted out during our data-processing, in addition to 80S elongating complexes (~9%) (Supplementary Figure S2B). The cryo-EM map of the major class corresponding to stalled 40S complexes was analyzed (Figures 1B-D). As a reference control we have also purified and analyzed by cryo-EM native 40S complexes, which were not stalled by GMP-PNP and the 40S fractions were simply collected and imaged. When compared to the cryo-EM reconstruction of native 40S subunits (Figures 1E-G), our segmentation indicates clearly the presence of an extra mass of density at the translation

GTPase binding-site. Surprisingly, we also observed a second mass of density at the intersubunit face, on the platform, presenting partially a helical shape that is also found in native 40S subunits (Figure 1).

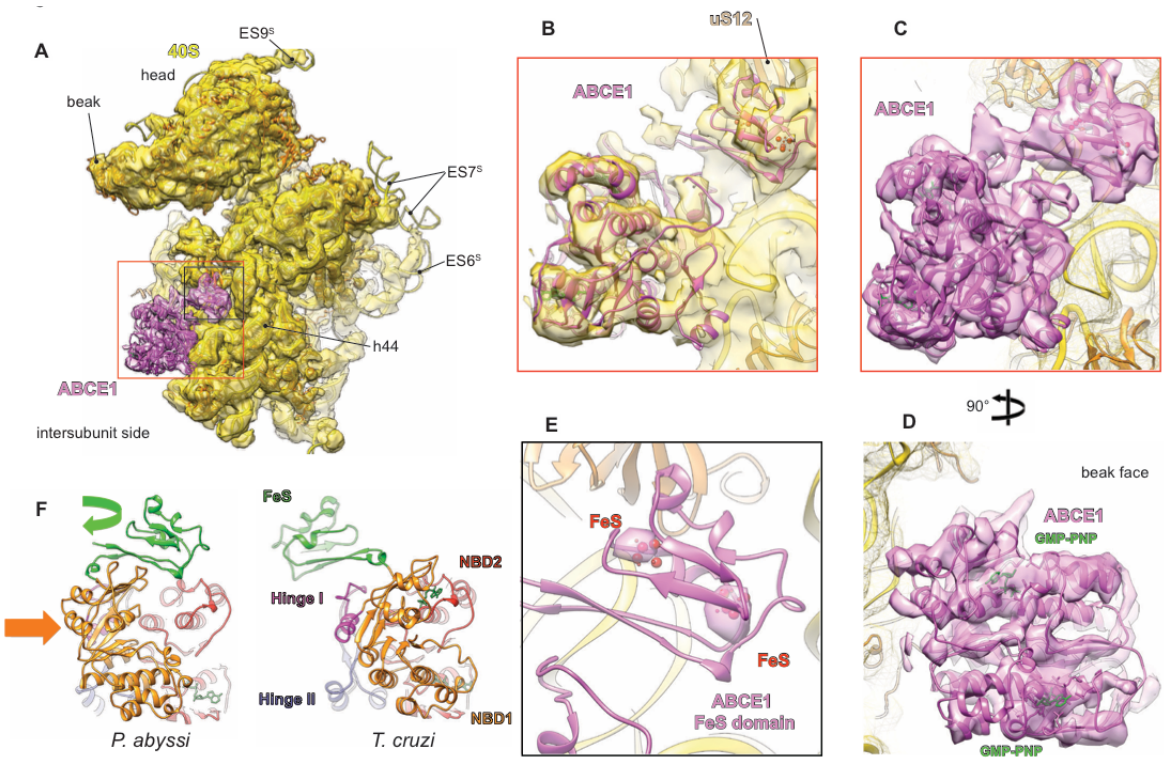


**Figure 2: Atomic model of *T. cruzi* ABCE1-bound 40S fitted into its cryo-EM segmented map.** (A-C) Atomic model of ABCE1-bound 40S, focused on the ABCE1 (pink), seen from the intersubunit side, in the context of the segmented (A and C) and unsegmented map (B). (D) Atomic model of GMP-PNP bound ABCE1 fitted into 40S segmented Cryo-EM map, viewed from beak face. (E) Close-up view of the atomic model of the iron–sulfur (FeS) cluster domain of ABCE1, fitted into its cryo-EM segmented map, illustrating the density clouds (pink) corresponding to the two FeS clusters of ABCE1 (red) shown at high contour levels. (F) Conformational changes undertaken by the *T. cruzi* GMP-PNP-bound ABCE1 by comparison to X-ray structure of *Pyrococcus abyssi* ADP-bound ABCE1(Karcher et al., 2008). The FeS domain from *T. cruzi* ABCE1 rotates by ~180° when compared to *P. abyssi* crystal structure. The green arrow shows the direction of the FeS domain rotation. The orange arrows indicate the direction of NDB2 domain movement. Each domain is colored variably.

We unambiguously assigned the density located at the GTPase binding-site to ABCE1 on the basis of its size and shape (Figures 2A-C), as it displays a local resolution sufficiently high to clearly distinguish secondary structure elements (Figures 2A-E and Supplementary S2A). Most importantly, our MS/MS and SDS-PAGE analyses show the presence of high levels of ABCE1 on stalled 40S when compared to the control native 40S (Supplementary Table 1 and Figure 1A). Using available crystal structures of archaeal ABCE1 from *Pyrococcus abyssi*

(Karcher et al., 2005) and taking into account the degree of conservation of this protein among archaeal and eukaryotic species (more specifically *T. cruzi*, see alignment in Supplementary Figure S3) we assigned and placed the functional domains of ABCE1 into our cryo-EM map, including its N-terminal [4Fe-4S] cluster (FeS), both NBDs and hinge regions (Figure 2). A near-complete atomic model of *T. cruzi* ABCE1 (side chains are not resolved at this resolution) was built by homology based on its archaeal counterpart X-ray structure (See Methods). After fitting *T. cruzi* ABCE1 model into its cryo-EM corresponding density, we can conclude that in our structure *T. cruzi* GMP-PNP-bound ABCE1 adopts a fully closed conformation, similarly to that observed for the yeast AMP-PNP-bound ABCE1 (Heuer et al., 2017), and differs to that found in eukaryotic and archaeal recycling complexes (Becker et al., 2012). According to previous work (Becker et al., 2012), the adoption of the fully closed ATP-bound state would demand a shift of the FeS cluster domain of 8 Å to impair a steric clash with the NBD2. In our structure, the FeS cluster domain of ABCE1 displays a very ample rotation that not only impairs the above-mentioned steric clash but also enable it to acquire a proper orientation for binding to 40S (Figure 2F). Importantly, the high electron density of the FeS spots of ABCE1 are distinctly discernible even at high-density thresholds in our map reconstruction (Figure 2E), validating our model that displays a ~180° rotation of the N-terminal FeS domain with respect to the archaeal crystal structure of ABCE1 (Figures 2F and Supplementary S4). This conformation is similar to that displayed in the binary ABCE1/40S complex from yeast (Heuer et al., 2017). Consistently, a previous report has demonstrated that the FeS cluster domain of archaeal AMP-PNP-stalled ABCE1 is essential for 30S small subunit interaction (Barthelme et al., 2011). On the basis of the aforementioned observations, the 40S particles in our reconstruction may represent the first structure reported so far of a very early stage of preinitiation, just before the beginning of the assembly of the preinitiation complex from the eukaryotic pathogen *T. cruzi*.

**ABCE1-40S ribosomal subunit interaction**

After fitting our model in its corresponding density, we determined the residues of ABCE1 that directly contact components of the 40S, at the intersubunit face where translation-related GTPases typically bind (Figures 3A-E). Although the resolution is insufficient to display side-chains and single bases, it was possible to pinpoint an ensemble of residues at the interface between ABCE1 and the 40S. Direct 40S interactions of the FeS cluster domain, NBD1 and hinges I and II from ABCE1 were observed (Figures 3A-E). Hinges I and II contact h14 and h8 respectively (Figure 3C). These interactions occur via highly conserved regions

within these hinge domains spanning the residues R315-N319 of hinge I, and residues R575, N586, S590 and K592 from hinge II (Figure 3C). The highly conserved HLH domain within the NBD1 of ABCE1 contacts the h5–h15 junction vicinity, whereas residues P111, Q113, K273 and E274 (less conserved when compared to archaeal ABCE1), provide additional contacts with the C-terminal part of ribosomal protein eS24 spanning amino acids 128-134 (Figure 3E), which is less conserved when compared to other eukaryotes and absent in archaea.

In our resulting ABCE1 model, hinges I and II, as well as its NBD1, create a wide contact area which seems to be essential for 40S binding. These results are consistent with the findings of a previous report showing that the same conserved residues are engaged in ABCE1/small ribosomal subunit interaction during eukaryotic and archaeal ribosome recycling (Becker et al., 2012). However, in spite of this similarity, the FeS cluster domain of NTP-closed ABCE1 from *T. cruzi* interacts directly with the small ribosomal subunit (Figure 3D), similar to that proposed in the binary ABCE1/40S complex from yeast (Heuer et al., 2017). Our results reveal that a drastic conformation switch of ABCE1 regarding the FeS cluster is also required for small ribosomal subunit binding in kinetoplastids. Accordingly, the *T. cruzi* FeS-containing N-terminal core of ABCE1 is rotated by ~180° away from the rest of the structure towards the intersubunit side of 40S, particularly the h44 and the ribosomal protein uS12 (Figure 3D). We observed that residues surrounding L69 of ABCE1 interact with a well-conserved loop within the ribosomal protein uS12 created by residues I75 and L74. It seems that the FeS cluster domain/uS12 interaction network observed in our structure is similar to that reported in archaea using cross-linking and a low resolution cryo-EM reconstruction (Kiosze-Becker et al., 2016), except that in our model the residues of uS12 implicated in this interaction are different. In addition to these contact points, we found that the residues S28-S31 of ABCE1 face the nucleotides G2196/G2195 and A2125/U2126 from h44 (Figure 3D). Such observations offer new molecular basis for the role of ABCE1 right after ribosomal subunit dissociation upon recycling and even perhaps during the assembly of initiation complexes.

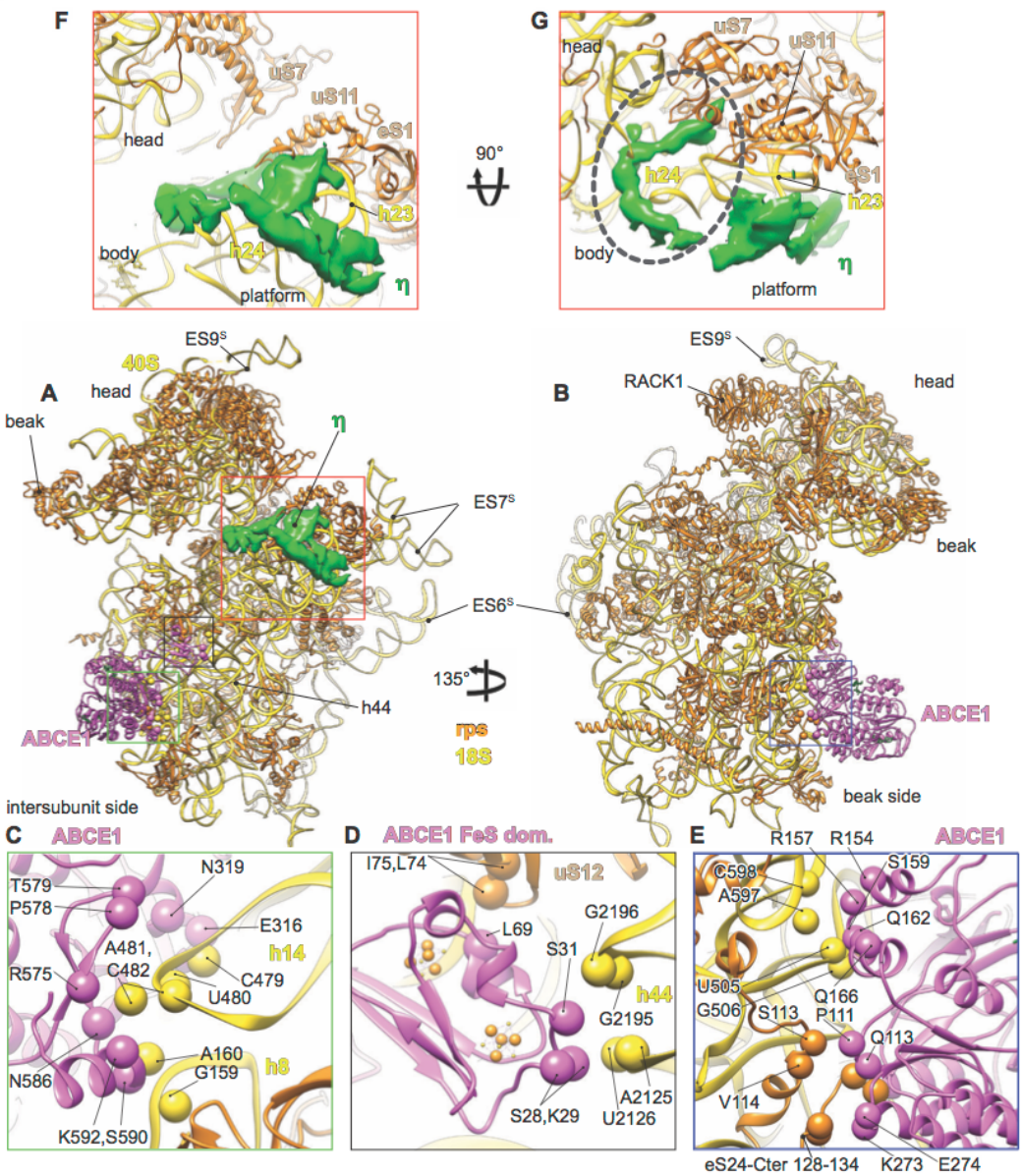


**Figure 3: Atomic model of *T. cruzi* ABCE1-bound 40S and ηF segmented density.** (A-B) Atomic model of ABCE1 bound at the intersubunit side of 40S, seen from the intersubunit and beak sides. (C) Close-up view of the interaction points between ABCE1, h8 and h14. (D) Close-up view of the interaction between the FeS-containing N-terminal core of ABCE1 and h44 and the ribosomal protein uS12. (E) Close-up view showing the interaction points between the NBD1 of ABCE1 and the h5–h15 junction vicinity and C-ter of eS24. (F-G) Close-up view showing the ηF segmented density (green) and its binding to h23, h24 and 40S ribosomal proteins. Dashed black circle highlights a discontinued density fragment that could be part of the main core of ηF in both the native and stalled 40S maps.

**An uncharacterized factor binds to the platform of native and stalled 40S subunits**

Notably, we observe an uncharacterized factor, termed here η factor ("eta" in Greek; ηF), which binds to the intersubunit face of the 40S platform in close proximity of the E site, of both the native and stalled 40S complexes. Its conformation does not display significant changes between the two complexes, suggesting that this factor is naturally associated under these particular growing conditions to 40S subunits *in vivo* and its binding does not depend on the presence of GMP-PNP (Figure 1). The overall shape is similar in both maps and its location could indicate a critical role in regulation of translation in kinetoplastids (Figures 1 and 3). At this resolution, we can distinguish putative helical and β-sheet domains that primarily bind to the intersubunit face of h24 and h23 (Figure 3F). Because the densities corresponding to ηF are similar in both complexes, we describe its features in the higher resolution stalled 40S complex (Figures 3F and G). The main helical domain of this density displays a mast-like shape and binds to h24, protruding towards the head of 40S. A smaller putative multi-helical domain is located in close proximity to h23 and ribosomal protein eS1 and establishes numerous contact points on a wide area of the platform (Figures 3F and G). However, we will not dwell more on the structure and interactions of ηF as we have not identified it yet with certainty and we do not dispose of its high-resolution structure. Remarkably, we see an additional density discontinued from the main one, which threads into the mRNA channel from the intersubunit interface near h24 till the mRNA channel exit. This density fragment circumvents a loop of h24 and goes towards ribosomal proteins uS7 and uS11 (Figure 3G, dashed black circle). Despite the fact that at low contour density levels this density fragment in the mRNA channel seems to be connected with the main core of ηF in both the native and stalled 40S maps, we cannot unambiguously affirm that is part of it. Strikingly, in both 40S complex reconstructions the mRNA channel latch is fully opened (Figures 1 and 3), as observed previously in py48S scanning initiation complexes (Llácer et al., 2015). We believe that this conformation should allow mRNA to be loaded directly to the mRNA channel and may represent an early state of initiation in which 40S subunits are ready to bind the 43S initiation factors.

Even if the location of ηF could have lead us to think that this factor is a peripheral extension of one of the eIF3 subunits (e.g. eIF3c-NTD, eIF3a-CTD; see Llácer et al., 2015 for details) or a ribosome assembly factor (e.g. Dim1, Rio2; see Strunk et al., 2011 for details), our MS/MS data did not reveal any strong evidence that could validate this assumption (Supplementary Table 1), as because none of these eF3 subunits were detected in sufficient abundance in the native 40S. It is worthy to mention that the NTD of eIF3c and the CTD eIF3a from *Trypanosoma cruzi* are shorter and present a low level of conservation when compared to

other eukaryotes, and the above-mentioned eukaryotic ribosomal assembly factors (Strunk et al., 2011) are not encoded by the *Trypanosoma cruzi* genome (except Dim1, but it is absent in our MS/MS data from both 40S complexes). Moreover, the fact that the visible domains of ηF interact with the small subunit rRNA moiety belonging to h24 and h23 (Figures 3F-G) strongly suggests that this protein could possess RNA binding-like motifs. However, more evidence is required to attribute it with certainty to any of the potential candidates found in our MS/MS data, leaving the way wide open for future research.

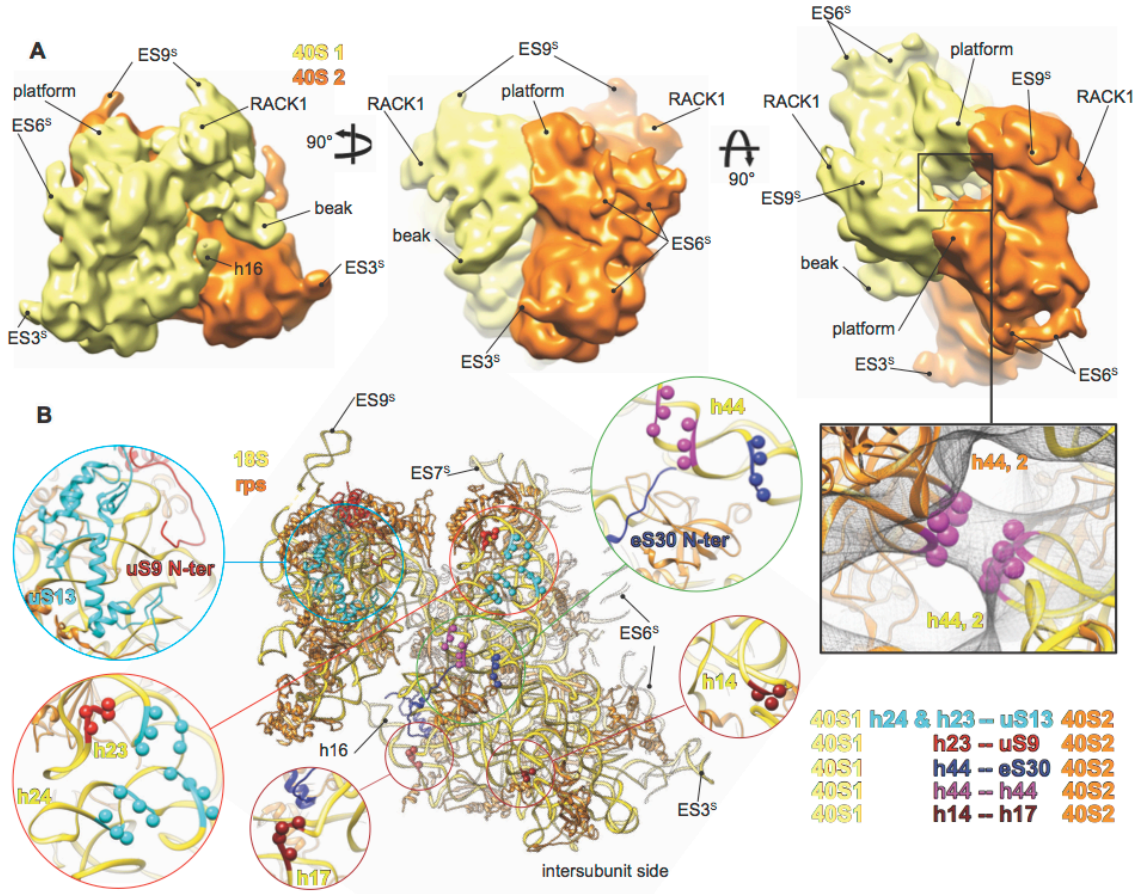


**Figure 4: Cryo-EM reconstruction of 40S dimers from *T. cruzi* and their related intersubunit bridges**
(A) Segmented cryo-EM map of *T. cruzi* 40S dimers viewed from different orientations. Density segments correspond to 40S subunit 1 (yellow) and 40S subunit 2 (orange). All the displayed maps are low-pass filtered according to the average resolution of each reconstruction. (B) Atomic model of 40S dimers and close-up view showing the five intersubunit bridges implicated in 40S dimerization. The bridges were highlighted with variable colors. The atomic model was simply generated by rigid-body-fitting the previously published model of the 40S subunit from *T. brucei* into the *T. cruzi* 40S densities.

## η factor interferes with 40S dimers

The intrinsic tendency of naked 40S subunits to dimerize has been reported since the early 70's (Goumans et al., 1980; Henshaw et al., 1973; Peterson et al., 1979). After particle sorting of the cryo-EM data from native and stalled 40S fractions, we obtained minor classes that represent 40S dimers (Supplementary S2). The binding pattern observed displays an antiparallel dimeric geometry that engages mainly the head and the platform of each 40S subunit symmetrically (head of monomer 1 "40S1" on platform of monomer 2 "40S2" and vice-versa) (Figure 4A). Our structure enables us to pinpoint five 40S intersubunit bridges implicated in 40S dimerization (Figure 4B). All these bridges are symmetrical. The first bridge is the strongest of all and is formed by h23 and h24 from 40S1 with uS13 and uS9 from 40S2, and vice-versa (Figure 4B). The second bridge is formed by 40S1 h44 and ribosomal protein eS30 on 40S2 (Figure 4B), and vice-versa. The third bridge is very peculiar as it is formed by 40S1 h44 and 40S2 h44 (Figure 4B) at the A-site and represents the central axis of these dimers. This bridge involves two large RNA/RNA interactions that could engage up to five nucleotides from each helix. Finally, the fourth bridge is formed by 40S1 h14 and 40S2 h17 (Figure 4B), and vice-versa.

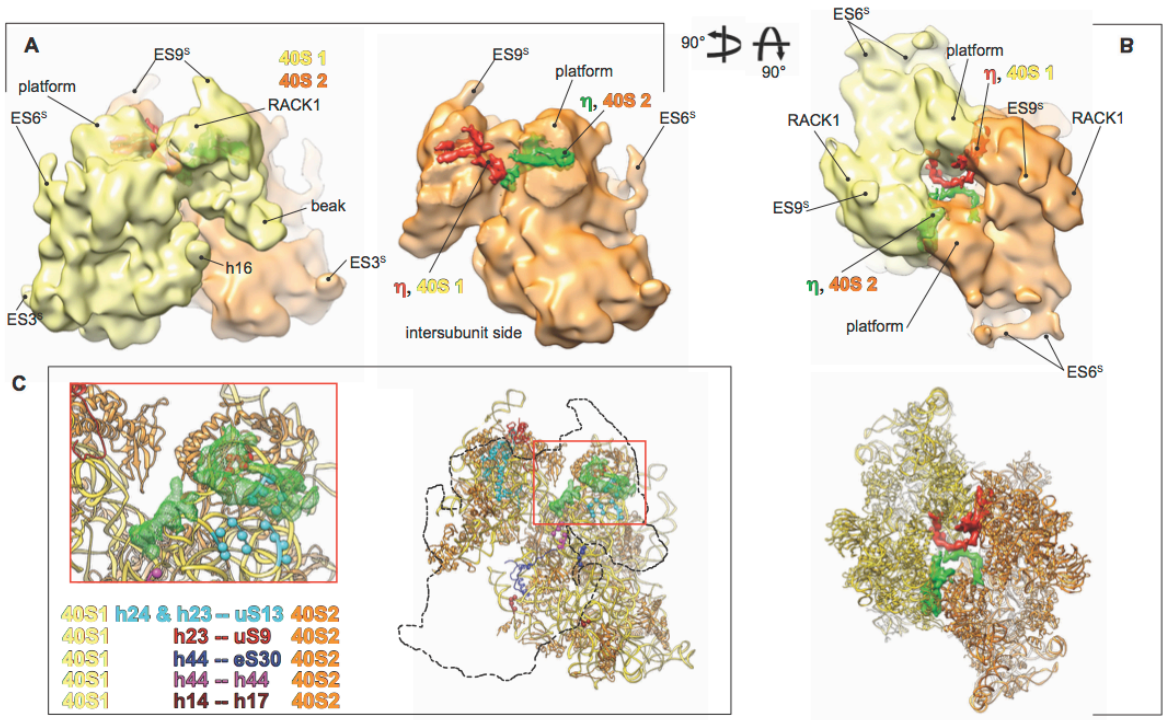


**Figure 5: The overlapping of ηF with *T. cruzi* 40S dimer intersubunit bridges**. (A-B) Docking of ηF (colored in red and green for 40S 1 and 40S 2, respectively) onto the cryo-EM reconstruction of *T. cruzi* 40S dimer, viewed from different orientations highlighting ηF steric clash with the 40S dimer main bridges located on the head and the platform. (C) Close-up view showing the ηF segmented density and its binding to h23, h24 and 40S ribosomal proteins, which clashes with the main bridges involved in 40S dimerization.

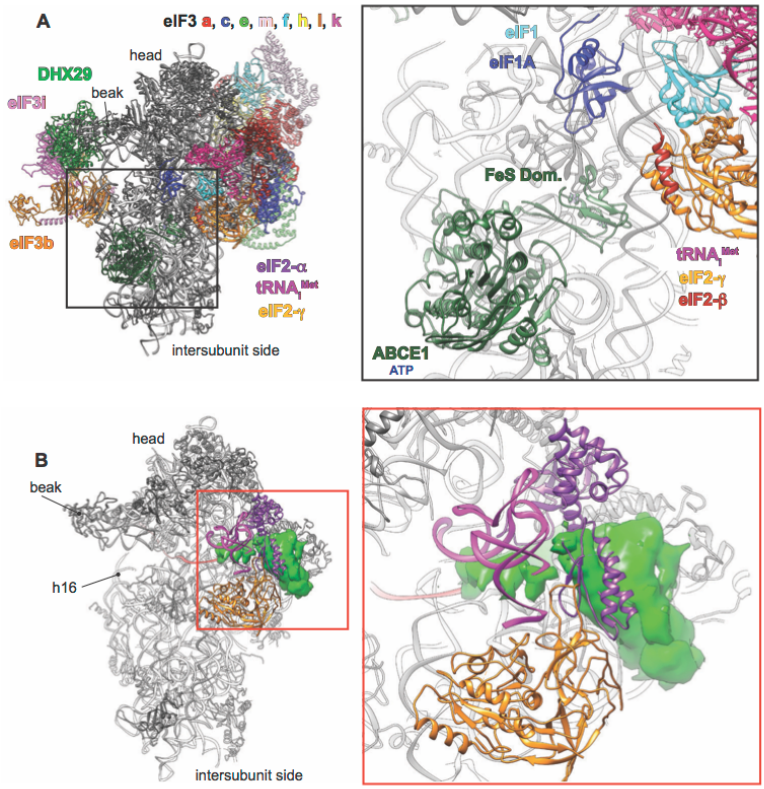


**Figure 6: Docking of ABCE1 and ηF on mammalian 43S PIC and 48S IC structures.** (A) *T. cruzi* GMP-PNP-bound ABCE1 atomic model (green), was docked on mammalian 43S PIC (des Georges et al., 2015) at the exact position observed in our stalled ABCE1/40S complex (left). Close-up view (right) showing that none of the constituents of the 43S PIC occupies the 40S GTPase binding-site, where ABCE1 interacts, hence all these factors can in theory co-exist at this stage of initiation. (B) *T. cruzi* ηF related density from our cryo-EM reconstruction was docked on mammalian 48S IC(Simonetti et al., 2016) (left). Close-up view (right) showing the clash between *T. cruzi* ηF (green) and eIF2-TC from mammalian 48S initiation complex.

Early reports revealed that native mammalian 40S subunits have low tendency for dimerization. Nevertheless salt washed native 40S subunits tend to dimerize strongly, suggesting that extra proteins serve to maintain them as individual subunits (Henshaw et al., 1973). However, in *T. cruzi* 40S dimers we observe no extra electron density on the 40S dimer interface suggesting that intersubunit binding is not mediated by any factors, at least in the present complexes. Indeed, we found that ηF was not present in 40S dimers and its location, within both the native and stalled 40S monomers, occludes the first and strongest 40S dimer bridge on the head and the platform (Fig. 5). Taken together, our data indicate that ηF could play the important role by preventing the 40S dimerization in kinetoplastids and make them available to commence the initiation process.

**Discussion**

Our reconstructions of native and stalled 40S complexes from *Trypanosoma cruzi* lysates enable us to unveil the presence of a persistently bound factor named here ηF, which displays a ling α-helix and can interact with the intersubunit side of the 40S platform, close to the E site, establishing protein/rRNA interactions with h24 and h23. The main helical domain of this molecule points out from the bottom of the platform towards the mRNA channel latch, and its position interferes with the formation of two key contact points that are involved in 40S dimerization. Notably, 40S dimers were sorted out after processing cryo-EM images from native and stalled 40S samples and their binding geometry is well defined, which seems to differ from that of mammalian 40S dimers, as to date, no clear pattern of assembly of 40S dimers was definable from structural data of the numerous initiation-related studies by cryo-EM. Our results point out that this dimeric arrangement may be specific for kinetoplastids.

Taking into account that the density assigned to ηF interferes with the formation of 40S dimer, we believe that this factor can play a role in preventing the latter. Hence, 40S dimer splitting could be a critical regulation checkpoint in kinetoplastidian mRNA translation, in which 40S subunits are made available for initiation after their "parking" as dimers upon ribosomal recycling. This may suggest that this process could be one of the first stages of the initiation cycle in kinetoplastids. Previous reports have shown that factors involved in 43S PIC formation can promote dissociation of 40S dimers (Goumans et al., 1980). However, in our case, since no eukaryotic initiation factor was consistently detected in the native 40S and none of them could be attributed to the intersubunit density named ηF, we could not make a link between those initiation factors and 40S dimers.

Unexpectedly, once we have stalled our *T. cruzi* 40S complexes with GMP-PNP we were not able to obtain *bona fide* 48S ICs as widely reported by us an others in the literature for other eukaryotes (Anthony and Merrick, 1992; Gray and Hentze, 1994; Simonetti et al., 2016). Instead of inducing the formation of a canonical 48S IC by the presence of GMP-PNP, in which the eIF2-TC is at the AUG codon, we captured a snapshot of the kinetoplastid translation cycle that involves a 40S ribosomal complex with the recycling NTPase ABCE1.

Notably, this stalled ABCE1-containing 40S complexes highly resemble the native 40S subunits in the way that the ηF-related density is present in both complexes and they display the same fully open conformation of their mRNA channel latches, suggesting that ABCE1 binding does not trigger substantial rearrangements in the native 40S subunits.

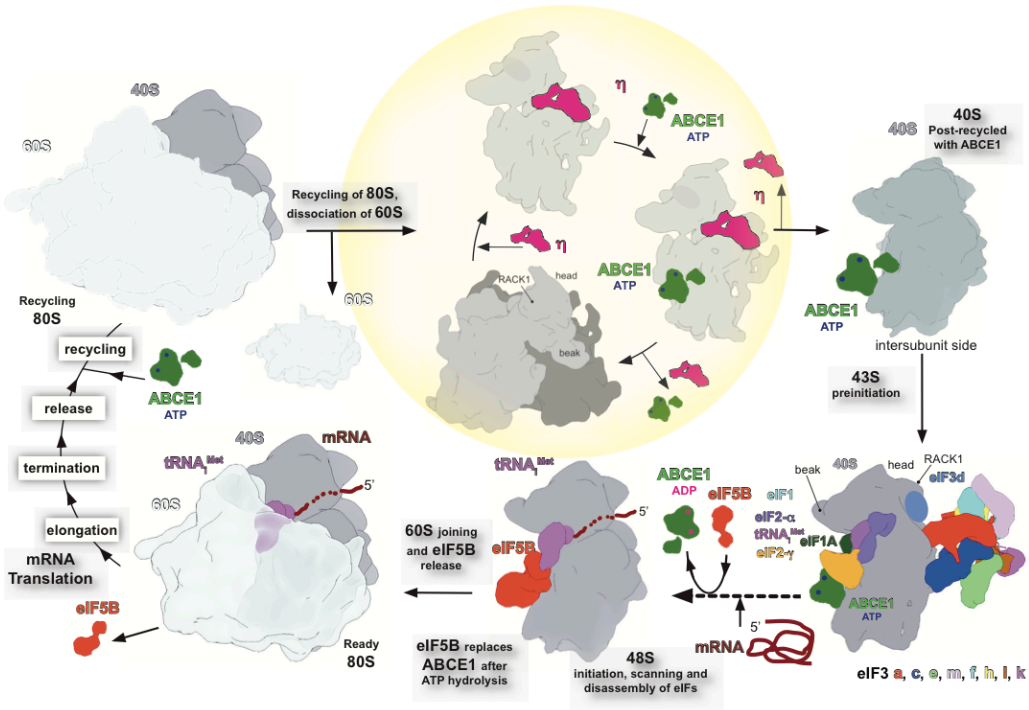


**Figure 7: Model of ABCE1 and ηF activities during translation initiation**. We propose that the kinetoplastid-specific factor, ηF (depicted in pink), is one of the first factors to participate in early kinetoplastid translation initiation, by preventing 40S dimerization to make 40S subunits available for initiation. After 40S dimer splitting, when bound to NTP, ABCE1 (green) can take its place on 40S, and together with ηF, act as critical ribosomal anti-association factors. Due to the presence of ηF, we believe that this early stage of the initiation cycle is specific for kinetoplastids (yellow bubble). After this stage, ηF should leave the complex since it clashes with the eIF2-TC. Therefore, after ηF departure, ABCE1 would be the only anti-association factor preventing the premature 60S subunit joining during 43S PIC assembly. We believe that the 40S/ABCE1 complex is common for all eukaryotes because ABCE1 does not clash with any 43S-related initiation factor, at least at this stage of the process. Finally, we propose that eIF5B (red) replaces ABCE1 on the 40S before 60S subunit and eIF5B release at a later stage of the process.

It was observed that the presence of extra-ribosomal intersubunit binding proteins are required to prevent either the re-association of freshly recycled 60S with 40S subunits or the formation of 40S dimers, suggesting that ribosomal dissociation is only transient in nature and re-association is the by-default ribosomal behavior (Goumans et al., 1980; Henshaw et al., 1973). It is plausible to presume that in the presence of NTP, ABCE1 can take its location on the GTPase binding pocket of the 40S, and together with ηF, to play a role in early initiation as probably the first anti-association factors of ribosomal subunits. Thus, with the prevention of premature 60S joining and 40S dimerization, the 43S preinitiation complex can assemble efficiently without being out-competed by naked ribosomal subunits. Moreover, the 43S PIC assembly cannot be hindered by ABCE1, as the latter does not collide with any 43S-related

initiation factor (Fig. 6A). Hence, the lack of the expected 48S IC could not be due to the presence of ABCE1. In addition, it was demonstrated that ABCE1 can bind to *in vitro* and *in vivo* assembled preinitiation complexes (Pisarev et al., 2010, Heuer et al., 2017). Nevertheless, the location of ηF-related density is incompatible with the binding of the eIF2-TC (des Georges et al., 2015; Hashem et al., 2013a; Hussain et al., 2014; Llácer et al., 2015; Simonetti et al., 2016) (Fig. 6B), thus probably inhibiting the formation of a *bona fide* 48S initiation complex in *Trypanosoma cruzi*. This observation suggests that the departure of ηF is required for the proper proceeding of the initiation cycle at this early stage (Fig. 7), highlighting again its relevance in the regulation of kinetoplastid translation. However, the trigger for ηF release from the 40S is completely unknown and future investigation will aim at deciphering this exciting question. Importantly, after the detachment of ηF, ABCE1 would be the only factor preventing the premature 60S subunit joining for the formation of the 43S PIC (Fig. 7). Therefore, we suggest that ABCE1 participates actively in translation initiation in kinetoplastids and more generally in eukaryotes.

Moreover, in the light of our structure, along with the 40S-ABCE1 binary complex (Heuer et al., 2017), we also revisit the interpretation of our late-stage mammalian 48S initiation complex (m48S IC) structure (Simonetti et al., 2016) and we propose a model for ABCE1 in mRNA translation initiation in eukaryotes. Indeed, our cryo-EM reconstruction of a mammalian late-stage 48S initiation complex captured the rabbit 48S initiation complex right after start-codon recognition showing a density on the GTPase binding site of the 40S subunit that was attributed to eIF3i and the RNA recognition motif (RRM) of eIF3g at the exact binding site of ABCE1 in the *Trypanosoma cruzi* ABCE1/40S complex (Supplementary S5). Rigid-body fitting of the atomic model of *T. cruzi* ABCE1 shows a quasi-perfect compatibility with the densities previously attributed to eIF3i and eIF3g. Although both interpretations, eIF3i+g and ABCE1, appear possible at the local resolution of the intersubunit densities (~9 to 10Å), ABCE1 is more consistent with the general shape of the densities and correspond closely to the resolved secondary structure elements.

Since ABCE1 does not clash with any element of the 43S PIC (Fig. 6A), another putative role of ABCE1 in enhancing initiation, which can be complementary to the anti-association function, could be the stimulation of the 43S PIC assembly. Indeed, it was reported previously (Andersen and Leevers, 2007; Chen et al., 2006; Dong et al., 2004) that the depletion of yeast, mammalian and fruit fly ABCE1 induces a reduction in polysomes and accumulation of vacant 80S monomers, typically symptomatic of initiation defects. Thus, it has been suggested that ABCE1 could participate in 43S PIC assembly by interacting directly with

components of the eukaryotic translation initiation machinery such as eIF2, eIF5 and eIF3 (via its subunits a and j) (Andersen and Leevers, 2007; Chen et al., 2006; Dong et al., 2004; Gavin et al., 2002; Kispal et al., 2005; Yarunin et al., 2005), and contribute to their recruitment to the 40S. This role is suggested by our MS/MS data of the stalled 40S in comparison with the native 40S, where we can observe a significant enrichment in eIF3 subunits in the stalled samples. However, kinetoplastidian eIF3 could not be directly observed in our structure, probably due to the putative collision of the eIF3c N-terminal domain with ηF, but it remains associated to ABCE1/40S complexes (probably loosely) after gel filtration assays, (Supplementary Fig. 1 and Table 1, and Fig. 1) suggesting that eIF3 and ABCE1 could coexist within a preinitiation complex in *Trypanosoma cruzi* similarly to other eukaryotic counterparts (Andersen and Leevers, 2007; Chen et al., 2006; Dong et al., 2004; Yarunin et al., 2005). Accordingly, the structure of the stalled 40S initiation complex from GMP-PNP-treated *Trypanosoma cruzi* lysates reported here could readily represent a snapshot of an early eukaryotic initiation complex in which ABCE1 may participate in early 43S PIC assembly by inhibiting ribosomal subunit premature association, and perhaps also by promoting eIF3, eIF5 and eIF2-TC recruitment (Fig 6B and 7). The exact moment when ABCE1 departure occurs remains unknown, however even if it persists after mRNA binding, it should be released upon the binding of eIF5B (as they bind the same GTPase binding site), which promotes the association of the 60S subunit marking the end of the initiation stage by producing the elongation-competent 80S ribosome (Fig. 7).

Finally, it is appropriate to discuss the NTP hydrolysis activity of ABCE1. Since ATP-like enzymes can bind and hydrolyze any natural nucleotide triphosphate and their NBDs are highly similar among all these enzymes, this raises the possibility that under physiological conditions ATP is not always the nucleotide triphosphate bound and hydrolyzed by ABCE1. We observed that after nucleotide triphosphate binding, ABCE1 displays the fully closed conformation, in which its NBDs dimerize and its FeS cluster domain rotates towards the h44 of 40S. According to this viewpoint, ATP-bound ABCE1 binds the 40S complexes till the ATP is hydrolyzed to ADP. ADP-bound ABCE1 has low affinity to the 40S complexes (Pisarev et al., 2010), probably due to the radical conformational changes characterized by the NDBs opening and the rotation of the FeS binding domain away from h44. An important question is whether the ATP hydrolysis is triggered or spontaneous. It is possible that the ATP hydrolysis is triggered by a certain geometry imposed by the stage of the initiation complex assembly such as the reported interactions with eIF3, eIF2 or eIF5 at any stage of the process. However, no

solid arguments seem to support this hypothesis to our knowledge and further studies will have to be conducted in this direction.

We believe that ABCE1 can bind efficiently to the GTPase binding-site on the 40S whenever the latter is available where ABCE1 acts as a place keeper in synergy with other initiation factors till the next step, and it is very likely that there are more undiscovered intermediate steps where ABCE1 can assure its roles. Our data reinforces the paradigm for the underexplored role of ABCE1 in translation initiation that is consistent with previous reports, but naturally the molecular rationale proposed here must be critically evaluated in future research on eukaryotes.

**Acknowledgments**

We would like to warmly thank Dr Leos S. Valasek for careful reading of the manuscript and enlightening discussion that substantially contributed to the manuscript. This work was supported in part by the ANR grant @RAction program 'ANR CryoEM80S' and the LABEX: ANR-10-LABX-0036_NETRNA and benefits from a funding from the state managed by the French National Research Agency as part of the Investments for the future program (to Y.H.). Netherlands Centre for Electron Nanoscopy, Institute of Biology, Ensteinweg 55, 2333 RC Leiden, The Netherlands for the electron microscope facility. Dr Aline Ribeiro for technical assistance with parasite and the proteomic platform of Strasbourg-Esplanade for conducting the nanoLC-MS/MS analysis. Finally we would like to thank the HPC computer center of the Meinau, Strasbourg for hosting the computer cluster and precious help in computing related-issues.

**Author Contributions**

J.B.Q., E.M-M., A.S. and Y.H. interpreted the data and wrote the manuscript. J.B.Q. purified the complexes. Y.H. performed the cryo-EM experiments and data processing. J.C. performed the MS experiment. J.B.Q. and M.S.S. performed the biological characterization of the *T. cruzi* strain. Y.H. performed the atomic modeling and directed research.

**Reference**

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 btt429.

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.

Andersen, D.S., and Leevers, S.J. (2007). The Essential Drosophila ATP-binding Cassette Domain Protein, Pixie, Binds the 40 S Ribosome in an ATP-dependent Manner and Is Required for Translation Initiation. J. Biol. Chem. 282, 14752–14760.

Anthony, D.D., and Merrick, W.C. (1992). Analysis of 40 S and 80 S complexes with mRNA as measured by sucrose density gradients and primer extension inhibition. J. Biol. Chem. 267, 1554–1562.

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

Bangs, J.D., Crain, P.F., Hashizume, T., McCloskey, J.A., and Boothroyd, J.C. (1992). Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. J. Biol. Chem. 267, 9805–9815.

Barthelme, D., Dinkelaker, S., Albers, S.-V., Londei, P., Ermler, U., and Tampé, 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. U. S. A. 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.

Bindereif, A. (2012). RNA Metabolism in Trypanosomes (Springer Science & Business Media).

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4, 187–217.

Chen, Z., Dong, J., Ishimura, A., Daar, I., Hinnebusch, A.G., and Dean, M. (2006). The essential vertebrate ABCE1 protein interacts with eukaryotic initiation factors. J. Biol. Chem. 281, 7452–7457.

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.

Dong, J., Lai, R., Nielsen, K., Fekete, C.A., Qiu, H., and Hinnebusch, A.G. (2004). The Essential ATP-binding Cassette Protein RLI1 Functions in Translation by Promoting Preinitiation Complex Assembly. J. Biol. Chem. 279, 42157–42168.

Fendley, G.A., Urbatsch, I.L., Sutton, R.B., Zoghbi, M.E., and Altenberg, G.A. (2016). Nucleotide dependence of the dimerization of ATP binding cassette nucleotide binding domains. Biochem. Biophys. Res. Commun. 480, 268–272.

Freire, E.R., Dhalia, R., Moura, D.M.N., da Costa Lima, T.D., Lima, R.P., Reis, C.R.S., Hughes, K., Figueiredo, R.C.B.Q., Standart, N., Carrington, M., et al. (2011). The four trypanosomatid eIF4E homologues fall into two separate groups, with distinct features in primary sequence and biological properties. Mol. Biochem. Parasitol. 176, 25–36.

Freire, E.R., Vashisht, A.A., Malvezzi, A.M., Zuberek, J., Langousis, G., Saada, E.A., Nascimento, J.D.F., Stepinski, J., Darzynkiewicz, E., Hill, K., et al. (2014). eIF4F-like complexes formed by cap-binding homolog TbEIF4E5 with TbEIF4G1 or TbEIF4G2 are implicated in post-transcriptional regulation in Trypanosoma brucei. RNA 20, 1272–1286.

Gabaldón, T., and Huynen, M.A. Prediction of protein function and pathways in the genome era. Cell. Mol. Life Sci. CMLS 61, 930–944.

Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141–147.

Georges, A. des, 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.

Goumans, H., Thomas, A., Verhoeven, A., Voorma, H.O., and Benne, R. (1980). The role of eIF-4C in protein synthesis initiation complex formation. Biochim. Biophys. Acta 608, 39–46.

Gray, N.K., and Hentze, M.W. (1994). Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs. EMBO J. 13, 3882–3891.

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. (2013a). High-resolution cryo-electron microscopy structure of the Trypanosoma brucei ribosome. Nature 494, 385–389.

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

Henshaw, E.C., Guiney, D.G., and Hirsch, C.A. (1973). The ribosome cycle in mammalian protein synthesis. I. The place of monomeric ribosomes and ribosomal subunits in the cycle. J. Biol. Chem. 248, 4367–4376.

Hershey, J.W., and Monro, R.E. (1966). A competitive inhibitor of the GTP reaction in protein synthesis. J. Mol. Biol. 18, 68–76.

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. advance online publication.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. 14, 33–38, 27–28.

Hussain, T., Llácer, J.L., Fernández, 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.

Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M.-A., Adlem, E., Aert, R., et al. (2005). The Genome of the Kinetoplastid Parasite, Leishmania major. Science 309, 436–442.

Jacobson, A. (2005). The end justifies the means. Nat. Struct. Mol. Biol. 12, 474–475.

Karcher, A., Bu¨ttner, K., Märtens, B., Jansen, R.-P., and Hopfner, K.-P. (2005). X-Ray Structure of RLI, an Essential Twin Cassette ABC ATPase Involved in Ribosome Biogenesis and HIV Capsid Assembly. Structure 13, 649–659.

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

Kerr, I.D. (2004). Sequence analysis of twin ATP binding cassette proteins involved in translational control, antibiotic resistance, and ribonuclease L inhibition. Biochem. Biophys. Res. Commun. 315, 166–173.

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.

Kiosze-Becker, K., Ori, A., Gerovac, M., Heuer, A., Nürenberg-Goloub, E., Rashid, U.J., Becker, T., Beckmann, R., Beck, M., and Tampé, R. (2016). Structure of the ribosome post-recycling complex probed by chemical cross-linking and mass spectrometry. Nat. Commun. 7, 13248.

Kispal, G., Sipos, K., Lange, H., Fekete, Z., Bedekovics, T., Janáky, T., Bassler, J., Aguilar Netz, D.J., Balk, J., Rotte, C., et al. (2005). Biogenesis of cytosolic ribosomes requires the essential iron–sulphur protein Rli1p and mitochondria. EMBO J. 24, 589–598.

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. 113, 12174–12179.

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

Lundgren, D.H., Hwang, S.-I., Wu, L., and Han, D.K. (2010). Role of spectral counting in quantitative proteomics. Expert Rev. Proteomics 7, 39–53.

MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J. Phys. Chem. B 102, 3586–3616.

Meleppattu, S., Kamus-Elimeleh, D., Zinoviev, A., Cohen-Mor, S., Orr, I., and Shapira, M. (2015). The eIF3 complex of Leishmania—subunit composition and mode of recruitment to different cap-binding complexes. Nucleic Acids Res. 43, 6222–6235.

Nakamura, Y., and Ito, K. (2003). Making sense of mimic in translation termination. Trends Biochem. Sci. 28, 99–105.

Perry, K.L., Watkins, K.P., and Agabian, N. (1987). Trypanosome mRNAs have unusual “cap 4” structures acquired by addition of a spliced leader. Proc. Natl. Acad. Sci. U. S. A. 84, 8190–8194.

Peterson, D.T., Merrick, W.C., and Safer, B. (1979). Binding and release of radiolabeled eukaryotic initiation factors 2 and 3 during 80 S initiation complex formation. J. Biol. Chem. 254, 2509–2516.

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.

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.

Rezende, A.M., Assis, L.A., Nunes, E.C., da Costa Lima, T.D., Marchini, F.K., Freire, E.R., Reis, C.R., and de Melo Neto, O.P. (2014). The translation initiation complex eIF3 in trypanosomatids and other pathogenic excavates – identification of conserved and divergent features based on orthologue analysis. BMC Genomics 15, 1175.

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.

Roy, F.L., Salehzada, T., Bisbal, C., Dougherty, J.P., and Peltz, S.W. (2005). A newly discovered function for RNase L in regulating translation termination. Nat. Struct. Mol. Biol. 12, 505–512.

Salas-Marco, J., and Bedwell, D.M. (2004). GTP hydrolysis by eRF3 facilitates stop codon decoding during eukaryotic translation termination. Mol. Cell. Biol. 24, 7769–7778.

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.

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.

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.

Strunk, B.S., Loucks, C.R., Su, M., Vashisth, H., Cheng, S., Schilling, J., Brooks, C.L., Karbstein, K., and Skiniotis, G. (2011). Ribosome Assembly Factors Prevent Premature Translation Initiation by 40S Assembly Intermediates. Science 333, 1449–1453.

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. Struct. Lond. Engl. 1993 16, 673–683.

Yarunin, A., Panse, V.G., Petfalski, E., Dez, C., Tollervey, D., and Hurt, E.C. (2005). Functional link between ribosome formation and biogenesis of iron–sulfur proteins. EMBO J. 24, 580–588.

Yoffe, Y., ZUBEREK, J., LEWDOROWICZ, M., ZEIRA, Z., KEASAR, C., ORR-DAHAN, I., JANKOWSKA-ANYSZKA, M., STEPINSKI, J., DARZYNKIEWICZ, E., and SHAPIRA, M. (2004). Cap-binding activity of an eIF4E homolog from Leishmania. RNA 10, 1764–1775.

Yoffe, Y., Léger, M., Zinoviev, A., Zuberek, J., Darzynkiewicz, E., Wagner, G., and Shapira, M. (2009). Evolutionary changes in the Leishmania eIF4F complex involve variations in the eIF4E–eIF4G interactions. Nucleic Acids Res. 37, 3243–3253.

Zamudio, J.R., Mittra, B., Campbell, D.A., and Sturm, N.R. (2009). Hypermethylated cap 4 maximizes Trypanosoma brucei translation. Mol. Microbiol. 72, 1100–1110.

Zeiner, G.M., Sturm, N.R., and Campbell, D.A. (2003). The Leishmania tarentolae Spliced Leader Contains Determinants for Association with Polysomes. J. Biol. Chem. 278, 38269–38275.

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.

Zhao, Z., Fang, L.L., Johnsen, R., and Baillie, D.L. (2004). ATP-binding cassette protein E is involved in gene transcription and translation in Caenorhabditis elegans. Biochem. Biophys. Res. Commun. 323, 104–111.

Zinoviev, A., and Shapira, M. (2012). Evolutionary Conservation and Diversification of the Translation Initiation Apparatus in Trypanosomatids. Int. J. Genomics 2012, e813718.

Zinoviev, A., Léger, M., Wagner, G., and Shapira, M. (2011). A novel 4E-interacting protein in Leishmania is involved in stage-specific translation pathways. Nucleic Acids Res. 39, 8404–8415.

## Material and methods

### Trypanosoma culture and ribosome isolation

Epimastigotes form of *Trypanosoma cruzi* Y strain were grown at 28ºC in liver infusion tryptose (LIT) medium, supplemented with 10% heat-inactivated fetal bovine serum. The cells were grown to a maximum density of 9.5 x $10^6$ cells/mL in a 400 mL of total medium volume, and they were collected by centrifugation at 1,000 x g for 15 min. The parasite pellets were washed three times with a buffer containing 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 (Promega). The parasite suspension was freeze-thawed three times and the lysate was centrifuged twice for 30 min and 5 min at 12,000g to remove the cells debris present in the suspension.

The 40S and 80S ribosomes were purified from a clean *T. cruzi* lysate suspension. Briefly, 400 uL of clean lysates at an $OD_{260}$ of between 30-40 units were loaded on the top of a sucrose gradient (10%-30%) prepared in the same buffer mentioned above, and then centrifuged at 210,000 x g for 5h30. The gradient was fractionated from the bottom to the top, and the 40S and 80S related fractions were collected according the UV absorbance profile. After the fractionation, 40S and 80S ribosomes were pelleted by ultracentrifugation at 624,000 x g for 2h and resuspended in a buffer containing 10 mM HEPES-KOH pH 7.4, 50 mM KOAc, 10 mM $NH_4Cl$, 5 mM $Mg(OAc)_2$, 2 mM DTT and EDTA free protease inhibitor cocktail. The quality of the sample was analyzed by SDS-PAGE, agarose gel (after phenol chloroform rRNA extraction) and mass spectrometry.

### *Ex vivo* isolation of the stalled ABCE1-containing 40S complex for Cryo-EM

The *T. cruzi* lysate was obtained according to the protocol mentioned above, and incubated with 20 mM GMP-PNP (Sigma-Aldrich) for 10 min on ice and then at 28ºC for 10 min, to stall ABCE1 on ribosomal small subunit as described before (Andersen and Leevers, 2007; Barthelme et al., 2011). The stalled 40S was purified on sucrose gradient by using the same protocol mentioned above, and resuspended in a buffer containing 10 mM HEPES-KOH pH 7.4, 50 mM KOAc, 10 mM $NH_4Cl$, 5 mM Mg $(OAc)_2$, 2 mM DTT. Native 40S sapmples where prepared similarly with the exception of adding GMP-PNP. Samples were then deposited on grids for their analysis by cryo-EM. In order to have high pure native 40S and stalled 40S, the samples were additionally purified by gel filtration, using Superose 6/300 column (GE healthcare) coupled to a HPLC system (Agilent systems). The pure samples were analyzed by SDS-PAGE (15% polyacrylamide gel) followed by mass spectrometry.

**NanoLC-MS/MS analysis**

The 40S, 80S and stalled 40S complexes were analyzed by liquid mass spectrometry (MS) as previously described[42]. Briefly, liquid samples were precipitated with 0.1 M ammonium acetate in 100% methanol. After a reduction-alkylation step (Dithiothreitol 5 mM - Iodoacetamide 10 mM), proteins were digested overnight with 1/25 (W/W) of sequencing-grade porcine trypsin (Promega). The mixtures of peptide were resuspended in water containing 0.1% FA and injected on nanoLC-MS/MS (NanoLC-2DPlus system with nanoFlex ChiP module; Eksigent, ABSciex, Concord, Ontario, Canada, coupled to a TripleTOF 5600 mass spectrometer). 5%-40% gradients of acetonitrile were used to elute the peptides from the C-18 analytical column (75 $\mu$m ID x 15 cm ChromXP; Eksigent). Peptides were identified with Mascot algorithm (version 2.2, Matrix Science, London, UK) through the ProteinScape 3.1 package (Bruker). They were validated with a minimum score of 30, a p-value<0.05 and proteins were validated respecting a false discovery rate FDR<1%. The data were searched against the T. cruzi proteome set from the UniProt database (release from July 2015, with decoy sequences) and complemented by search on TriTrypDB (http://tritrypdb.org/tritrypdb/). For the MS/MS analysis of the polyacrylamide gel, the SDS-PAGE bands of stalled 40S sample were excised from the gel, digested with trypsin, and subjected to nanoLC-MS/MS, using the same procedure described above.

**Cryo-EM data acquisition**

4 µL of 120 nM 40S or 200 nM stalled 40S, were applied to 400 mesh holey carbon Quantifoil 2/2 grids (Quantifoil Micro Tools), as described in our previous work (Simonetti et al., 2016). The grids were blotted with filter paper from both sides for 1.5 s at 4ºC, 100% humidity, blot force 5 and waiting time 30 s (Vitrobot apparatus Mark IV). The data were collected on a spherical aberration (Cs) corrected Titan Krios S-FEG instrument (FEI) operating at 300 kV acceleration voltage and at a nominal underfocus of $\Delta z = -0.6$ to $-4.5$ $\mu$m 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 microscope was carefully aligned as well as the Cs corrector. The Falcon II camera was calibrated at nominal magnification of 59,000 x. The calibrated magnification on the 14 $\mu$m pixel camera was 127,272 x, resulting in 1.1 Å pixel size at the specimen level. The camera was set up to collect 7 frames (starting from the second one) out of 22 possible. Total exposure was 1.5 s, with a dose of 57 /Å$^2$ (or 2.2 /Å$^2$ per frame).

**Image processing**

We used the SCIPION (Abrishami et al., 2015; de la Rosa-Trevín et al., 2016; Sorzano et al., 2015) package for image processing and 3D reconstruction, which integrates several data processing software. The movie alignment was performed with the Optical Flow algorithm integrated in Xmipp3 (de la Rosa-Trevín et al., 2013). Then, an average image of the whole stack was used to determine the contrast transfer function by CTFFIND4 (Rohou and Grigorieff, 2015) and to select (semi-automatically) ~197,000 particles for the stalled 40S and ~188,500 particles for the native 40S in SCIPION (Abrishami et al., 2013). RELION (Scheres, 2012) was used for particle sorting through 3D classification via SCIPION, leading to two major 3D classes comprising ~73,000 particles for the stalled 40S and ~53,000 particles for the native 40S (Supplementary Fig. S2) showing structural features attributable to the Trypanosome 40S and stalled 40S, immediately obvious thanks to the presence of large RNA expansion segments 6 (Fig. 2). Other minor 3D classes were derived: 40S dimers, stalled 80S and a low-resolution 40S/ABCE1 class that does not display the additional platform density, which were found in the stalled sample (Supplementary Fig. S2B), and 40S dimers, empty 80S, and a low-resolution native 40S (Supplementary Fig. S2E). We have attempted additional classification that did not yield any additional classes. The major 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 5.9 Å for the stalled 40S complex and 6.75 Å for the native one.

**Segmentation and display of density maps.** Cryo-EM reconstructions were segmented in UCSF Chimera (Pettersen et al., 2004) using the SEGGER module (Pintilie et al., 2010). Segments counting fewer than 1,000 voxels were discarded.

**Atomic model**

Atomic model of the small ribosomal subunit from *Trypanosoma brucei* (Hashem et al., 2013a) was rigid-body fitted into our cryo-EM map for its analysis and figures rendering only. A near-complete atomic model of ABCE1 from *T. cruzi* was derived by homology to the crystal structures of archaeal ABCE1 from *Pyrococcus abyssi*(Karcher et al., 2005) using SWISS-PDB (Arnold et al., 2006). After the construction of the atomic model, it was placed into its cryo-EM density map segment of ABCE1 by rigid- body fitting using Chimera. Starting from this

system including only *T. cruzi* ABCE1, the atomic model was flexibly fitted into its map segment using MDFF (molecular dynamics flexible fitting) (Trabuco et al., 2008) as described in previous work (Georges et al., 2015). Briefly, the initial system was prepared for MDFF using VMD (Humphrey et al., 1996) by embedding ABCE1 in a solvent box of TIP3P water molecules, with an extra 12 Å padding in each direction, and the model's charge was neutralized by using KCl. The trajectories were run in NAMD using CHARMM force field parameters (Combined CHARMM All-Hydrogen Topology File for CHARMM22 Proteins and CHARMM27 Lipids) (Brooks et al., 1983; MacKerell et al., 1998). The runs were stopped at 350 ps of simulation time, after convergence.

**Local resolution measurement of the cryo-EM map.** The local resolution of the cryo-EM reconstruction was measured using RESMAP (Kucukelbir et al., 2014) and represented as a variable color scale using UCSF Chimera (Pettersen et al., 2004).

# Supplementary

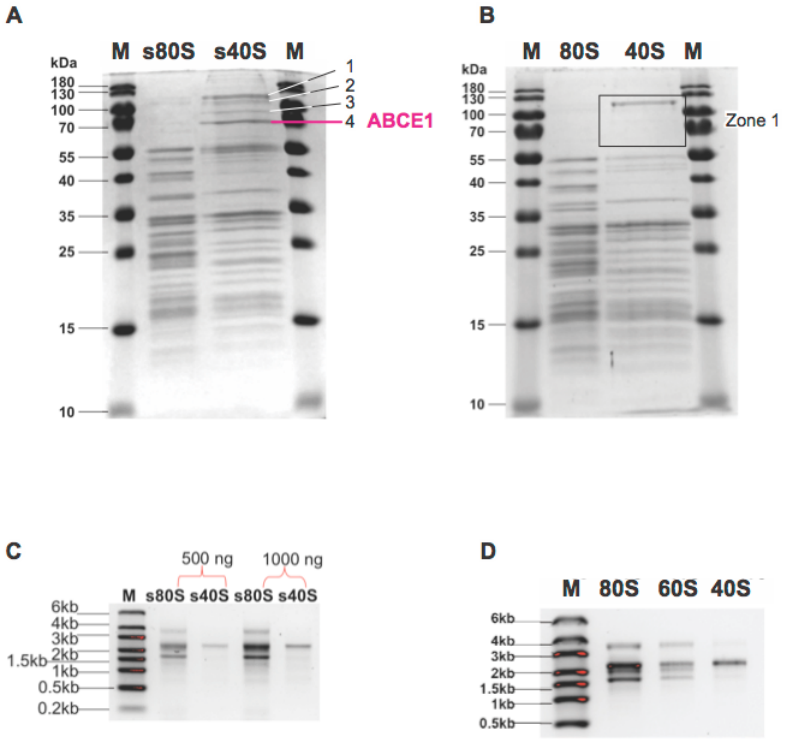


**Figure S 1: nanoLC–MS/MS characterization of native and stalled 40S complexes, related to Figure 1.** (A-B) After pelleting sucrose gradient peak fractions and subsequent gel filtration chromatography (Fig. 1), purified 40S samples (stalled with GMP-PNP and native) were analyzed by SDS-PAGE followed MS/MS analysis on the sliced bands indicated by numbers 1 to 4 for the stalled 40S (s40S) and by zone 1 for the native 40S complexes. Stalled and native 80S samples were used as controls for ribosomal proteins. (A) Five µg of either s80S or s40S were loaded in 15% SDS-PAGE gels. Bands 1 to 4 from s40S sample were individually excised from the gel and characterized by nanoLC- MS/MS. The identities of each band are reported in Supplementary Table 1. ABCE1 was by far the strongest hit in band 4 and it is highlighted in pink. (B) Five µg of either 80S or 40S were loaded in 15% SDS-PAGE gels. The highlighted zone 1 was excised from the gel and characterized by nanoLC- MS/MS. The identities of proteins found in this zone are reported in Supplementary Table 1. (C-D) After rRNA phenol-chloroform extraction, rRNA quality control on stalled or native ribosomal samples was performed using non-denaturing agarose gels. (C) 500 or 1000 ng of rRNA from s80S or s40S samples were loaded in 1% agarose gels. (D) 1000 ng of rRNA from either native 80S, 60S or 40S were loaded in 1% agarose gels. Related to figure 1.

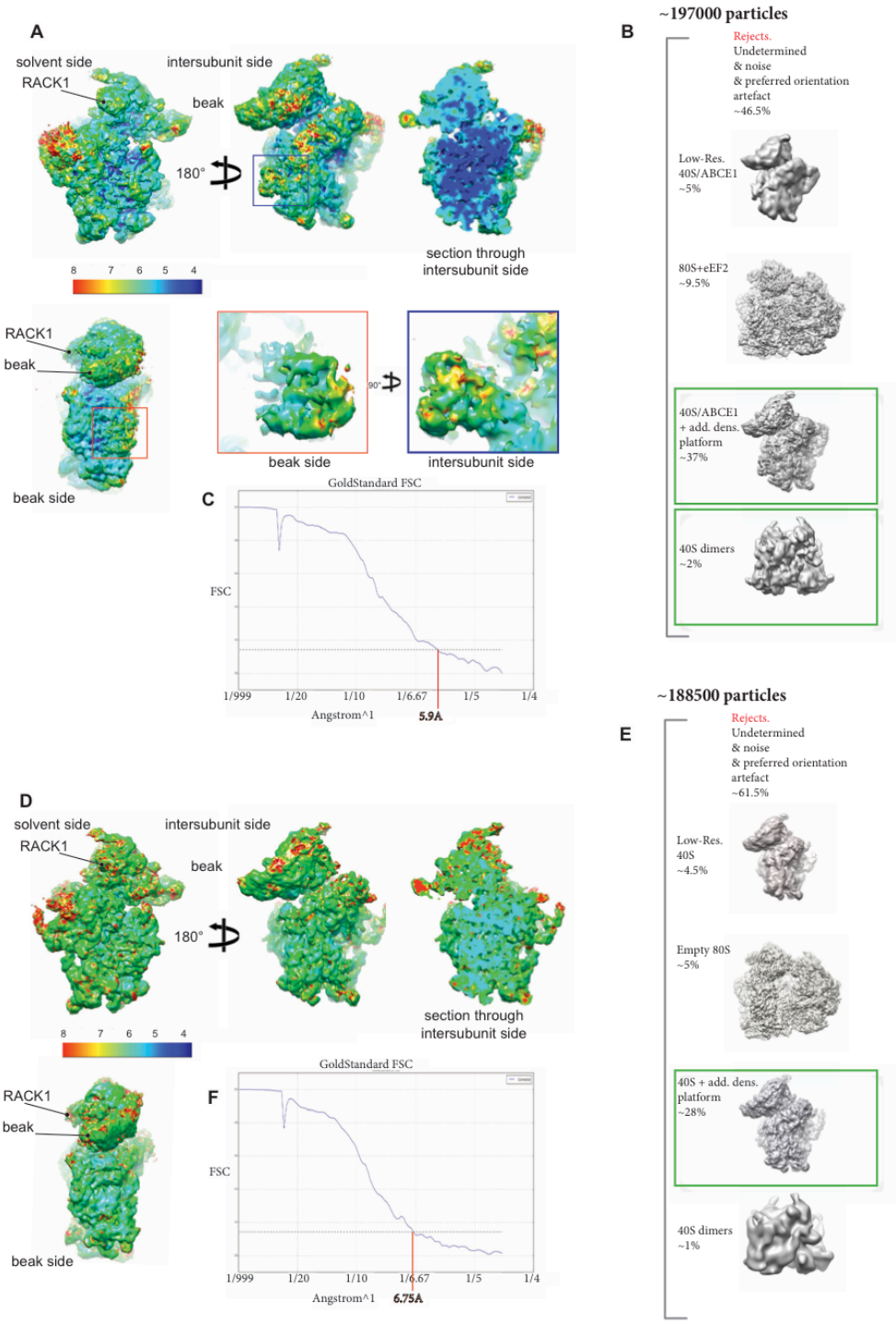


**Figure S 2: Particle sorting of the cryo-EM images and the local and average resolutions of native and stalled 40S.** (A) Local resolution of the cryo-EM map from class 40S/ABCE1/ηF, displayed from different orientations, colored according to the local resolution as estimated by RESMAP. Close- up views (red and blue squares) highlight local resolution of the ABCE1 density located at the GTPase binding site. According to RESMAP the local resolution of ABCE1 is roughly between 5.5 and 7 Å. (B) Overview of the particle

classes obtained after particle sorting. The populations of all classes are indicated by their particle numbers and percentages over the total number of particles. The ABCE1/40S/ηF complex was processed from a total of ~197000 particle images (see Methods) using the eukaryotic small ribosomal subunit as a reference. Green rectangle indicates the retained classes for further refinement in RELION. (C) Gold-standard FSC curve for the three-dimensional reconstructions from class 40S/ABCE1, marking the corresponding resolution in angstrom, 5.9. (D) Local resolution of the cryo-EM map from class native 40S, displayed from different orientations, colored according to the local resolution as estimated by RESMAP. (E) Overview of the particle classes obtained after particle sorting. The populations of all classes are indicated by their particle numbers and percentages over the total number of particles. The native 40S complex was processed from a total of ~1885000 particle images (see Methods) using the same method mentioned above. Green rectangle indicates the retained class for further refinement in RELION. (F) Gold-standard FSC curve for the three-dimensional reconstructions from class native 40S, marking the corresponding resolution in angstrom, 6.75 Å. Related to figure 1.

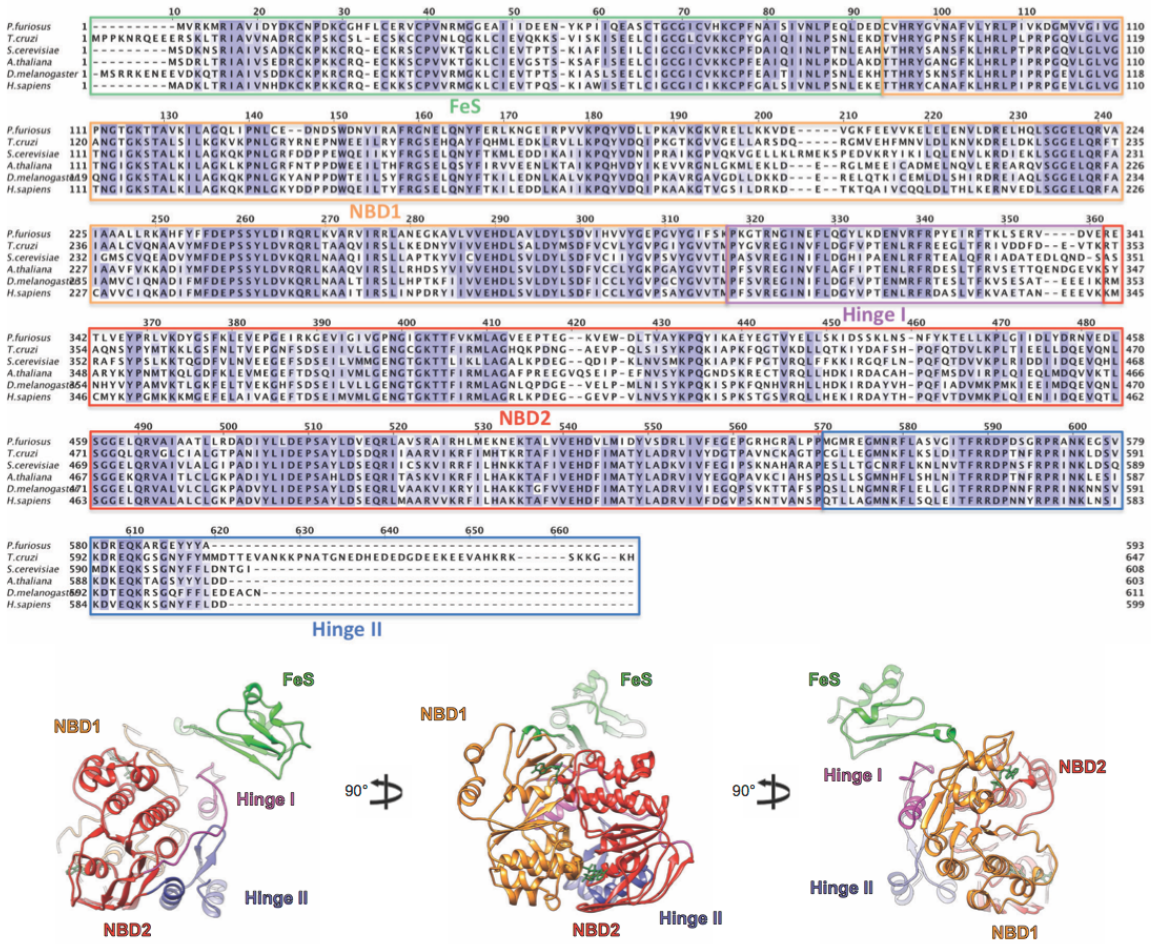


**Figure S 3: Sequence alignment of ABCE1 from different species highlighting the conservation of ABCE1 domains.** A multiple sequence alignment colored by BLOSUM 62 score of ABCE1 orthologues highlighting high conserved regions of T. cruzi ABCE1 when compared to its archaeal, plant, yeast, drosophila and human counterparts. The darkest color represents the highest conserved residues. Different domains of ABCE1 are framed in variably colored rectangles according to the color code used on the atomic model shown on the bottom. Related to figures 2 and 3.

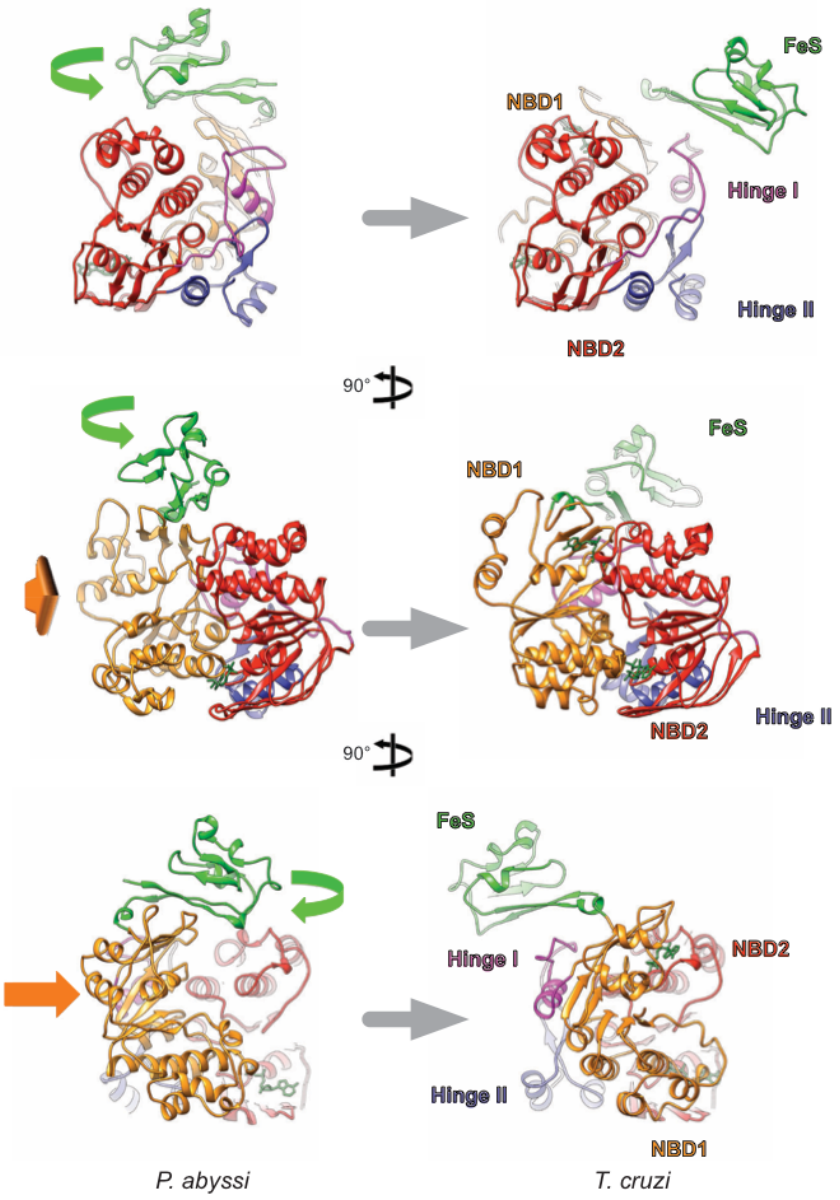


**Figure S 4: Conformational changes undertaken by the T. cruzi ABCE1 by comparison to X-ray structure of archaeal ABCE1.** Comparison of the ADP-bound ABCE1 crystal structure from Pyrococcus abyssi24 (shown on the left) with the GMP-PNP-bound ABCE1 in complex with 40S found in the T. cruzi (shown on the right) viewed from three different orientations. The FeS domain from T. cruzi ABCE1 rotates by ~180° when compared to the archaea crystal structure. The green arrow shows the direction of the FeS domain rotation. The orange arrows indicate the direction of NDB2 domain movement. Each domain is colored variably. Related to figure 2.

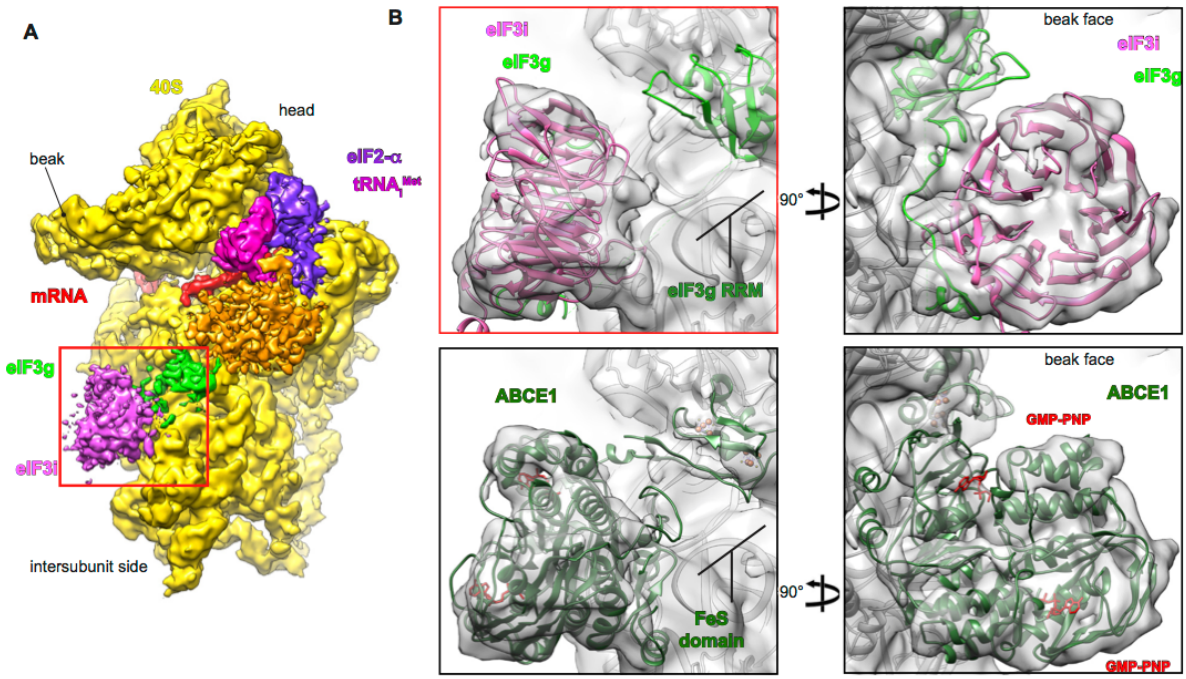


**Figure S 5: Comparison between models of ABCE1 and eIF3i and eIF3g fitted in Cryo-EM segmented map of mammalian 48S late-stage initiation complex (Simonetti et al., 2016).** (A) Cryo-EM structure of the m48S late-stage IC viewed from intersubunit side, highlighting the densities segments that previously was attributed to eIF3i (pink), and eIF3g (green). (B) Atomic model of the m48S late-stage IC fitted in its cryo-EM segmented map (Simonetti et al., 2016), focused on the eIF3 peripheral subunits i (pink) and g (green), seen from the intersubunit and beak sides. Bottom panel, same as B, but with ABCE1 fitted in the intersubunit density instead of eIF3i+g, showing the better fit of ABCE1 for that latter density. Related to figure 7.

## Article 2: The cryo-EM structure of a novel 40S kinetoplastid-specific ribosomal protein

### In Brief

Querido et al., presented a cryo-EM structure of a novel kinetoplastid-specific ribosomal (r-) protein (KSRP) bound to the ribosome. Recent advance in cryo-EM field have allowed the description of nearly all the rRNA chains, including most of the kinetoplastid-specific ESs and elements and r-protein extension. However, r-protein assignments were performed mainly according to homology to known r-proteins from other eukaryotes such as *Saccharomyces cerevisiae*, therefore the discovery of novel kinetoplastid-specific r-proteins has been out of reach. Here we demonstrate the existence of a novel r-protein in kinetoplastids, which binds at the left foot of 40S. The general features of KSRP highlight its potential as a new target for the development of safer and more specific anti-kinetoplastid therapeutic agents.

### Highlights

- Cryo-EM structures of ribosomal complexes unveiled the first bona fide kinetoplastid-specific ribosomal protein reported so far.
- NRBD/KSRP is an essential 40S ribosomal protein that stabilizes kinetoplastid-specific rRNA elements.
- It also interacts with the kinetoplastid-specific C-terminal region of ribosomal protein eS6.
- The entire interacting network of NRBD/KSRP is kinetoplastid-specific.

**Article 3: Cryo-EM structure of native translation initiation complex from human pathogen *Trypanosoma cruzi***

## In Brief

Querido et al., presentes here the cryo-EM structure of 48S preinitiation complex from human pathogen *Trypanosoma cruzi*. Kinetoplastids possess some unique features such as larger rRNA expansion segments, protein extension and cap-4 structure in all mature mRNA. These specificities suggest that translation initiation in kinetoplastids may be regulated differently compared to other know eukaryotes, and present kinetoplastid-specific aspects. However, little is known about translation regulation in these organisms. The structure of the near complete 48S initiation complex from *T. cruzi* reveals the binding pattern of most eIFs and highlights the important role of their larger expansion segments for the binding and stabilization of eIF3 core, as well as the binding of eIF3d. Furthermore, structural and proteomic analysis reveals an unsuspected role of DDX60 in translation initiation in kinetoplastids. The kinetoplastid-specific aspects of translation initiation revealed here highlight the potential of translation for the development of new therapeutic agents against kinetoplastids

## Highlights

- Cryo-EM structure together with proteomic analysis indicated DDX60 as an important player during translation initiation in kinetoplastids.
- The biding patter of kinetoplastidian eIF3 core to the 40S is different from that observed in their mammalian host.
- eIF3c is sandwiched by ES7S-hA and ES7S-hB.
- $ES9^S$ binds and stabilizes eIF3d.

**Cryo-EM structure of native translation initiation complex from human pathogen *Trypanosoma cruzi***

Jailson Brito Querido[1], Anthony Bochler[1], Lauriane Kuhn[2], Angelita Simonetti[1], and Yaser Hashem[1,*]

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

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

*Correspondence should be sent to: y.hashem@ibmc-cnrs.unistra.fr (Y.H.)

**SUMMARY**

Kinetoplastids are responsible for several human vector-borne diseases, that together threaten more than 400 million people world-wide, include that of Chagas disease caused by *Trypanosoma cruzi*. They possess unusual features, such as the presence of a spliced-leader in all mature mRNAs. Moreover, ribosomal RNA expansion segments (ESs), such as $ES3^S$, $ES6^S$, $ES7^S$ and $ES9^S$ are larger than that in their mammalian host counterparts. The location of their larger $ES6^S$, $ES7^S$ and $ES9^S$ close to the mRNA exit channel suggest an involvement in the translation initiation process. Here, we purified from *T. cruzi* cell lysates native initiation complexes that we then analyzed by cryo-EM. The structure of the native initiation complexes reveals several key kinetoplastid-specific aspects of translation initiation, such as the peculiar accomodation of eIF3 core in the 48S, through the interaction of eIF3a, eIF3c and eIF3e with $ES6^S$ and $ES7^S$. Indeed, our structure reveals that eIF3c is sandwiched by $ES7^S$-helix A ($ES7^S$-hA) and $ES7^S$-hB. Furthermore, $ES9^S$ binds and stabilizes eIF3d. Notably, our structural and proteomic analysis reveals that kinetoplastidian DDX60 may play an important role in translation initiation by unwinding the spliced-leader mRNAs. The peculiarities revealed here in this first structure of initiation complex from protozoan parasite, provides new perspectives for the development of translation-targeted therapeutic agents against kinetoplastids.

**Keywords**

Kinetoplastids, translation, ribosome, initiation complex

## INTRODUCTION

Kinetoplastid is a group of flagellated protozoans, pathogens of plants, invertebrate, and vertebrate, including human pathogens such as *Trypanosoma cruzi* (etiologic agent of Chagas disease), *Trypanosoma brucei spp*. (etiologic agent of sleeping sickness) and *Leishmania spp*. (etiologic agent of leishmaniosis). Together they threaten more than 400 million people world-wide (World Health Organization, 2012), however, preventive vaccines are lacking and the treatment is based on highly toxic drugs (Field et al., 2017; Clayton, 2010). As early-diverged from other eukaryotes, kinetoplastids share some unique features, such as unusual large rRNA expansion segments (ESs) on the small ribosomal subunit, ribosomal protein (r-protein) extension (Hashem et al., 2013a; Shalev-Benami et al., 2016; Zhang et al., 2016; Liu et al., 2016) and a hypermethylated cap structure (cap-4) (Perry et al., 1987). Furthermore, the kinetoplastidian polycistronic pre-mRNA maturation occurs by trans-splicing of a Splice leader (high conserved 39 nt mine exon) at the 5' end. The spliced-leader confers a hypermethylated cap structure (cap-4) to all mature mRNA (Perry et al., 1987). The presence of an unusual 5' cap-4 structure, suggests a different mechanism for the 43S preinitiation complex (43S PIC) recruitment to the mRNA. Indeed, although orthologues of most of the canonical initiation factors (eIFs) are encoded in kinetoplastidian genomes, it has been shown that they present low sequence conservation when compared to those from their mammalian host counterparts (Rezende et al., 2014; Meleppattu et al., 2015; Li et al., 2017).

Cryo-EM structure of kinetoplastids ribosome (Hashem et al., 2013a; Zhang et al., 2016) revealed that their unusual large expansion segments $ES6^S$, $ES7^S$ and $ES9^S$ localize close to the mRNA exit channel (Hashem et al., 2013a; Hashem et al., 2013b; Zhang et al., 2016). Due to that location and their size, it has been proposed that they may play a role in translation initiation by interacting with eIF3 (Hashem et al., 2013b). Nevertheless, little is known about the translation initiation in these early eukaryotes.

Translation initiation in eukaryotes is a dynamic, cyclical and complex process, regulated for more than 12 eukaryotic eIFs and some auxiliary factors (reviewed in Hinnebusch, 2017). This process begins with the assembly of the eIF2-ternary complex composed by eIF2, GTP and Met-$tRNA_i^{Met}$ (eIF2-TC). The recruitment of eIF2-TC to the 40S is enhanced by other eIFs, such as eIF1, eIF1A, eIF3 and eIF5. Together they lead to the formation of 43S PIC. Recent data have indicated that also ATP-binding cassette subfamily E member 1 (ABCE1; Rli1 in yeast) plays role during 43S PIC assembly (Dong et al., 2004; Heuer et al., 2017; Mancera-Martínez et al., 2017).

Once assembled, the 43S PIC is recruited to the eIF4F-bound mRNA complex. eIF4F complex is composed by the cap-binding protein (eIF4E), a scaffold protein (eIF4G) and a DEAD-box RNA helicase (eIF4A). The helicase activity of eIF4A is enhanced by the activity of eIF4B, and the later binds to the single-stranded mRNA to prevents the re-annealing (reviewed in Marintchev et al., 2009). The multifactor complex eIF4F/4B is able to unwind the 5′ UTR mRNA, and promotes its attachment to the 43S likely via eIF4G-eIF3 interaction. Nevertheless, some RNA requires the activity of other DEAD or DEAH helicases in addition to eIF4A. These two groups of helicases form the largest family of RNA helicases and they are characterized by the presence of an Asp-Glu-Ala-Asp (DEAD) or Asp-Glu-Ala-His (DEAH) motifs. They are implicated in a number of cellular processes involving RNA binding and unwinding of the secondary structure. Remarkably, it has been shown that several DEAD and DEAH box helicases such as DHX29, Ded1/DDX3, Dhh1/DDX6, Vasa/DDX4 and RHA/DHX9 can play roles during translation initiation (reviewed in Parsyan et al., 2011). The mechanisms by which these helicases influence translation are still unclear. Nevertheless, the role of DHX29 on the translation of high structure 5′-UTR mRNA (Pisareva et al., 2008), as well as its binding site on the ribosome have been extensively described. DHX29 binds at the vicinity of the mRNA entry channel and interacts with eIF3i and eIF3b located at the solvent side of 43S PIC (Hashem et al., 2013b; des Georges et al., 2015; Pisareva and Pisarev, 2016).

The role of several DEAD box helicases such as DDX60 are still unclear. It has been shown that human DDX60 is a sentinel helicase for cytoplasmic viral RNA sensor RIG-I (retinoic acid-inducible gene-I) (Miyashita et al., 2011; Oshiumi et al., 2015), however the same activity through the RIG-I signaling has also been attributed to DHX29 in human airway epithelial cells (Sugimoto et al., 2014). This cross-correlation, suggest that a RNA helicase can play different roles in processes that involve RNA, either by promoting the mRNA translation or by acting as a sentinel helicase for exogenous RNA.

As mention above, some mRNAs require other helicases in addition to eIF4A for an efficient translation initiation, however the identification of auxiliary helicase that can assist translation is still out of reach in some human pathogens, such as kinetoplastids. Indeed, we expect that translation in these organisms present several unusual features. Accordingly, we have reported the structure of a novel kinetoplastid-specific ribosomal protein (KSRP) bound at the left foot of 40S (Querido et al., 2017). KSRP interacts with the $ES3^S$, $ES6^S$ and also with the kinetoplastid-specific C-terminal extension of the r-protein eS6. These findings highlight the uniqueness of the translational machinery in kinetoplastids and suggest the functional importance of their large ESs during the translation.

Here, we purified from *T. cruzi* cell lysates native translation initiation complex, that we then analyzed by cryo-EM. The structure of native initiation complexes reveals several kinetoplastid-specific aspects of translation initiation, such as an intricate interaction between eIF3 and $ES6^S$ and $ES7^S$. Furthermore, our data revels an unsuspected role of a DDX60 during translation initiation in kinetoplastids.

## RESULTS AND DISCUSSION

### Proteomic analysis of native kinetoplastids initiation complex revels the presence of DDX60

To purify native initiation complexes from *T. cruzi*, we treated the whole cell lysate (See Methods) with a non-hydrolysable analog of GTP (GMP-PNP). The use of GMP-PNP is known to synchronize translation initiation at late stage, after start-codon recognition and prior Pi release (Simonetti et al., 2016; Merrick, 1979). Furthermore, the use of GMP-PNP also prevents the nucleotide hydrolyses and the consequent recycling of several auxiliary factors, including helicases (Mallam et al., 2012; Chen et al., 2008; Liu et al., 2008). The native initiation complex was purified by sucrose gradient, as described before (Simonetti et al., 2016). As we have expected, the use of GMP-PNP lead to an increase of initiation complexes (Figure 1A). In order to characterize the purified initiation complexes, we also run a second gradient without treating the lysate with GMP-PNP (Figure 1A). Proteomic analysis (MS/MS) have indicated that the initiation complex is enriched with eIFs, including eIF1, eIF1A, eIF2, eIF3 and eIF5 (Figures 1B, S1).

Consistent with our recent report in mammals (Simonetti et al., 2016; Mancera-Martínez et al., 2017), we found that the use of GMP-PNP to treat the cell extracts lead to the arrest of ABCE1-containing translation initiation complex. Indeed, it is also in agreement with a recent publication that also suggested that Rli1 (ortholog of ABCE1 in yeast) can bind to 40S even after the subunit splitting and plays a role in translation initiation (Heuer et al., 2017).

Surprisingly, proteomic analysis also has indicated that the initiation complex is enriched with DDX60 (Figures 1B, S1). The binding of DDX60 to the initiation complex was validated by size-exclusion chromatography followed by liquid MS/MS analyses (Figure S1). There is no information about the role of DDX60 in translation initiation, therefore in order to understand the specie-specificity of the role of DDX60 in translation initiation, we also purified native initiation complex from mammals (rabbit reticulocyte lysate), yeast *Saccharomyces cerevisiae* and *Leishmania tarentolae*. Proteomic analysis revels that the presence of DDX60

is exclusive to the two kinetoplastids we have analyzed, *T. cruzi* and *L. tarentolae* (Experimental Procedure; Figure S2). This finding suggests that the role of DDX60 in translation initiation is a kinetoplastid-specific aspect of translation initiation.

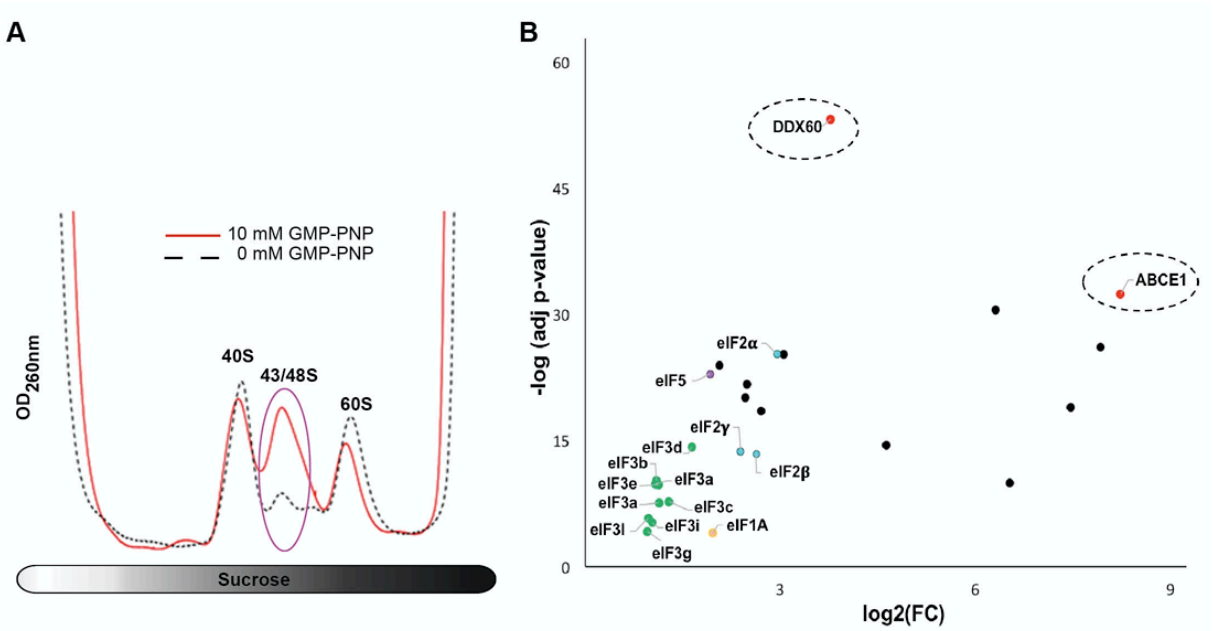


**Figure 1: Purification of initiation complex and proteomic analysis. A,** Sucrose-gradient profile of *T. cruzi* cell lysate treated (10 mM) and untreated with GMP-PNP. Fractions containing 48S particles (magenta circle) were pooled and used for proteomic analysis and cryo-EM reconstruction. **B,** Volcano plot of the 48S fractions (10 mM GMP-PNP) compared against the native 40S fractions (0 mM GMP-PNP). The plot is the log2 ratios plotted against the adjusted negative log p values (adjusted p value < 0.05 and log2 fold change >1). Black dots are proteins that we could not validated their co-purification with 48S by size exclusion chromatography. See also supplementary figure S1.

## The overall structure of the native initiation complex from kinetoplastids

The cryo-EM reconstruction was derived after particle sorting and 3D classification (Experimental Procedures). The structure of native initiation complexes purified from *T. cruzi* cell lysate was determined to overall resolutions of 6.0 Å (Figure S3).

The cryo-EM reconstruction of the main eIF3 core-contain class reveals an open conformation, scanning-competent stage, in which the head of 40S is tilted upwards to open the mRNA entry channel latch, as described before in yeast (Llácer et al., 2015). Therefore, following the terminology from yeast complexes, we hereafter referred to our complex as 48S pre-initiation complex (48S-open).

We segmented the cryo-EM map based on the structure of mammalian 43S pre-initiation complex (des Georges et al., 2015) and yeast 48S-open (Llácer et al., 2015). Our segmentation displayed densities corresponding to 40S, eIF1A, eIF1, eIF2α, eIF2β, eIF2γ, Met-tRNA$_i^{Met}$, eIF3d and eIF3 core (Figures 2A-D). Furthermore, we also observed the helix-rich segment of eIF3c-NTD, formed by a cluster of five helices in a pocket formed by h11, h24, h27, h44, and uS15 (Figure 3A), as described before in yeast 48S-open (Llácer et al., 2015).

The structure of eIF2α either in context of ternary complex or in context of translation initiation complexes has been characterized (Hussain et al., 2014; des Georges et al., 2015; Schmitt et al., 2012). It contains three domains: a N-terminal β-barrell domain (D1), a helical domain (D2) and a C-terminal α–β domain (D3). The structure of kinetoplastidian eIF2α is not known yet. However, the sequence analysis indicated that kinetoplastidian eIF2α contains additional 120 amino acids at the N-ter region when compared to that from other eukaryotes (Figure S4A). Therefore, as anticipated, there is an additional density connected to the domain 1 and 2 of eIF2α (Figures 2 and S4B), which is also interacting with the T-loop of the Met-tRNA$_i^{Met}$. This density has not been described before in any eukaryotes (Figure S4B). Furthermore, its general shape is compatible with the helix-rich secondary structure predicted by SYMPRED (Lin et al., 2005). Therefore, we uniquely attributed this density to the additional N-terminal elements of kinetoplastidian eIF2α.

It is not clear what is the main role of the kinetoplastidian additional domain of eIF2α. Nevertheless, considering that it also interacts with the head of 40S subunit, it can stabilize the binding of the ternary complex during the scanning process.

The structure of eIF2β has also been described in several studies, such as in archaea and yeast (Schmitt et al., 2012; Llácer et al., 2015). However, due to its flexibility, the structure of the unstructured N-terminal region has not been described so far. Interestingly, here we have an unidentified density, attached to the N-terminal region of eIF2β, which is also interacting with eIF2γ. Due to the limitation imposed by the resolution, it is not possible to unambiguously assign this density (Figures 2A-B). However, the size and the location suggest that it may belong to the N-terminal domain of eIF2β that could not been solved when the discovered of the structure of eIF2β in the context of translation initiation complex (Llácer et al., 2015).

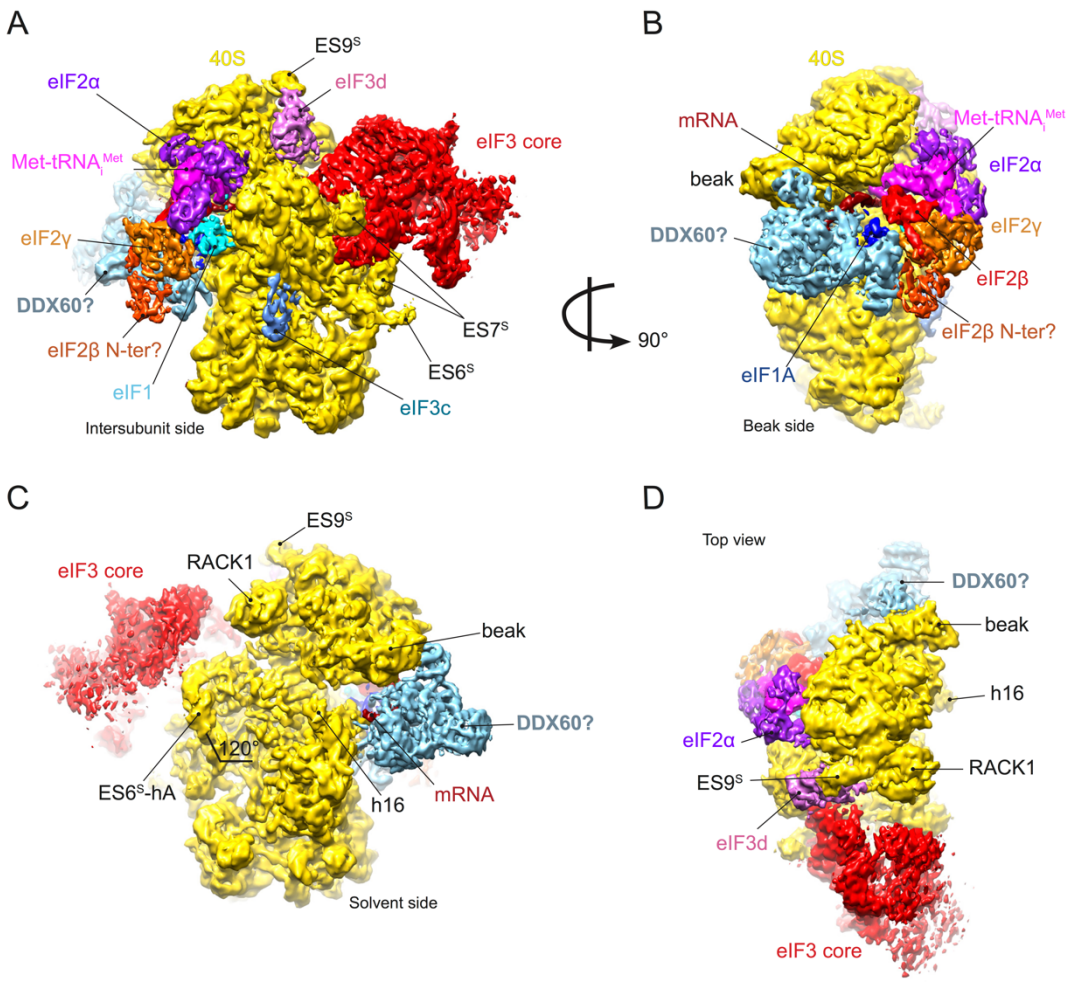


**Figure 2: Cryo-EM structure of kinetoplastids 48S pre-initiation complex.** Density segments corresponding to different initiation factors, 40S, Met-tRNA$_i^{Met}$, and mRNA are colored variably. **A**, Initiation complex viewed from the intersubunit side highlighting all the initiation factors. **B**, Beak side view highlighting the additional density at the intersubunit side, close to the mRNA channel. **C**, Solvent side view highlighting the conformation of the ES6$^S$-hA. D, top view showing the distribution of the factors around the head.

Finally, the cryo-EM reconstruction revels also an additional density that seems to be the mRNA. It starts near the helix 16 and threads into the mRNA channel towards the exit channel (Figures 2B-C). At the P-site it is interacting with the Met-tRNA$_i^{Met}$. Unexpectedly, our reconstruction revels a large mass of density at the 40S intersubunit face, in close vicinity to the mRNA channel (Figures 2B-D). Remarkably, the mRNA passes through the core of this unidentified density, which suggest that it may belong to a helicase (DDX60), as discussed later in this manuscript.

**The interaction between the mRNA and the anticodon stem–loop within 48S-open**

It has been shown that the treatment of mammalian cell lines (RRL) with GMP-PNP, yields 48S-close with the Met-tRNA$_i^{Met}$ full accommodated in the P-site ($P_{IN}$) (Simonetti et al., 2016). However, following the same procedure with *T. cruzi* cell lysate leads to an accumulation of 48S-open, in which the Met-tRNA$_i^{Met}$ is not fully engaged in the P-site, $P_{OUT}$ conformation (Figures 3A-B). Indeed, here the anticodon stem–loop (ASL) is lifted upwards when compare to that observed in m48S IC. In contrast to the Met-tRNA$_i^{Met}$ in context of m48S-close here the Met-tRNA$_i^{Met}$ does not interact with any elements from the body of 40S (Figure 3B). The Met-tRNA$_i^{Met}$ within kinetoplastidian 48S-open present a conformation similar to that observed in yeast 48S-open (Llácer et al., 2015).

Genetic studied indicated that eIF2β interacts with Met-tRNA$_i^{Met}$ through the residues S202 and K214. That interaction is important for the stabilization of the open conformation, which prevents initiation at Near-Cognate (Llácer et al., 2015). Accordingly, here we also observed several interactions between the ASL and eIF2β, which is preventing the full accommodation of the Met-tRNA$_i^{Met}$ in the P-site and prevents translation initiation at Near-Cognate codons. This observation comforts the role of eIF2β in ensuring the fidelity of translation initiation.

eIF1 plays an important role in the stabilization of the open conformation (Llácer et al., 2015; Pestova and Kolupaeva, 2002). Nevertheless, due to its location adjacent to the P-site, its release is mandatory for the formation of stable codon-anticodon base pairing and the close conformation (Lomakin et al., 2003). Accordingly, our particles image sorting revealed a class displaying the 48S complex in the closed conformation (48S-close) at lower resolution (~8), which does not present neither eIF2β nor eIF1 (Figure S5). Our structures suggest that the role of eIF1 in preventing the premature hydrolysis of eIF2-bound GTP is also conserved in kinetoplastids.

A remarkable difference between kinetoplastids and their mammalian host lies at the codon-anticodon base pairing level. The use of GMP-PNP to treat mammalian cell lysate lead the arrest of translation initiation complex with a full codon-anticodon base-pairing (Simonetti et al., 2016). Here, only a single base pair of the codon-anticodon interaction appears to be established (Figure 3C). This interaction occurs between the C base of the ASL with a base within the 5' UTR mRNA. This kinetoplastid singular behavior can be explained by the presence of kinetoplastid-specific player during translation initiation, i.e. as mentioned above,

in kinetoplastids DDX60 seems to play a role during translation initiation, perhaps different from any other known eukaryotes.

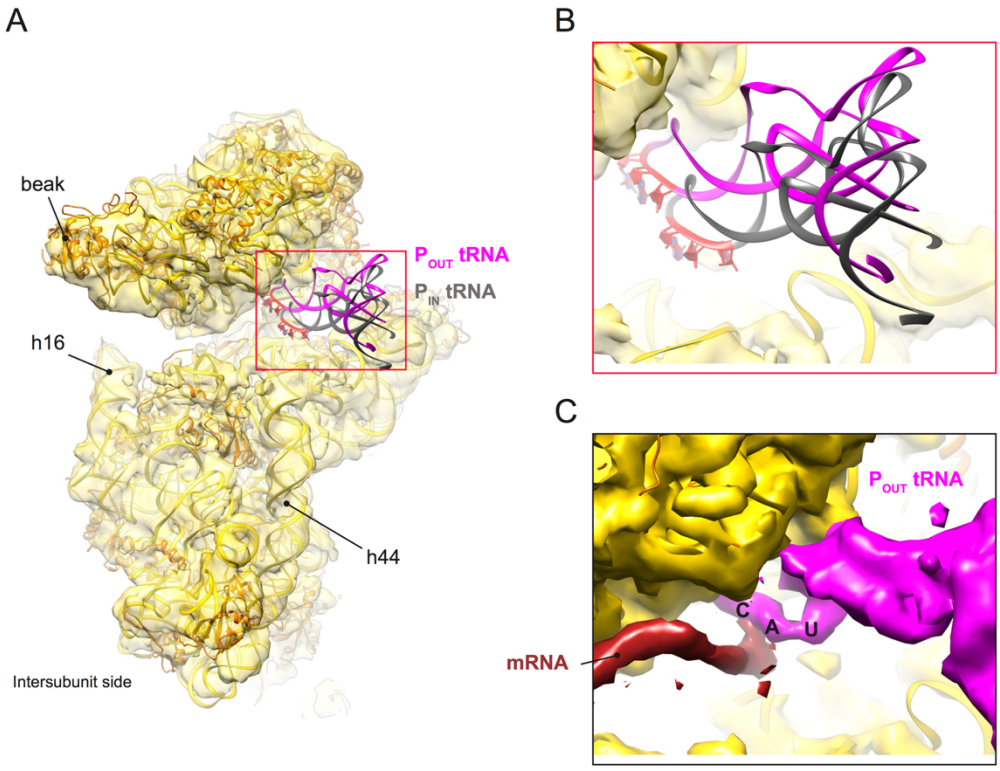


**Figure 3: Met-tRNA$_i^{Met}$ accommodation and interactions within 48S-open. A**, Superposition of kinetoplastid $P_{OUT}$ Met-tRNA$_i^{Met}$ with mammalian $P_{IN}$ Met-tRNA$_i^{Met}$ (Simonetti et al., 2016). **B**, Close-up view of the P-site, highlighting the distinct position of the two tRNA. **C**, Close-up view of the P-site, highlighting the interaction between the Met-tRNA$_i^{Met}$ with the mRNA.

## Kinetoplastid unusual interactions between eIF3 and expansion segments

Kinetoplastids do not require eIF3m subunit, therefore the left leg of the heptameric eIF3 structural core is residual (Figures 4A) when compared to that five-lobed conformation observed in mammalian octameric eIF3 structural core (Figures 4B) (Hashem et al., 2013b; des Georges et al., 2015). Furthermore, the binding of kinetoplastidian eIF3 to the 40S is different from that observed in mammals.

ES6$^S$ and ES7$^S$ are positioned close to the mRNA exit channel, in close vicinity to the eIF3 binding site (Figure 4C-D). In mammals, the only interaction between eIF3 and ES6$^S$ or ES7$^S$ is that observed between the apical loop of ES7$^S$ (Figure 4D) with eIF3a and eIF3c (des Georges et al., 2015). Kinetoplastids is known for their larger ESs, including ES6$^S$, ES7$^S$ and ES9$^S$ (Hashem et al., 2013a). Indeed, kinetoplastidian ES7$^S$ is about 7 times larger than its

homolog in mammals. Therefore, in contrast to what is observed in mammals, in kinetoplastids the ES7$^S$ is the main binding point of eIF3 core. Compared to its conformation in the context of native 40S (Querido et al., 2017), ES7$^S$ within 48S-open is bent to the intersubunit side and acts as a cradle for the accommodation of eIF3 (Figure 4C). Indeed, unlike in any other eukaryotes, eIF3c is sandwiched by ES7$^S$-helix A (ES7$^S$-hA) and ES7$^S$-hB (Figures 4C), thus forming a kinetoplastid-specific interaction (Figure 4D).

The size of ES6$^S$ varies among different species of kinetoplastids (429nt to 512nt). However, is about twice lager than that observed in most of in mammals (~230nt). Here, we report an interaction between kinetoplastidian ES6$^S$ and the eIF3 core, through ES6$^S$-hE and eIF3e contact point (Figure S6). Because of their unusual larger and structurally different ES6$^S$, is very unlikely that this interaction can also occur in their mammalian host. These intricate interactions between eIF3 core and ES6$^S$ and ES7$^S$ seem to be the main mechanism through which kinetoplastids can bind and stabilize the eIF3 core even without the eIF3m subunit, inexistent in kinetoplastids.

Another remarkable difference between kinetoplastids and their mammalian hosts lies in the kinetoplastid-specific interaction between eIF3d and ES9$^S$. Kinetoplastidian eIF3d binds close to the mRNA exit channel (Figure 3E), as described before in mammals (Figure 3F) (des Georges et al., 2015). However, kinetoplastidian eIF3d binds ES9$^S$ (Figure 3E). It has been shown that kinetoplastids and *Drosophila melanogaster* possess unusual larger ES9$^S$ when compared to its counterpart observed in mammals (Hashem et al., 2013a; Anger et al., 2013). In *D. melanogaster* the larger ES9$^S$ is in a conformation reaching towards the 40S beak (Anger et al., 2013), while in kinetoplastids ES9$^S$ points towards the mRNA exit channel (Hashem et al., 2013a). That architecture allows kinetoplastidian ES9$^S$ to plays an important role in translation initiation by binding and stabilize eIF3d, which in these organisms might be a mandatory subunit that is required for translation of all mRNA, in contrast to other eukaryotic organisms.

Finally, another remarkable difference between kinetoplastids and other known eukaryotic similar complexes relies in the conformation of their larger ES6$^S$-hA. In mammals ES6$^S$-hA binds and stabilize eIF3b (des Georges et al., 2015). However, here ES6$^S$-hA is reoriented by 120° in the direction of the 40S head, when compared to that in mammals and yeast (Figures 2C). This unusual conformation makes any interaction between ES6$^S$-hA and eIF3b unlikely to occurs. Accordingly, our particles image sorting revealed one class of 48S-open at lower resolution (~13 Å), which presents additional densities at the solvent site (Figure S7) close to the h16, at the exact binding site of the eIF3b-i-g module in mammals (des Georges

et al., 2015). However, even in that class the conformation of ES6$^S$-hA is exactly the same as in the main 6.0 Å resolution complex presented here. Therefore, our results indicated that kinetoplastidian ES6$^S$-hA does not bind eIF3b.

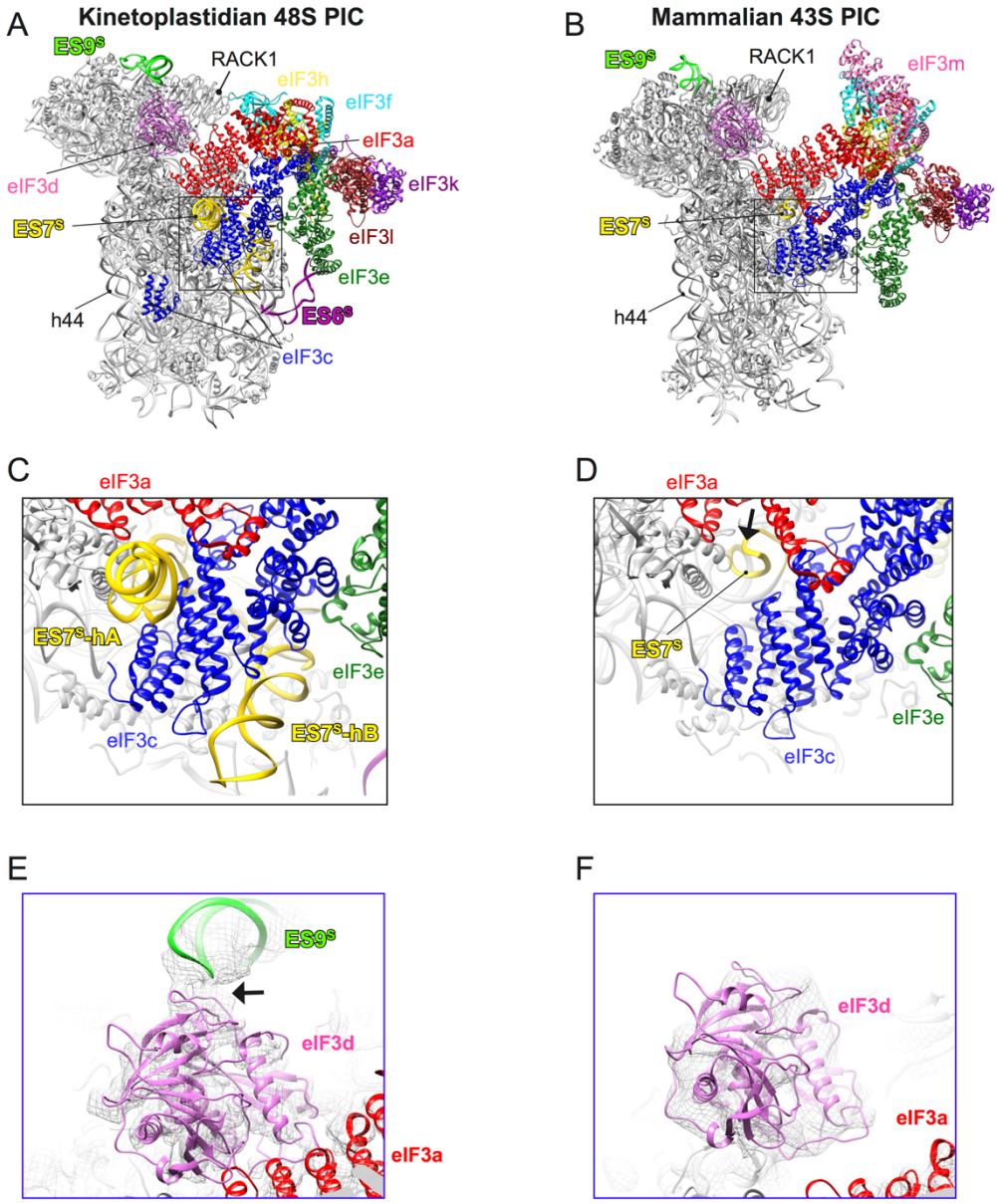


**Figure 4: Side-by-side comparison of eIF3-40S subunit interactions within kinetoplastidian 48S-open and mammalian 43S PIC. A-B,** Structure of kinetoplastidian 48S-open (left) and mammalian 43S PIC (right) as presented previously (des Georges et al., 2015). **C**, Close-up view of the contacts between ES7$^S$ and eIF3a and eIF3c in kinetoplastids (left). **D**, Close-up view of the interaction between the apical loop of ES7$^S$ with mammalian eIF3a (right). **E-F**, Close-up view highlighting the interaction between kinetoplastid ES9$^S$ and eIF3d, which is not present in mammals (right). Black arrows indicate sites of interaction.

Put together, the unusual larger ESs from kinetoplastids seems to be an important player during translation initiation, by binding and stabilizing the eIF3 core but also eIF3d, probably tuning translation initiation in one preferential setup specialized for initiating on spliced-leader mRNAs of kinetoplastids.

**Putative location of DDX60 and its interaction with mRNA**

Our cryo-EM reconstruction revels an unexpected density mass at the intersubunit side in close vicinity to the mRNA channel latch (Figure 4A-B). This density between eIF2β and 40S beak has not been described in any other equivalent complex from other eukaryotic organisms. The fitting of atomic models from known eIFs into this density enables us to discard its attribution to any known eIFs found in our MS/MS analysis. Indeed, we were able to identify the biding site of the majority of eIFs, including that of peripheral subunits b, i and g of eIF3 (Figure S7). However, rigid-body fitting of the conserved structural core of helicases indicated that this density contains a helicases core (Figure S8). Furthermore, the density that corresponds to the mRNA passes through the hypothetical core of the helicase in its trajectory from the vicinity of h16 towards the mRNA exit channel (Figure 4C). Therefore, based on rigid-body fitting, size, shape and proteomic analysis (Figure S1), we have tentatively assigned this density to DDX60. However, due to limitation imposed by the resolution, we still need to validate this assignment.

RNA helicases of the DEAD or DEAH box families have been reported in all eukaryotic cells. They are involved in several process, including RNA transcription, translation and degradation. All the enzymes which belong to these families are NTPase, able to unwind RNA. Indeed, several structural studies have indicated that members of the DEAD box enzymes possess a highly conserved structural core, composed by an ATP and RNA biding domains (Del Campo and Lambowitz, 2009; Sengoku et al., 2006; Andersen et al., 2006). The conserved core harbors two recombinase A (RecA)-like helicase domains. Furthermore, the structures of RNA-bound DEAD box enzymes reveals clearly a conserved binding mode of RNA, which normally involve only the RNA sugar phosphate backbone (Del Campo and Lambowitz, 2009; Sengoku et al., 2006; Andersen et al., 2006). However, in spite of their high conservation, the functions of DEAD box proteins are very distinct (Linder and Jankowsky, 2011). The conserved core is surrounded by variable auxiliary domains, which determine the function of each DEAD box helicase (Linder and Jankowsky, 2011). Human DDX60 is known for its role in RIG-I-mediated innate immune response against viral RNA (Oshiumi et al., 2015). Nevertheless,

kinetoplastidian DDX60 (~246 kDa) is larger than that in humans (~197 kDa), which may confer them additional auxiliary domains that are not present in human. Together, these observations indicate that kinetoplastidian DDX60 is able to play different role from that described in mammals.

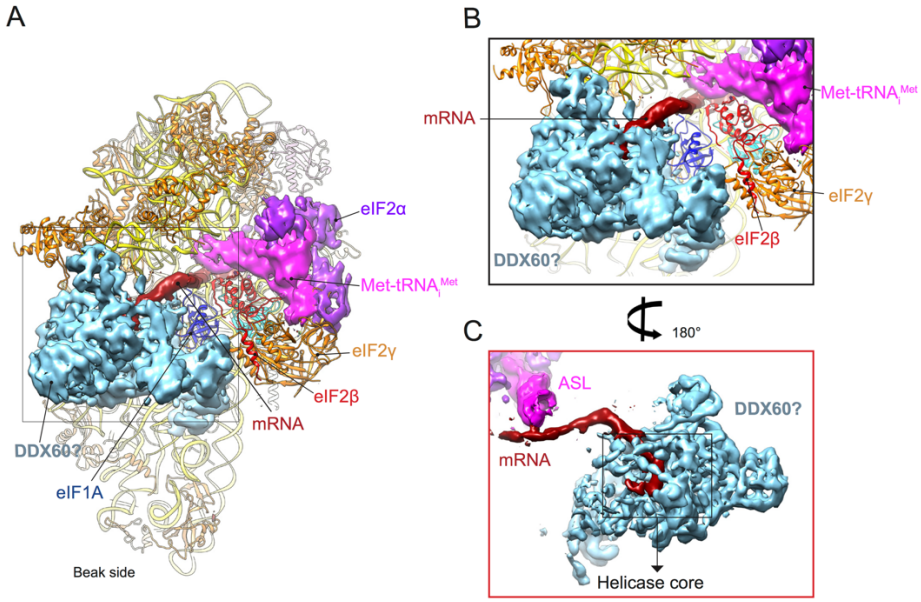


**Figure 5: Mass density potentially from DDX60 and its interaction with mRNA. A,** The location of the additional density that potentially belongs to DDX60. **B-C**, Close-up view highlighting the interaction between mRNA and the density that we tentatively assigned to DDX60.

DDX60 is not the only helicase that may play different roles according to the cell type. Mammalian DHX29 also can play a dual role. For years, DHX29 has been studied for its role in translation initiation (Pisareva et al., 2008; Hashem et al., 2013b). Nevertheless, a recent study has indicated that DHX29 can play a role in RIG-I-mediated innate immune response against viral RNA in human airway epithelial cells (Sugimoto et al., 2014). Therefore, we believe that is plausible to consider that the activity of DDX60 is also cell-specific, either by unwinding the mRNA during translation initiation in kinetoplastids or by co-sensing the viral mRNA and trigger the antiviral immunity in mammals.

Analysis of spliced-leader RNA from *T. brucei* indicates the presence of several short RNA duplex structure (Harris et al., 1995). Therefore, based on our results and previous reports, we propose that kinetoplastids may require an auxiliary helicase for the unwinding of the secondary structure of the spliced leader mRNA. It is known that most of the DEAD box helicases require the biding of ATP to catalyze the unwinding of short RNA duplexes by local

strand separation (Yang et al., 2007; Mallam et al., 2012). The binding of ATP is important for the closed conformation of the core, which is able to unwinding the RNA duplex (Mallam et al., 2012). Although the hydrolysis of ATP is not required for the RNA unwinding by DEAD box helicase, it was found that it is necessary for the release of the helicase from its RNA substrate (Liu et al., 2008). Accordingly, here we observe that the use of a non-hydrolysable nucleotide (GMP-PNP) arrests the helicase DDX60 bound to its mRNA substrate (Figure 4C).

## Conclusions

The cryo-EM reconstruction of kinetoplastidian 48S pre-initiation complex presented here enabled us to characterize the near complete structure of the 48S complex and determine the exact ribosomal position of majority of eIFs (Figure 5), excepting for eIF5.

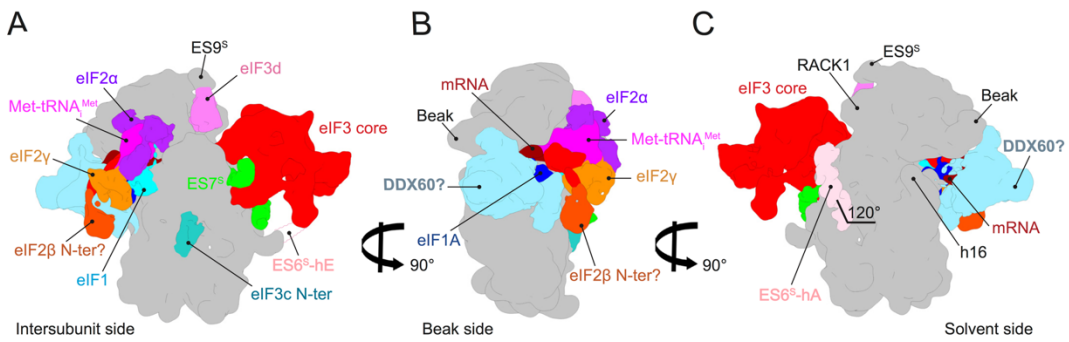


**Figure 6: Schematic representation of the kinetoplastidian 48S pre-initiation complex. A-C,** The present here only included the eIFs and subunits present in the reconstruction obtainded at 6Å resolution. The model is shown in different orientation, highlighting the intricate interaction between ESs and eIF3 core and eIF3d. Furthermore, it is also highlighting the possible position of DDX60 in context of kinetoplastidian initiation complex. **A-B**, Is also highlighting the position of the kinetoplastid-specific N-terminal domain of eIF2α as well as the possible location of the N-terminal residues of eIF2β. **C**, Is highlighting the kinetopasd-specific conformation of ES6$^{S}$-hA, which prevents its biding to the eIF3b.

The structure of auxiliary helicases in context of translation initiation complex was solved for mammalian DHX29 (des Georges et al., 2015; Hashem et al., 2013b). However, these reconstructions were obtained in a complex void of mRNA (43S pre-initiation complex). Therefore, here we present for the first time a cryo-EM reconstruction showing the binding site of a hypothetical auxiliary helicase and its interaction with mRNA in context of 48S translation pre-initiation complex. DDX60 seems to play a specific role in translation initiation in kinetoplastids by unwinding the 5′ UTR of the spliced-leader mRNA. It is worth mentioning

that the binding of DDX60 prevents the rotation of the 40S head toward the intersubunit side and the closure of the mRNA latch. Therefore, its release is mandatory for the formation of 48S-close conformation and full codon-anticodon base-pairing. Considering, that DEAD box helicases are poorly processive helicases, after each unwinding of short duplex they should be released from their substrate RNA. In this context, after the unwinding of the short duplexes within the spliced-leader mRNA, the nucleotide bound to DDX60 is hydrolyzed. The hydrolysis leads to its release from the 48S-open and the rotation of the 40S head to forms 48S-close initiation complex.

Kinetoplastids present several specie-specificities, however our reconstructions revel that they have conserved the global architecture of eIF3 core composed seven PCI/MPN subunit (ceIF3a, eIF3c, eIF3e, eIF3f, eIF3h, eIF3k and eIF3l) (Masutani et al., 2007). Furthermore, the architecture of eIF3 is comparable to that observed in mammals with the exception of eIF3m that doesn't exist in kinetoplastids. The architecture of eIF3 in the context of kinetoplastidian translation initiation complexes revels that the binding and the stabilization of the septamer core is guaranteed only by the central subunits, eIF3a, eIF3c and eIF3e, which may explain the absence eIF3m that would play a minor structuring role in this context. Most importantly, the accommodation of eIF3c within $ES7^S$ and the interaction of eIF3e with $ES6^S$ are kinetoplastid-specific. Furthermore, the binding and the stabilization of eIF3d by $ES9^S$ appears so far to be kinetoplastids-specific.

Put together, our results present the first structural insights into the mechanism of translation initiation and scanning in the human pathogen *T. cruzi*. The kinetoplastidian peculiarities described here may offer new perspectives for the research and development of translation-targeted therapeutic agents against kinetoplastids.

**Author Contributions**

J. B. Q., and Y. H. designed, analyzed and interpreted molecular and biochemical data experiments. J. B. Q., purified the complexes and performed the biochemical experiments. Y. H. carried out cryo-EM data processing. J.B.Q and A. B. performed the preliminary molecular modeling. J.B.Q., A. B., A.S., and Y.H. interpreted the structure. J. B. Q. and Y. H. wrote the manuscript. L.K. performed MS/MS analysis. All authors contributed to the discussion and the writing of the manuscript. Y. H. directed research.

**References**

Andersen, C.B.F., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L.P., Pedersen, J.S., Séraphin, B., Le Hir, H., and Andersen, G.R. (2006). Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. Science 313, 1968–1972.

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.

Chen, Y., Potratz, J.P., Tijerina, P., Del Campo, M., Lambowitz, A.M., and Russell, R. (2008). DEAD-box proteins can completely separate an RNA duplex using a single ATP. Proc. Natl. Acad. Sci. U. S. A. 105, 20203–20208.

Clayton, J. (2010). Chagas disease: pushing through the pipeline. Nature 465, S12–S15.

Del Campo, M., and Lambowitz, A.M. (2009). Structure of the Yeast DEAD box protein Mss116p reveals two wedges that crimp RNA. Mol. Cell 35, 598–609.

Dong, J., Lai, R., Nielsen, K., Fekete, C.A., Qiu, H., and Hinnebusch, A.G. (2004). The essential ATP-binding cassette protein RLI1 functions in translation by promoting preinitiation complex assembly. J. Biol. Chem. 279, 42157–42168.

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. advance online publication.

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.

Harris, K.A., Crothers, D.M., and Ullu, E. (1995). In vivo structural analysis of spliced leader RNAs in Trypanosoma brucei and Leptomonas collosoma: a flexible structure that is independent of cap4 methylations. RNA 1, 351–362.

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

Hashem, Y., des Georges, A., Dhote, V., Langlois, R., Liao, H.Y., Grassucci, R.A., Hellen, C.U.T., Pestova, T.V., and Frank, J. (2013b). Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. Cell 153, 1108–1119.

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. advance online publication.

Hinnebusch, A.G. (2017). Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation. Trends Biochem. Sci.

Hussain, T., Llácer, J.L., Fernández, 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.

Li, K., Zhou, S., Guo, Q., Chen, X., Lai, D.-H., Lun, Z.-R., and Guo, X. (2017). The eIF3 complex of Trypanosoma brucei: composition conservation does not imply the conservation of structural assembly and subunits function. RNA N. Y. N 23, 333–345.

Lin, K., Simossis, V.A., Taylor, W.R., and Heringa, J. (2005). A simple and fast secondary structure prediction method using hidden neural networks. Bioinforma. Oxf. Engl. 21, 152–159.

Linder, P., and Jankowsky, E. (2011). From unwinding to clamping — the DEAD box RNA helicase family. Nat. Rev. Mol. Cell Biol. 12, 505–516.

Liu, F., Putnam, A., and Jankowsky, E. (2008). ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. Proc. Natl. Acad. Sci. U. S. A. 105, 20209–20214.

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. U. S. A. 113, 12174–12179.

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

Lomakin, I.B., Kolupaeva, V.G., Marintchev, A., Wagner, G., and Pestova, T.V. (2003). Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. Genes Dev. 17, 2786–2797.

Mallam, A.L., Del Campo, M., Gilman, B., Sidote, D.J., and Lambowitz, A.M. (2012). Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. Nature 490, 121–125.

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. 0.

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.

Masutani, M., Sonenberg, N., Yokoyama, S., and Imataka, H. (2007). Reconstitution reveals the functional core of mammalian eIF3. EMBO J. 26, 3373–3383.

Meleppattu, S., Kamus-Elimeleh, D., Zinoviev, A., Cohen-Mor, S., Orr, I., and Shapira, M. (2015). The eIF3 complex of Leishmania—subunit composition and mode of recruitment to different cap-binding complexes. Nucleic Acids Res. 43, 6222–6235.

Merrick, W.C. (1979). Evidence that a single GTP is used in the formation of 80 S initiation complexes. J. Biol. Chem. 254, 3708–3711.

Miyashita, M., Oshiumi, H., Matsumoto, M., and Seya, T. (2011). DDX60, a DEXD/H Box Helicase, Is a Novel Antiviral Factor Promoting RIG-I-Like Receptor-Mediated Signaling▽. Mol. Cell. Biol. 31, 3802–3819.

Oshiumi, H., Miyashita, M., Okamoto, M., Morioka, Y., Okabe, M., Matsumoto, M., and Seya, T. (2015). DDX60 Is Involved in RIG-I-Dependent and Independent Antiviral Responses, and Its Function Is Attenuated by Virus-Induced EGFR Activation. Cell Rep. 11, 1193–1207.

Parsyan, A., Svitkin, Y., Shahbazian, D., Gkogkas, C., Lasko, P., Merrick, W.C., and Sonenberg, N. (2011). mRNA helicases: the tacticians of translational control. Nat. Rev. Mol. Cell Biol. 12, 235–245.

Perry, K.L., Watkins, K.P., and Agabian, N. (1987). Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. Proc. Natl. Acad. Sci. U. S. A. 84, 8190–8194.

Pestova, T.V., and Kolupaeva, V.G. (2002). The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. Genes Dev. 16, 2906–2922.

Phan, L., Zhang, X., Asano, K., Anderson, J., Vornlocher, H.P., Greenberg, J.R., Qin, J., and Hinnebusch, A.G. (1998). Identification of a translation initiation factor 3 (eIF3) core complex, conserved in yeast and mammals, that interacts with eIF5. Mol. Cell. Biol. 18, 4935–4946.

Pisareva, V.P., and Pisarev, A.V. (2016). DHX29 and eIF3 cooperate in ribosomal scanning on structured mRNAs during translation initiation. RNA N. Y. N 22, 1859–1870.

Pisareva, V.P., Pisarev, A.V., Komar, A.A., Hellen, C.U.T., and Pestova, T.V. (2008). Translation initiation on mammalian mRNAs with structured 5′-UTRs requires DExH-box protein DHX29. Cell 135, 1237–1250.

Querido, J.B., Mancera-Martínez, E., Vicens, Q., Bochler, A., Chicher, J., Simonetti, A., and Hashem, Y. (2017). The cryo-EM Structure of a Novel 40S Kinetoplastid-Specific Ribosomal Protein. Structure 0.

Rezende, A.M., Assis, L.A., Nunes, E.C., da Costa Lima, T.D., Marchini, F.K., Freire, E.R., Reis, C.R., and de Melo Neto, O.P. (2014). The translation initiation complex eIF3 in trypanosomatids and other pathogenic excavates – identification of conserved and divergent features based on orthologue analysis. BMC Genomics 15.

Schmitt, E., Panvert, M., Lazennec-Schurdevin, C., Coureux, P.-D., Perez, J., Thompson, A., and Mechulam, Y. (2012). Structure of the ternary initiation complex aIF2-GDPNP-methionylated initiator tRNA. Nat. Struct. Mol. Biol. 19, 450–454.

Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., and Yokoyama, S. (2006). Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. Cell 125, 287–300.

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.

Sugimoto, N., Mitoma, H., Kim, T., Hanabuchi, S., and Liu, Y.-J. (2014). Helicase proteins DHX29 and RIG-I cosense cytosolic nucleic acids in the human airway system. Proc. Natl. Acad. Sci. U. S. A. 111, 7747–7752.

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

Yang, Q., Del Campo, M., Lambowitz, A.M., and Jankowsky, E. (2007). DEAD-box proteins unwind duplexes by local strand separation. Mol. Cell 28, 253–263.

Xiol, J., Spinelli, P., Laussmann, M.A., Homolka, D., Yang, Z., Cora, E., Couté, Y., Conn, S., Kadlec, J., Sachidanandam, R., et al. (2014). RNA Clamping by Vasa Assembles a piRNA Amplifier Complex on Transposon Transcripts. Cell 157, 1698–1711.

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.

## Materials and methods

### Kinetoplastids cultures

*Trypanosoma cruzi* epimastigotes were grown at 28ºC 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.

### Yeast culture

*Saccharomyces cerevisiae* strain JD1370 were grown in YPD media (yeast extract 10 g/L, peptone 20 g/L, glucose 20 g/L) at 30°C with orbital agitation of 150 rpm. The cells were harvested when the OD is bellow 1.5.

### Native initiation complex isolation

*T. cruzi, L. tarent*olae and *Saccharomyces cerevisiae* native initiation complex 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 incubated with 10 mM GMP-PNP (Sigma) for 10 min on ice, followed by 10 min incubation at 28 °C. The suspension 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 43S and 48S initiation complex were collected and pooled according the UV absorbance profile and pelleted. The pelleted ribosomal complex was re-suspended in sucrose-free buffer II (10 mM HEPES-KOH pH 7.4, 50 mM KOAc, 10 mM NH4Cl, 5 mM Mg(OAc)2, and 2 mM DTT).

Mammalian near native initiation complex was purified as described before (Simonetti et al., 2016). In brief, 48S was purified from 60 µl of nuclease-treated RRL (Promega) supplemented with 13 µg of in vitro transcribed human β-globin mRNA and fractionated in 5%–25% linear sucrose gradient.

**NanoLC-MS/MS analysis**

The mass spectrometry (MS/MS) analysis of liquid translation initiation complexes samples were performed as described before (Chicher et al., 2015). Briefly, the precipitated proteins were subjected to reduction and alkylation, followed by digestion 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*, *L. major*, *Saccharomyces cerevisiae* and *Oryctolagus cuniculus* 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 for kinetoplastidian samples.

The volcano plot was obtained from the analysis of three independent replicates. The spectral count was analyzed by GLM regression statistical test using EdgeR package. For each protein identified here, the adjusted P-value and the fold-change (FC) were calculated and plotted. Each protein found in 48S PIC was manually validated by comparing with MS/MS results of the sample analyzed after size-exclusion chromatography.

**Cryo-EM data acquisition**

The cryo grids were prepared by apply 4 µL of 130 nM initiation complex onto mesh holey carbon Quantifoil 2/2 grids (Quantifoil Micro Tools) coated with an additional continuous thin layer of carbon. The grids were previously submitted to plasma cleaning. They were blotted for 1.5 sec (blot force 5) at 4ºC, 100% humidity and using waiting time 30 s (Vitrobot Mark IV). The data acquisitions were performed using a Talos Arctica instrument (FEI) operated at 200 kV acceleration voltage and at a nominal underfocus of $\Delta z = -0.6$ to $-4.0$ $\mu$m. The microscope is equipped with the third-generation back-thinned direct electron detector CMOS (III) 4,096 x 4,096 camera and automated data collection with EPU software (FEI). The Falcon III camera was calibrated at nominal magnification of 75,000 x. The calibrated magnification on the 14 $\mu$m pixel camera is 134,615 x, resulting in 1.16 Å pixel size at the specimen level. The camera was set up to collect 20. Total exposure was 1.5 s, with a dose of 24 /Å2.

**Image processing**

The data process and 3D reconstruction was performed using SCIPION (Abrishami et al., 2015; de la Rosa-Trevín et al., 2016; Sorzano et al., 2015) package which integrates several data

processing software. The movie alignment of ~6000 images were performed using Optical Flow algorithm integrated in Xmipp3 (de la Rosa-Trevín et al., 2013). CTFFIND4 (Rohou and Grigorieff, 2015) was used for the estimation of defocus for particles selection. After manual selection of the images by analyzing their corresponding Fourier transform, approximately 950,000 particles were selected using SCIPION (Abrishami et al., 2013). RELION (Scheres, 2012) was used for particle sorting through 3D classification via SCIPION, refer to Extended Data Figure S3 for particle sorting details. The refinement of the selected classes was performed using RELION's 3D autorefine. 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 6.0 Å (Figure S3).

The cryo-EM map was 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 40S ribosome from *Trypanosoma cruzi* (Querido et al., 2017) was used to derive a preliminary atomic model of the *T. cruzi* 48S initiation complex ribosomal complexes. The atomic models of the initiation factors were derived by homology to the crystal and cryo-EM structures of yeast and mammals initiation complex (Llácer et al., 2015; des Georges et al., 2015) using SWISS-MODEL (Arnold et al., 2006). The atomic models of eIFs were placed into its corresponding cryo-EM density map segment by rigid- body fitting using Chimera.

# Supplementary Figures

| | | | BASIC Spectral Count (# spectra) | |
|---|---|---|---|---|
| | | | BEFORE Gel Filtration | AFTER Gel Filtration |
| Q4E5Z1\|Q4E | DDX60 | Uncharacterized protein OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508153.1050 PE=4 SV=1 | 263 | 96 |
| Q4DLI2\|Q4DL | ABCE1 | Ribonuclease L inhibitor, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508637.150 PE=3 SV | 103 | 31 |
| | | | | |
| | | | | |
| **40S ribosomal proteins:** | | | | |
| | | | BASIC Spectral Count (# spectra) | |
| | | | BEFORE Gel Filtration | AFTER Gel Filtration |
| accession | | description | 48S | 48S |
| Q4D5P4\|Q4D | | 40S ribosomal protein S4 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053509683.117 PE=3 SV=1 | 131 | 93 |
| Q4DTN2\|Q4D | | Activated protein kinase C receptor, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511211.12 | 96 | 48 |
| Q4E0Q3\|Q4E | | 40S ribosomal protein S5, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506297.150 PE=3 SV | 51 | 40 |
| Q4DZ41\|RS3 | | 40S ribosomal protein S3a-2 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511001.9 PE=3 SV=1 | 84 | 46 |
| Q4E093\|Q4E | | 40S ribosomal protein S18, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506679.100 PE=3 S | 73 | 64 |
| Q4DSU0\|Q4D | | 40S ribosomal protein S6 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053510769.49 PE=3 SV=1 | 72 | 58 |
| Q4CLU9\|Q4C | | 40S ribosomal protein S8 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511069.20 PE=3 SV=1 | 60 | 46 |
| Q4D4L4\|Q4D | | 40S ribosomal protein S11, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053507837.50 PE=3 SV | 47 | 39 |
| Q4D6I5\|Q4D6 | | 40S ribosomal protein S14, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053409117.20 PE=3 SV | 61 | 37 |
| Q4CUL0\|Q4C | | 40S ribosomal protein S3, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053430605.29 PE=3 SV | 72 | 37 |
| Q4D6N9\|Q4D | | Ribosomal protein S19, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053504013.100 PE=4 SV= | 36 | 36 |
| Q4DY30\|Q4D | KSRP | RNA-binding protein, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511727.290 PE=4 SV=1 | 72 | 29 |
| Q4D4S1\|Q4D | | 40S ribosomal protein S9, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053504163.30 PE=3 SV | 38 | 28 |
| Q4CUC8\|Q4C | | Ribosomal protein S7, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506593.19 PE=4 SV=1 | 80 | 25 |
| Q4CQU0\|Q4C | | 40S ribosomal protein SA OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053503719.20 PE=3 SV=1 | 58 | 22 |
| Q4D916\|Q4D | | 40S ribosomal protein S16, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053507513.60 PE=4 SV | 52 | 19 |
| Q4E0N6\|Q4E | | 40S ribosomal protein S15a, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506297.330 PE=3 | 34 | 15 |
| Q4DIZ9\|Q4DL | | 40S ribosomal protein S2, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053503833.40 PE=3 SV | 74 | 27 |
| Q4CXN0\|Q4C | | Ubiquitin/ribosomal protein S27a, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053510293.40 PE | 40 | 14 |
| Q4DTX6\|Q4D | | Ribosomal protein S25, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053504105.94 PE=4 SV=1 | 44 | 8 |
| Q4DK39\|Q4D | | 40S ribosomal protein S17, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508827.79 PE=3 SV | 57 | 16 |
| Q4E088\|Q4E | | 40S ribosomal protein S10, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506679.140 PE=4 S | 54 | 22 |
| Q4DW69\|Q4D | | 40S ribosomal protein S12 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508231.20 PE=3 SV=1 | 39 | 13 |
| Q4CXV6\|Q4C | | 40S ribosomal protein S33, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506413.30 PE=3 SV | 31 | 17 |
| Q4DTQ1\|Q4D | | 40S ribosomal protein S23, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053504181.20 PE=3 SV | 28 | 28 |
| Q4D6H7\|Q4D | | Ribosomal protein S20, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508475.10 PE=3 SV=1 | 28 | 16 |
| Q4CWD6\|Q4C | | 40S ribosomal protein S13, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053510029.70 PE=3 SV | 30 | 18 |
| Q4DN73\|Q4D | | 40S ribosomal protein S27, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506963.14 PE=3 SV | 17 | 25 |
| Q4DW38\|Q4D | | 40S ribosomal protein S24 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053507681.150 PE=3 SV=1 | 26 | 15 |
| Q4CMS5\|Q4C | | Ribosomal protein S29, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511015.20 PE=4 SV=1 | 15 | 13 |
| Q4DGZ5\|Q4D | | 40S ribosomal protein S15, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511809.130 PE=3 S | 20 | 11 |
| Q4CYE4\|Q4C | | Ribosomal protein S26, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053503801.20 PE=4 SV=1 | 18 | 12 |
| Q4E3L9\|Q4E | | 40S ribosomal protein S21, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053510101.430 PE=4 S | 18 | 7 |
| Q4DA48\|Q4D | | 40S ribosomal protein S30, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053507019.83 PE=4 SV | 5 | |
| | | | | |
| | | | | |
| **Initiation factors:** | | | | |
| | | | BASIC Spectral Count (# spectra) | |
| | | | BEFORE Gel Filtration | AFTER Gel Filtration |
| accession | | description | 48S | 48S |
| Q4DL69\|Q4D | eIF3a | Uncharacterized protein OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508919.140 PE=4 SV=1 | 129 | 50 |
| Q4DSL1\|Q4D | eIF3b | Translation initiation factor, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511303.60 PE=4 SV | 159 | 41 |
| Q4E3G1\|Q4E | eIF3c | Eukaryotic translation initiation factor 3 subunit 8, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.10470 | 96 | 16 |
| Q4D7F2\|Q4D | eIF3e | Eukaryotic translation initiation factor 3 subunit E OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053509205 | 103 | 24 |
| Q4E620\|Q4E | eIF2 alpha | Elongation initiation factor 2 alpha subunit, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.10470535081 | 105 | 22 |
| Q4DCN0\|Q4D | eIF3d | Eukaryotic translation initiation factor 3 subunit 7-like protein, putative OS=Trypanosoma cruzi (strain CL Brener) GN= | 113 | 16 |
| Q4D452\|Q4D | eIF3i | Eukaryotic translation initiation factor 3 subunit I OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511229 | 69 | 14 |
| Q4D5W3\|Q4D | eIF3l | Eukaryotic translation initiation factor 3 subunit L OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508169 | 83 | 25 |
| Q4E3S5\|Q4E | eIF3h | Homology with eIF3H (InterPro), Uncharacterized protein OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705 | 58 | 8 |
| Q4CUG4\|Q4C | eIF3g | Eukaryotic translation initiation factor 3 subunit G OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508685 | 70 | 19 |
| Q4CSE1\|Q4C | eIF5 | Eukaryotic translation initiation factor 5, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053504119 | 115 | 49 |
| Q4DDK1\|Q4D | eIF3k | Homology with eIF3K (InterPro), Uncharacterized protein OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705 | 29 | 9 |
| Q4DH88\|Q4D | eIF2 beta | Translation initiation factor, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053503955.70 PE=4 SV | 45 | 25 |
| Q4DQZ2\|Q4D | eIF3f | Uncharacterized protein OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053510089.200 PE=4 SV=1 | 68 | 24 |
| Q4CPV7\|Q4C | eIF2 gamma | Eukaryotic translation initiation factor 2 subunit, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053 | 62 | 7 |
| Q4CQB1\|Q4C | eIF1A | Eukaryotic translation initiation factor 1A, putative (Fragment) OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104 | 25 | 4 |
| Q4DM75\|Q4D | eIF1 | Protein translation factor SUI1 homolog, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508515 | 18 | 5 |
| | | | | |
| | | | | |

**Figure S 1: Mass spectrometry characterization of native initiation complex from *T. cruzi* before and after size-exclusion chromatography (SEC).** The initiation complexes obtained from *T. cruzi* cell lysate were analysed by MS/MS. The figure contains only 40S and initiation factors. The full list of proteins identified here will be online. The spectral counts were used for relative quantification. Darker colors indicate the most abundant protein present in the sample. Related to Figure 1.

A

## Leishmania

| accession | Name | description | 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=ABC | 101 |

**40S ribosomal proteins:**

| accession | Name | description | Spectral Count IC |
|---|---|---|---|
| tr\|Q868B1\|Q | | 40S ribosomal protein S5 OS=Leishmania major GN=LMJF_11_0960 PE=4 SV=1 | 188 |
| tr\|Q4Q216\|Q | | Putative ubiquitin/ribosomal protein S27a OS=Leishmania major GN=LMJF_36_0600 PE=4 SV | 265 |
| tr\|Q4Q1Y2\|Q | | Putative 40S ribosomal protein S18 OS=Leishmania major GN=LMJF_36_0940 PE=3 SV=1 | 122 |
| tr\|Q4QG31\| | | 40S ribosomal protein S4 OS=Leishmania major GN=RS4 PE=2 SV=1 | 299 |
| tr\|Q4Q8H1\| | | 40S ribosomal protein S14 OS=Leishmania major GN=LMJF_28_0960 PE=3 SV=1 | 155 |
| tr\|Q4QC89\| | | Putative 40S ribosomal protein S23 OS=Leishmania major GN=LMJF_21_1060 PE=3 SV=1 | 90 |
| tr\|Q4Q4A0\| | | Putative 40S ribosomal protein S3 OS=Leishmania major GN=LMJF_15_0950 PE=4 SV=1 | 99 |
| sp\|P25204\| | | 40S ribosomal protein S8 OS=Leishmania major GN=RPS8A PE=3 SV=1 | 108 |
| sp\|Q9NE83\| | | 40S ribosomal protein S6 OS=Leishmania major GN=RPS6 PE=3 SV=1 | 175 |
| tr\|Q4Q817\|Q | | 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\|Q4Q0QC | | 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\|Q | | 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\|Q | | Putative 40S ribosomal protein S17 OS=Leishmania major GN=LMJF_28_2555 PE=3 SV=1 | 42 |
| tr\|Q4Q140\|Q | | Putative 40S ribosomal protein S27-1 OS=Leishmania major GN=LMJF_36_3750 PE=3 SV=1 | 63 |
| tr\|Q4Q8L6\|Q | | 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:**

| accession | Name | description | Spectral Count (# sp IC |
|---|---|---|---|
| tr\|Q4QEJ8\|Q | eIF3a | Uncharacterized protein OS=Leishmania major GN=LMJF_17_0010 PE=4 SV=1 | 278 |
| tr\|Q4QE62\|Q | eIF3b | Putative translation initiation factor OS=Leishmania major GN=LMJF_17_1290 PE=4 SV=1 | 175 |
| tr\|Q4Q6Y6\|Q | eIF3d | Eukaryotic translation initiation factor 3 subunit 7-like protein OS=Leishmania major GN=LMJF | 125 |
| tr\|Q4Q833\|Q | eIF3e | Eukaryotic translation initiation factor 3 subunit E OS=Leishmania major GN=LMJF_28_2310 P | 84 |
| tr\|Q4Q253\|Q | eIF3l | Eukaryotic translation initiation factor 3 subunit L OS=Leishmania major GN=LMJF_36_0250 P | 91 |
| tr\|Q4Q127\|Q | 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\|Q | eIF3h | Uncharacterized protein OS=Leishmania major GN=LMJF_07_0640 PE=4 SV=1 | 76 |
| tr\|Q4Q3H3\|Q | eIF5 | Putative eukaryotic translation initiation factor 5 OS=Leishmania major GN=LMJF_34_0350 PE | 75 |
| tr\|Q4Q9T0\|Q | eIF3f | Uncharacterized protein OS=Leishmania major GN=LMJF_25_1610 PE=4 SV=1 | 67 |
| tr\|Q4Q055\|Q | eIF3c | Putative eukaryotic translation initiation factor 3 subunit 8 OS=Leishmania major GN=LMJF_36 | 62 |
| tr\|Q4Q557\|Q | 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\|Q | eIF3g | Eukaryotic translation initiation factor 3 subunit G OS=Leishmania major GN=LMJF_34_2700 P | 46 |
| tr\|Q4QAL1\|Q | eIF1A | Putative translation factor sui1 OS=Leishmania major GN=LMJF_24_1210 PE=4 SV=1 | 33 |
| tr\|Q4QIB4\|Q | eIF2 beta | Translation initiation factor-like protein OS=Leishmania major GN=LMJF_08_0550 PE=4 SV=1 | 29 |
| tr\|Q4QF06\|Q | eIF1A | Putative eukaryotic translation initiation factor 1A OS=Leishmania major GN=LMJF_16_0140 P | 27 |

B

## Yeast

| accession | Name | description | Spectral Count IC |
|---|---|---|---|
| | DDX60 | | 0 |
| sp\|Q03195\|RLI1_YEAS | Rli1/ABCE1 | Translation initiation factor RLI1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Bake | 116 |

**40S ribosomal proteins:**

| accession | | description | Spectral Count IC |
|---|---|---|---|
| tr\|E7NL95\|E7NL95_YEA | | 40S ribosomal protein S1 OS=Saccharomyces cerevisiae (strain FostersO) (Baker's yeast) GN=RPS | 273 |
| sp\|A6ZM02\|RS3A2_YE | | 40S ribosomal protein S1-B OS=Saccharomyces cerevisiae (strain YJM789) (Baker's yeast) GN=RP | 273 |
| sp\|A7A1W1\|RS3A1_YE | | 40S ribosomal protein S1-A OS=Saccharomyces cerevisiae (strain YJM789) (Baker's yeast) GN=RP | 255 |
| sp\|P0CX37\|RS6A_YEA | | 40S ribosomal protein S6-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 166 |
| sp\|P05750\|RS3_YEAST | | 40S ribosomal protein S3 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's ye | 158 |
| sp\|P0CX35\|RS4A_YEA | | 40S ribosomal protein S4-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 157 |
| sp\|P07280\|RS19A_YEA | | 40S ribosomal protein S19-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 148 |
| sp\|P0CX39\|RS8A_YEA | | 40S ribosomal protein S8-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 144 |
| sp\|P05755\|RS9B_YEAS | | 40S ribosomal protein S9-B OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 137 |
| sp\|P0CX31\|RS24A_YE | | 40S ribosomal protein S24-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 123 |
| sp\|P0CX55\|RS18A_YE | | 40S ribosomal protein S18-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 112 |
| sp\|O13516\|RS9A_YEA | | 40S ribosomal protein S9-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 101 |
| sp\|P06367\|RS14A_YEA | | 40S ribosomal protein S14-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 93 |
| sp\|P39516\|RS14B_YEA | | 40S ribosomal protein S14-B OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 91 |
| sp\|P02407\|RS17A_YEA | | 40S ribosomal protein S17-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 91 |
| sp\|P26783\|RS5_YEAST | | 40S ribosomal protein S5 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's ye | 88 |
| sp\|P38701\|RS20_YEAS | | 40S ribosomal protein S20 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's y | 80 |
| sp\|P0CX51\|RS16A_YE | | 40S ribosomal protein S16-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 79 |
| sp\|P26786\|RS7A_YEAS | | 40S ribosomal protein S7-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 76 |
| sp\|P0C0T4\|RS25B_YE | | 40S ribosomal protein S25-B OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 74 |
| sp\|P05759\|RS27A_YEA | | Ubiquitin-40S ribosomal protein S31 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) ( | 41 |
| sp\|P0CX47\|RS11A_YE | | 40S ribosomal protein S11-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 60 |
| sp\|Q3E754\|RS21B_YE | | 40S ribosomal protein S21-B OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 61 |
| sp\|P0C0V8\|RS21A_YE | | 40S ribosomal protein S21-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 54 |
| sp\|P41057\|RS29A_YEA | | 40S ribosomal protein S29-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 66 |
| sp\|P25443\|RS2_YEAST | | 40S ribosomal protein S2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's ye | 65 |
| sp\|P05756\|RS13_YEAS | | 40S ribosomal protein S13 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's y | 62 |
| sp\|A7A0V3\|RSSA2_YE | | 40S ribosomal protein S0-B OS=Saccharomyces cerevisiae (strain YJM789) (Baker's yeast) GN=RP | 55 |
| sp\|P48164\|RS7B_YEAS | | 40S ribosomal protein S7-B OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 52 |
| sp\|P48589\|RS12_YEAS | | 40S ribosomal protein S12 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's y | 58 |
| sp\|Q01855\|RS15_YEAS | | 40S ribosomal protein S15 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's y | 37 |
| sp\|P41058\|RS29B_YEA | | 40S ribosomal protein S29-B OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 52 |
| sp\|Q08745\|RS10A_YEA | | 40S ribosomal protein S10-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 53 |
| sp\|P0C0W1\|RS22A_YE | | 40S ribosomal protein S22-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 67 |
| sp\|Q3E7Y3\|RS22B_YE | | 40S ribosomal protein S22-B OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 67 |
| sp\|P35997\|RS27_YEAS | | 40S ribosomal protein S27-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 25 |
| sp\|Q3E7X9\|RS28A_YE | | 40S ribosomal protein S28-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 27 |
| sp\|P0CX29\|RS23A_YE | | 40S ribosomal protein S23-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 25 |
| sp\|P0C0X0\|RS28B_YE | | 40S ribosomal protein S28-B OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 23 |
| sp\|P39938\|RS26A_YEA | | 40S ribosomal protein S26-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 12 |
| sp\|P39939\|RS26B_YEA | | 40S ribosomal protein S26-B OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 9 |
| sp\|P0CX33\|RS30A_YE | | 40S ribosomal protein S30-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's | 3 |

**60S ribosomal proteins:**

| accession | | description | SIC Spectral Co P1 |
|---|---|---|---|
| sp\|P05737\|RL7A_YEAST | | 60S ribosomal protein L7-A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's y | 7 |

**Initiation factors:**

| accession | | description | SIC Spectral Co P1 |
|---|---|---|---|
| tr\|A0A0C5MEG5\|A0A0C | eIF3a | Eukaryotic translation initiation factor 3 subunit A OS=Saccharomyces cerevisiae YJM1615 GN=RPG | 232 |
| sp\|P06103\|EIF3B_YEA | eIF3b | Eukaryotic translation initiation factor 3 subunit B OS=Saccharomyces cerevisiae (strain ATCC 2045 | 140 |
| sp\|P32481\|IF2G_YEAS | eIF2 gamma | Eukaryotic translation initiation factor 2 subunit gamma OS=Saccharomyces cerevisiae (strain ATCC | 128 |
| sp\|A6ZMK5\|EIF3I_YEA | eIF3i | Eukaryotic translation initiation factor 3 subunit I OS=Saccharomyces cerevisiae (strain YJM789) (Ba | 101 |
| sp\|Q05775\|EIF3J_YEA | eIF3j | Eukaryotic translation initiation factor 3 subunit J OS=Saccharomyces cerevisiae (strain ATCC 2045 | 100 |
| sp\|A6ZN26\|EIF3C_YEA | eIF3c | Eukaryotic translation initiation factor 3 subunit C OS=Saccharomyces cerevisiae (strain YJM789) (B | 72 |
| sp\|P09064\|IF2B_YEAS | eIF2 beta | Eukaryotic translation initiation factor 2 subunit beta OS=Saccharomyces cerevisiae (strain ATCC 20 | 63 |
| sp\|P20459\|IF2A_YEAS | eIF2 alpha | Eukaryotic translation initiation factor 2 subunit alpha OS=Saccharomyces cerevisiae (strain ATCC 2 | 62 |
| sp\|Q04067\|EIF3G_YEA | eIF3g | Eukaryotic translation initiation factor 3 subunit G OS=Saccharomyces cerevisiae (strain ATCC 2045 | 52 |
| sp\|P39730\|IF2P_YEAS | eIF5B | Eukaryotic translation initiation factor 5B OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288 | 33 |
| sp\|P38431\|IF5_YEAST | eIF5 | Eukaryotic translation initiation factor 5 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c | 26 |
| sp\|P38912\|IF1A_YEAS | eIF1A | Eukaryotic translation initiation factor 1A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288 | 15 |
| sp\|P32911\|SUI1_YEAS | eIF1A | Eukaryotic translation initiation factor eIF-1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S2 | 23 |
| sp\|P32501\|EI2BE_YEA | | Translation initiation factor eIF-2B subunit epsilon OS=Saccharomyces cerevisiae (strain ATCC 2045 | 5 |

C

## Rabbit

| accession | | description | IC |
|---|---|---|---|
| tr\|G1SQV4\|G1SQV4_RABIT | DDX60 | DExD/H-box helicase 60 OS=Oryctolagus cuniculus GN=DDX60 PE=4 SV=2 | 0 |
| tr\|G1SG72\|G1SG72_RABIT | ABCE1 | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=ABCE1 PE=1 SV | 46 |

**40S ribosomal proteins:**

| accession | | description | S Spectral Count IC |
|---|---|---|---|
| tr\|G1T3D9\|G1T3D9_RABIT | | 40S ribosomal protein S8 OS=Oryctolagus cuniculus (Rabbit) PE=3 SV=1 | 154 |
| tr\|G1U7M4\|G1U7M4_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=LOC100337712 P | 46 |
| tr\|G1T1F0\|G1T1F0_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=LOC100337691 P | 83 |
| tr\|G1T168\|G1T168_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) PE=1 SV=2 | 80 |
| tr\|G1SS70\|G1SS70_RABIT | | 40S ribosomal protein S3a OS=Oryctolagus cuniculus (Rabbit) GN=RPS3A PE=1 S | 176 |
| tr\|G1TM55\|G1TM55_RABIT | | 40S ribosomal protein S6 OS=Oryctolagus cuniculus (Rabbit) GN=LOC100339133 | 147 |
| tr\|G1TNM3\|G1TNM3_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) PE=1 SV=2 | 96 |
| tr\|G1SK22\|G1SK22_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=RPS27A PE=1 SV | 64 |
| tr\|G1TA47\|G1TA47_RABIT | | 40S ribosomal protein S4 OS=Oryctolagus cuniculus (Rabbit) PE=3 SV=2 | 128 |
| tr\|G1TFM5\|G1TFM5_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=RPS5 PE=1 SV= | 41 |
| tr\|G1SGX4\|G1SGX4_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=RPS16 PE=1 SV= | 61 |
| tr\|G1U482\|G1U482_RABIT | | 40S ribosomal protein S3a OS=Oryctolagus cuniculus (Rabbit) GN=LOC10035214 | 97 |
| tr\|G1TPG3\|G1TPG3_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=LOC100356144 P | 70 |
| tr\|G1TDB3\|G1TDB3_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=RPS25 PE=1 SV | 43 |
| tr\|G1TUT9\|G1TUT9_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=RPS2 PE=1 SV= | 65 |
| tr\|G1TTD6\|G1TTD6_RABIT | | 40S ribosomal protein S4 OS=Oryctolagus cuniculus (Rabbit) GN=LOC100355379 | 118 |
| tr\|G1TS23\|G1TS23_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) PE=3 SV=1 | 61 |
| tr\|G1SFR8\|G1SFR8_RABIT | | 40S ribosomal protein S12 OS=Oryctolagus cuniculus (Rabbit) GN=RPS12 PE=1 S | 51 |
| tr\|G1TIB4\|G1TIB4_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=RPS28 PE=1 SV | 45 |
| tr\|G1TL29\|G1TL29_RABIT | | 40S ribosomal protein S27 OS=Oryctolagus cuniculus (Rabbit) GN=RPS27L PE=3 | 35 |
| tr\|B7NZS8\|B7NZS8_RABIT | | Ribosomal protein S9 (Predicted) OS=Oryctolagus cuniculus (Rabbit) GN=RPS9 P | 67 |
| tr\|G1U2Z9\|G1U2Z9_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) PE=1 SV=2 | 32 |
| tr\|G1SP51\|G1SP51_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=RPS13 PE=1 SV | 32 |
| tr\|G1SZ47\|G1SZ47_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=RPS23 PE=1 SV | 42 |
| tr\|G1SIZ2\|G1SIZ2_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=LOC100354063 P | 41 |
| tr\|G1TS40\|G1TS40_RABIT | | 40S ribosomal protein S24 OS=Oryctolagus cuniculus (Rabbit) GN=RPS24 PE=1 S | 36 |
| tr\|G1TG89\|G1TG89_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=RPS15A PE=1 S | 68 |
| tr\|G1TZ76\|G1TZ76_RABIT | | 40S ribosomal protein S27 OS=Oryctolagus cuniculus (Rabbit) GN=RPS27 PE=1 S | 21 |
| tr\|G1SWM1\|G1SWM1_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) PE=1 SV=1 | 21 |
| tr\|G1SVB0\|G1SVB0_RABIT | | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=RPS7 PE=1 SV= | 25 |
| tr\|G1TQM9\|G1TQM9_RABIT | | 40S ribosomal protein SA OS=Oryctolagus cuniculus (Rabbit) GN=RPSA PE=3 SV | 27 |
| tr\|G1TM82\|G1TM82_RABIT | | 40S ribosomal protein S21 OS=Oryctolagus cuniculus (Rabbit) GN=RPS21 PE=1 S | 7 |

**Initiation factors:**

| accession | | description | Spectral Count IC |
|---|---|---|---|
| tr\|G1SRA8\|G1SRA8_RABIT | eIF2 gamma | Eukaryotic translation initiation factor 2 subunit 3 OS=Oryctolagus cuniculus (Rabbi | 55 |
| tr\|G1SMZ5\|G1SMZ5_RABIT | eIF3a | Eukaryotic translation initiation factor 3 subunit A OS=Oryctolagus cuniculus (Rabb | 207 |
| tr\|G1T2G4\|G1T2G4_RABIT | eIF2 alpha | Eukaryotic translation initiation factor 2 subunit 1 OS=Oryctolagus cuniculus (Rabbi | 62 |
| tr\|G1SZ03\|G1SZ03_RABIT | eIF3b | Eukaryotic translation initiation factor 3 subunit B OS=Oryctolagus cuniculus (Rabb | 85 |
| sp\|P41035\|IF2B_RABIT | eIF2 beta | Eukaryotic translation initiation factor 2 subunit 2 OS=Oryctolagus cuniculus (Rabbi | 97 |
| tr\|G1U971\|G1U971_RABIT | eIF3c | Eukaryotic translation initiation factor 3 subunit C OS=Oryctolagus cuniculus (Rabb | 106 |
| tr\|G1STF9\|G1STF9_RABIT | eIF3i | Eukaryotic translation initiation factor 3 subunit I OS=Oryctolagus cuniculus (Rabbit | 65 |
| tr\|G1SED9\|G1SED9_RABIT | eIF3l | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=EIF3L PE=1 SV= | 51 |
| tr\|G1SQV5\|G1SQV5_RABIT | eIF3d | Eukaryotic translation initiation factor 3 subunit D OS=Oryctolagus cuniculus (Rabb | 71 |
| tr\|G1SUC8\|G1SUC8_RABIT | eIF3e | Eukaryotic translation initiation factor 3 subunit E OS=Oryctolagus cuniculus (Rabb | 39 |
| tr\|G1SLW8\|G1SLW8_RABIT | eIF3m | Eukaryotic translation initiation factor 3 subunit M OS=Oryctolagus cuniculus (Rabb | 37 |
| tr\|G1ST95\|G1ST95_RABIT | eIF3h | Eukaryotic translation initiation factor 3 subunit H OS=Oryctolagus cuniculus (Rabb | 26 |
| sp\|P41111\|EI2BD_RABIT | | Translation initiation factor eIF-2B subunit delta OS=Oryctolagus cuniculus (Rabbit) | 11 |
| tr\|G1TCY7\|G1TCY7_RABIT | | Translation initiation factor eIF-2B subunit delta OS=Oryctolagus cuniculus (Rabbit) | 11 |
| tr\|G1SNT1\|G1SNT1_RABIT | eIF3j | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=EIF3J PE=3 SV= | 30 |
| tr\|G1TRL5\|G1TRL5_RABIT | eIF5B | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=EIF5B PE=1 SV= | 29 |
| sp\|P0CL18\|EIF2D_RABIT | | Eukaryotic translation initiation factor 2D OS=Oryctolagus cuniculus (Rabbit) GN=E | 13 |
| tr\|G1T3L2\|G1T3L2_RABIT | eIF3k | Eukaryotic translation initiation factor 3 subunit K OS=Oryctolagus cuniculus (Rabb | 13 |
| tr\|G1TGW1\|G1TGW1_RABIT | eIF5 | Uncharacterized protein OS=Oryctolagus cuniculus (Rabbit) GN=EIF5 PE=4 SV=1 | 12 |
| tr\|G1SLC2\|G1SLC2_RABIT | eIF3f | Eukaryotic translation initiation factor 3 subunit F OS=Oryctolagus cuniculus (Rabb | 10 |

**Elongation factors:**

**Figure S 2: Mass spectrometry characterization of native initiation complex purified from different organisms.** The native initiation complexes obtained from *L. tarentolae* (**A**), *Saccharomyces cerevisiae* (**B**) and rabbit reticulocyte lysate (**C**) were analyzed by MS/MS. The figure contains only 40S and initiation factors. The spectral counts were used for relative

quantification. Darker colors indicate the most abundant protein present in the sample. Related to Figure 1. The full list of proteins identified here will be online.

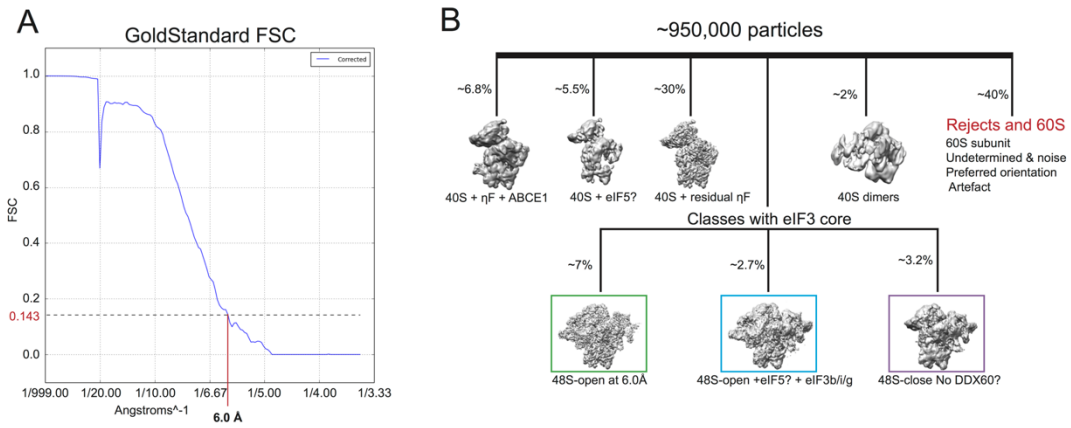


**Figure S 3: Particle sorting and resolution. A,** Gold-standard FSC curve for the three-dimensional reconstruction from class 48S with additional densities at the intersubunit side, and its corresponding resolution of 6.0 Å. **B**, Overview of the particle sorting of the sample purified from initiation complex fractions. The final reconstruction was performed from ~7% of the particles (see Methods), and using the eukaryotic small ribosomal subunit as a reference. Related to figure 2.

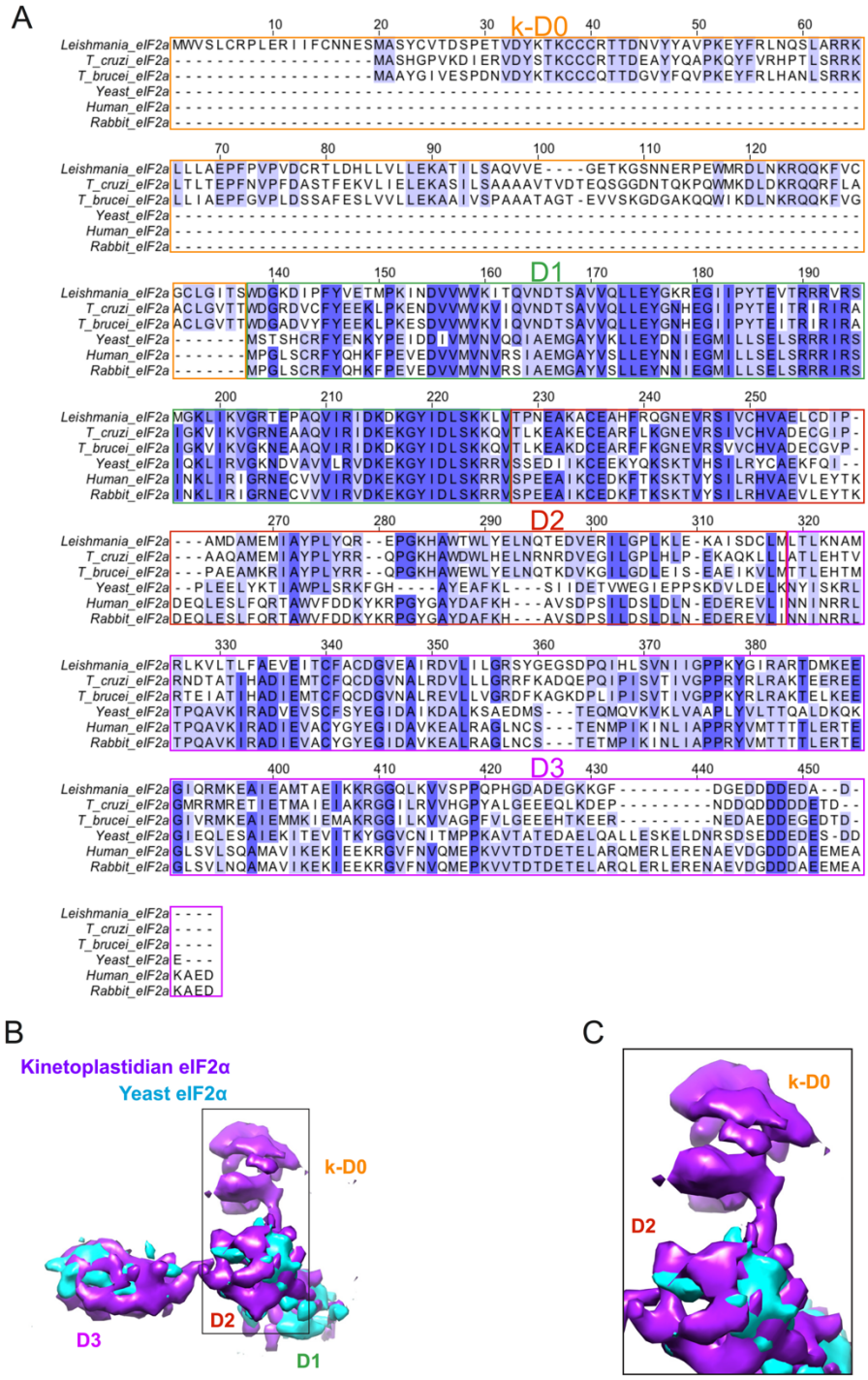


**Figure S 4: A multiple sequence alignment and superposition of eIF2α highlighting the kinetoplastidian-specific additional N-terminal domain. A,** Multiple sequence alignment colored by BLOSUM 62 score of *T. cruzi* eIF2α when compared to its *T. brucei*, *Leishmania major*, *Saccharomyces cerevisiae*, rabbit and human counterparts. **B-C**, Superposition of *T. cruzi* and yeast eIF2α highlighting the kinetoplastidian additional N-terminal domain.

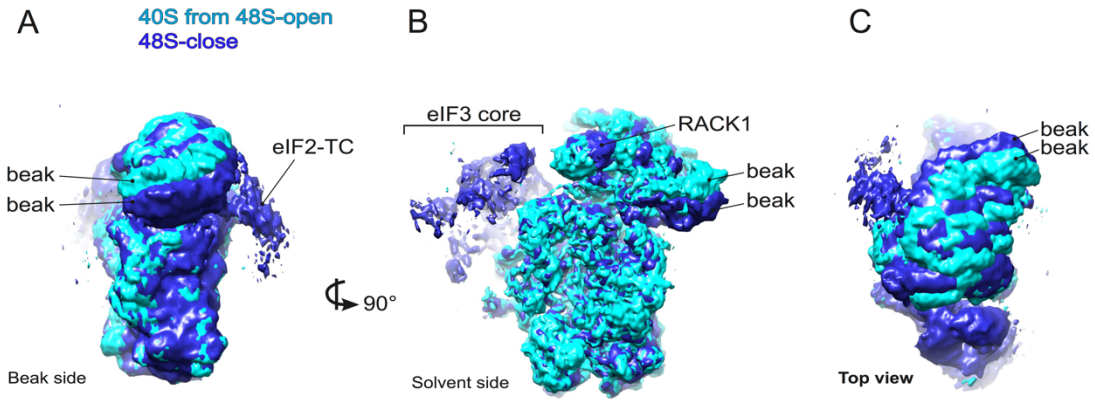


**Figure S 5: Superposition of the 40S from *T. cruzi* 48S-open and the *T. cruzi* 48S-close. A-C,** Highlighting the movement of the head and the lack of the density that may correspond to DDX60.

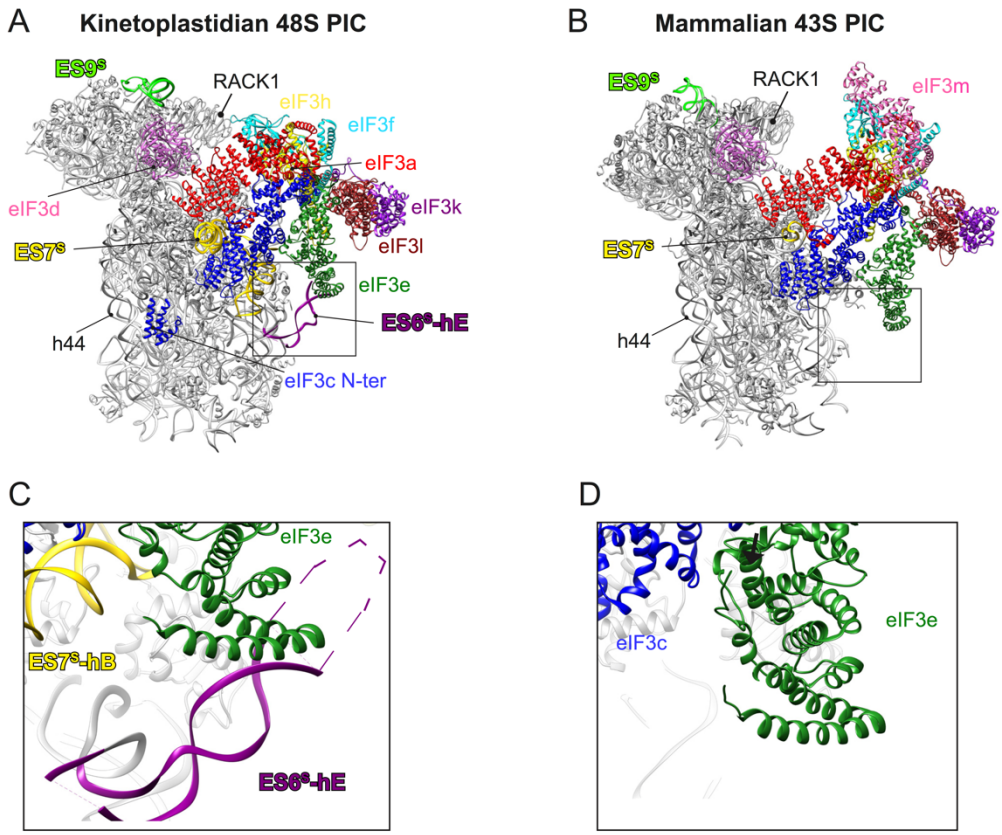


**Figure S 6: Side-by-side comparison of eIF3-40S subunit interactions within kinetoplastidian 48S-open and mammalian 43S PIC. A-B**, Structure of kinetoplastidian 48S-open (left) and mammalian 43S PIC (right) as presented previously (des Georges et al., 2015)**. C,** Close-up view of the contacts between ES7S and eIF3a and eIF3c in kinetoplastids (left). Related to figure 3.

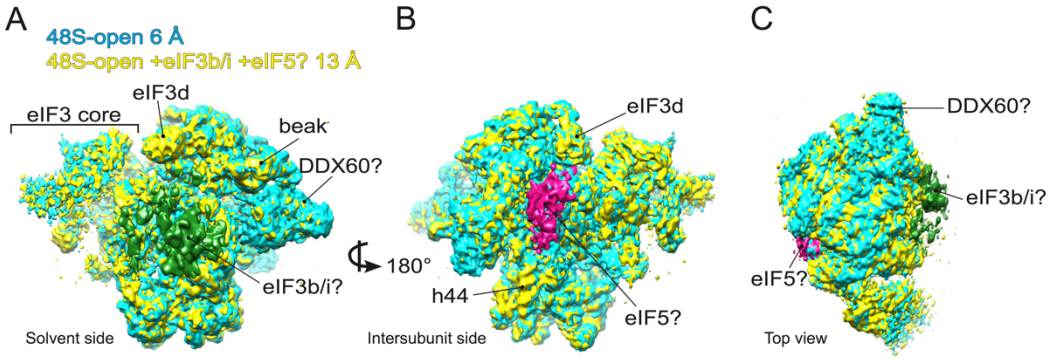


**Figure S 7: Superposition of the two classes of *T. cruzi* 48S-open. A-C,** Highlighting the two additional densities in the low resolution class. The additional density at the solvent side that may correspond to eIF3b is colored. The additional density at the intersubunit side that may correspond to eIF5 is colored in pink.

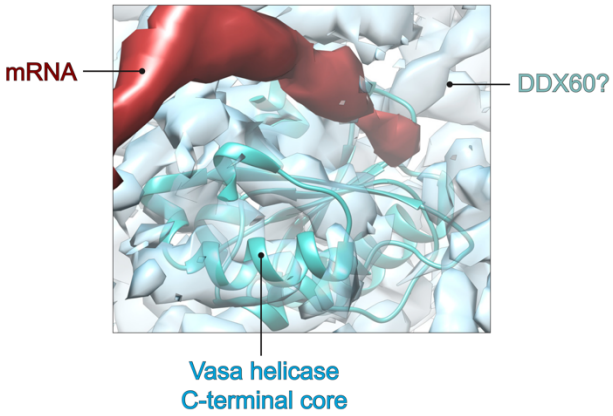


**Figure S 8: Rigid-body fitting of the atomic model of helicase into the uncharacterized density located at the intersubunit side.** Atomic model of the core of Vasa helicase (PDB: 4D25 from Xiol et al., 2014) fitted in the cryo-EM segmented map that may correspond to DDX60.

## CONCLUSION AND PERSPECTIVES

The characterization of native and salt-washed 40S allowed us to identified a new ribosomal protein specific for kinetoplastids, termed here KSRP (Figure 19). KSRP binds at the left foot of 40S where it interacts with ES3$^{S}$ and ES6$^{S}$. Furthermore, KSRP also interacts with kinetoplastid-specific extension of r-protein eS6. The binding site of the two RRM domains of KSRP suggest a role as rRNA chaperone as well as a scaffold for the binding of other regulatory proteins. The discovery of a new r-protein essential for the viability of kinetoplastids, introduced a new potential target for the development of safer and more specific anti-kinetoplastid therapeutic agents that block the mRNA translation in this pathogen.

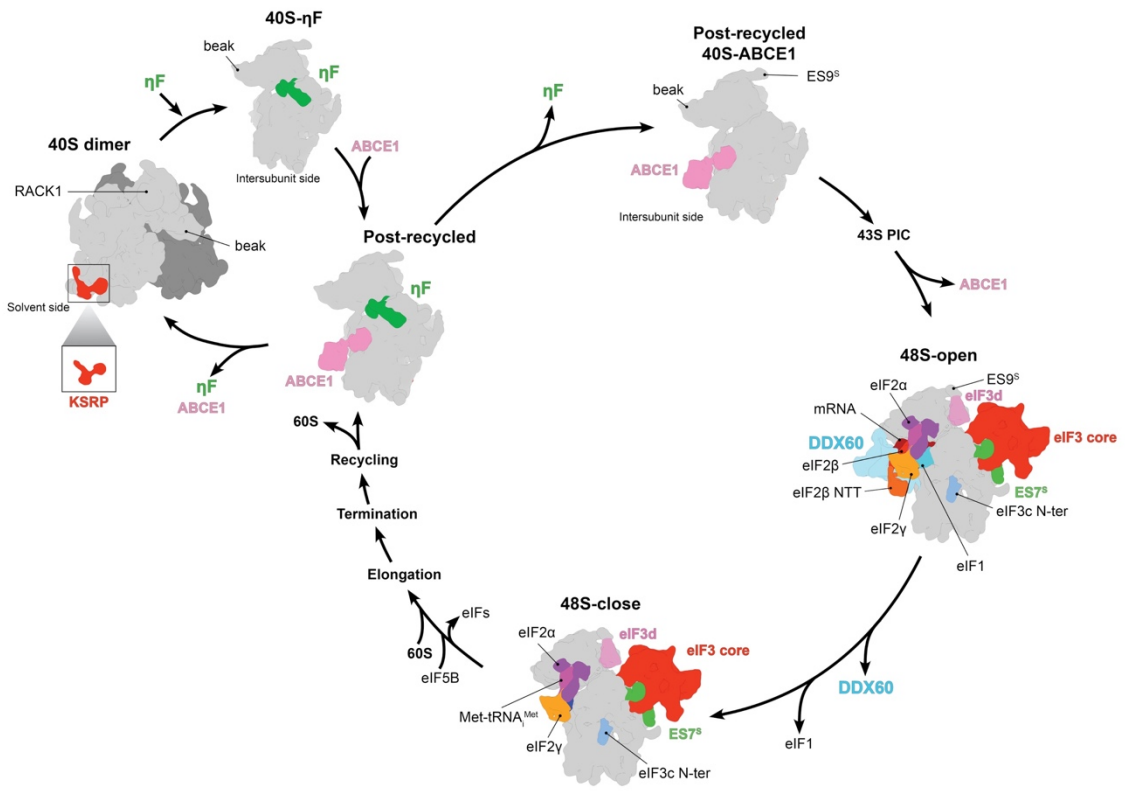


**Figure 19: Schematic model of mRNA translation in kinetoplastids.** The r-protein KSRP binds at left foot of 40S in its solvent side. Furthermore, we propose that the kinetoplastid-specific factor, ηF, prevents the 40S dimerization, which make 40S subunits available for translation. However, ηF must be released, because its binding prevents the recruitment of the mRNA and eIF2-TC. ABCE1 and ηF act as critical ribosomal anti-association factors. After the 43S assembly, kinetoplastidian ABCE1 must be released, because the FeS cluster domain clashes with DDX60 biding site. Finally, DDX60 itself also need to be released because its presence prevents the formation of 48S-close conformation.

Notably, our cryo-EM reconstructions of native and ABCE1-bound 40S complexes from T. cruzi lysates enable us to identify a new factor bound at the platform of 40S close to the E-site, establishing protein/rRNA interactions with h24 and h23. This factor, named here as ηF, also threads into the mRNA channel. The biding site of ηF clashes with eIF2-TC as well as with the mRNA. Therefore, ηF prevents the assembly of 43S PIC. Based on that, we proposed here that ηF is a kinetoplastid-specific translation regulation factor that must be released before the assembly of the initiation complex (Figure 19). Although our structural data have suggested a role for ηF in translation control, further investigation is required to confirm its role and understand the molecular mechanism through which ηF is released from the ribosome. Therefore, biochemical identification of ηF as well as its expression profile during different stages of the parasite life cycle will be mandatory to fully grasp and unravel the different kintoplastid-specific aspects of translational control.

Furthermore, here we have shown that the binding of ABCE1 to the 40S is fully compatible with the assembly of the 43S PIC. Indeed, it is in conformation with our recent publication concerning mammals (Mancera-Martínez et al., 2017), where we presented the structure of ABCE1 in context of mammalian late stage initiation complex. That structure has allowed us to conclude that ABCE1 act as an anti-association factor by preventing the premature joining of 60S (Figure 19). Accordingly, recent study found that the conformation of the ABCE1 FeS cluster domain in context of post-recycling complex prevent the intersubunit bridge formed between uL14 and h44 (Heuer et al., 2017).

Finally, the structure of a near complete 48S pre-initiation complex enables us to identify the binding pattern of most of eIFs (Figure 19), which reveals the architecture of eIF3 core in context of kinetoplastidian initiation complex. The unusual larger ESs in kinetoplastids play an important role in translation initiation either by work as a platform for the accommodation of eIF3 core or by binding and stabilize the eIF3d. Furthermore, our structural and proteomic analysis reveals an unexplored role of DDX60 in translation initiation, likely by binding and unwinding the spliced leader mRNA. The binding site of DDX60 to the 43S PIC is not compatible with the binding of ABCE1. Therefore, unlike in mammals where it was found that ABCE1 is present in the initiation even a late stage of initiation (Mancera-Martínez et al., 2017), in kinetoplastids ABCE1 must be release to allow the accommodation of DDX60 (Figure 19). Although our current structure of 48S IC presents large number of kinetoplastids-specific aspect of translation, an improved resolution combined with biochemical approaches, such us protein crosslinking and ribosome profiling are still required for a better assignment of the reconstruction.

In conclusion, our structural and biochemical analysis highlight the singularity of the translational machinery as well as the translation initiation mechanism (Figure 19) in the human pathogens kinetoplastids.

# 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 |
| GMP-PNP | Sigma | Cat#G0635 |
| **Cell lines** | | |
| *Trypanosoma cruzi* strain Y | This thesis | N/A |
| *Leishmania tarentolae* strain T7-TR | Jena Bioscience | Cat#LT-110 |
| *Saccharomyces cerevisiae strain* strain JD1370 | This thesis | N/A |
| Rabbit Reticulocyte Lysate | Promega | Cat#L4960 |
| **Commercial Assays** | | |
| Amersham ECL Prime Western Blotting Detection Reagent | GE Healthcare | Cat#RPN2232 |
| **Deposited Data** | | |
| Structure of *T. cruzi* 80S | This thesis, deposited at EMdatabank | EMDB- 3844 |
| Structure of *T. cruzi* 40S | This thesis, deposited at EMdatabank | EMDB- 3845 |
| Structure of *T. cruzi* KSRP | This thesis, deposited at PDB | PDB: 5OPT |
| Structure of *L. donovani* KSRP | This thesis, 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 |

## REFERENCE

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 btt429.

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.

Agabian, N. (1990). Trans splicing of nuclear pre-mRNAs. Cell *61*, 1157–1160.

Aitken, C.E., and Lorsch, J.R. (2012). A mechanistic overview of translation initiation in eukaryotes. Nat. Struct. Mol. Biol. *19*, 568–576.

Andersen, D.S., and Leevers, S.J. (2007). The Essential Drosophila ATP-binding Cassette Domain Protein, Pixie, Binds the 40 S Ribosome in an ATP-dependent Manner and Is Required for Translation Initiation. J. Biol. Chem. *282*, 14752–14760.

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.

Anthony, D.D., and Merrick, W.C. (1992). Analysis of 40 S and 80 S complexes with mRNA as measured by sucrose density gradients and primer extension inhibition. J. Biol. Chem. *267*, 1554–1562.

Armache, J.-P., Jarasch, A., Anger, A.M., Villa, E., Becker, T., Bhushan, S., Jossinet, F., Habeck, M., Dindar, G., Franckenberg, S., et al. (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*, 19748–19753.

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

Asano, K., Phan, L., Anderson, J., and Hinnebusch, A.G. (1998). Complex formation by all five homologues of mammalian translation initiation factor 3 subunits from yeast Saccharomyces cerevisiae. J. Biol. Chem. *273*, 18573–18585.

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*, 269–271.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science *289*, 905–920.

Bangs, J.D., Crain, P.F., Hashizume, T., McCloskey, J.A., and Boothroyd, J.C. (1992). Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. J. Biol. Chem. *267*, 9805–9815.

Barthelme, D., Dinkelaker, S., Albers, S.-V., Londei, P., Ermler, U., and Tampé, 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. U. S. A. *108*, 3228–3233.

Basile, L., Jansa, J.M., Carlier, Y., Salamanca, D.D., Angheben, A., Bartoloni, A., Seixas, J., Van Gool, T., Canavate, C., Flores-Chavez, M., et al. (2011). Chagas disease in European countries: the challenge of a surveillance system. Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull. *16*.

Battiste, J.L., Pestova, T.V., Hellen, C.U.T., and Wagner, G. (2000). The eIF1A Solution Structure Reveals a Large RNA-Binding Surface Important for Scanning Function. Mol. Cell *5*, 109–119.

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.

Ben-Shem, A., Garreau de Loubresse, N., 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.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B., et al. (2005). The genome of the African trypanosome Trypanosoma brucei. Science *309*, 416–422.

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*, e1003962.

Bindereif, A. (2012). RNA Metabolism in Trypanosomes (Springer Science & Business Media).

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. *4*, 187–217.

Campbell, D.A., Kubo, K., Clark, C.G., and Boothroyd, J.C. (1987). Precise identification of cleavage sites involved in the unusual processing of trypanosome ribosomal RNA. J. Mol. Biol. *196*, 113–124.

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.

Chen, Y., Potratz, J.P., Tijerina, P., Del Campo, M., Lambowitz, A.M., and Russell, R. (2008). DEAD-box proteins can completely separate an RNA duplex using a single ATP. Proc. Natl. Acad. Sci. U. S. A. *105*, 20203–20208.

Chen, Z., Dong, J., Ishimura, A., Daar, I., Hinnebusch, A.G., and Dean, M. (2006). The essential vertebrate ABCE1 protein interacts with eukaryotic initiation factors. J. Biol. Chem. *281*, 7452–7457.

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.

Choi, S.K., Lee, J.H., Zoll, W.L., Merrick, W.C., and Dever, T.E. (1998). Promotion of met-tRNAiMet binding to ribosomes by yIF2, a bacterial IF2 homolog in yeast. Science *280*, 1757–1760.

Clayton, C. (2016). Gene expression in Kinetoplastids. Curr. Opin. Microbiol. *32*, 46–51.

Clayton, J. (2010). Chagas disease: pushing through the pipeline. Nature *465*, S12–S15.

Clayton, C., and Shapira, M. (2007). Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol. Biochem. Parasitol. *156*, 93–101.

Coura, J.R., and Viñas, P.A. (2010). Chagas disease: a new worldwide challenge. Nature *465*, S6–S7.

Docampo, R., de Souza, W., Miranda, K., Rohloff, P., and Moreno, S.N.J. (2005). Acidocalcisomes ? conserved from bacteria to man. Nat. Rev. Microbiol. *3*, 251–261.

Dong, J., Lai, R., Nielsen, K., Fekete, C.A., Qiu, H., and Hinnebusch, A.G. (2004). The essential ATP-binding cassette protein RLI1 functions in translation by promoting preinitiation complex assembly. J. Biol. Chem. *279*, 42157–42168.

Eliaz, D., Doniger, T., Tkacz, I.D., Biswas, V.K., Gupta, S.K., Kolev, N.G., Unger, R., Ullu, E., Tschudi, C., and Michaeli, S. (2015). Genome-wide analysis of small nucleolar RNAs of Leishmania major reveals a rich repertoire of RNAs involved in modification and processing of rRNA. RNA Biol. *12*, 1222–1255.

El-Sayed, N.M., Myler, P.J., Bartholomeu, D.C., Nilsson, D., Aggarwal, G., Tran, A.-N., Ghedin, E., Worthey, E.A., Delcher, A.L., Blandin, G., et al. (2005a). The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. Science *309*, 409–415.

El-Sayed, N.M., Myler, P.J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E.A., Hertz-Fowler, C., et al. (2005b). Comparative genomics of trypanosomatid parasitic protozoa. Science *309*, 404–409.

Erickson, F.L., and Hannig, E.M. (1996). Ligand interactions with eukaryotic translation initiation factor 2: role of the gamma-subunit. EMBO J. *15*, 6311–6320.

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.

Fendley, G.A., Urbatsch, I.L., Sutton, R.B., Zoghbi, M.E., and Altenberg, G.A. (2016). Nucleotide dependence of the dimerization of ATP binding cassette nucleotide binding domains. Biochem. Biophys. Res. Commun. *480*, 268–272.

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. *advance online publication*.

Franco, J.R., Cecchi, G., Priotto, G., Paone, M., Diarra, A., Grout, L., Mattioli, R.C., and Argaw, D. (2017). Monitoring the elimination of human African trypanosomiasis: Update to 2014. PLoS Negl. Trop. Dis. *11*, e0005585.

Freire, E.R., Dhalia, R., Moura, D.M.N., da Costa Lima, T.D., Lima, R.P., Reis, C.R.S., Hughes, K., Figueiredo, R.C.B.Q., Standart, N., Carrington, M., et al. (2011). The four trypanosomatid eIF4E homologues fall into two separate groups, with distinct features in primary sequence and biological properties. Mol. Biochem. Parasitol. *176*, 25–36.

Freire, E.R., Vashisht, A.A., Malvezzi, A.M., Zuberek, J., Langousis, G., Saada, E.A., Nascimento, J.D.F., Stepinski, J., Darzynkiewicz, E., Hill, K., et al. (2014). eIF4F-like complexes formed by cap-binding homolog TbEIF4E5 with TbEIF4G1 or TbEIF4G2 are implicated in post-transcriptional regulation in Trypanosoma brucei. RNA *20*, 1272–1286.

Gabaldón, T., and Huynen, M.A. Prediction of protein function and pathways in the genome era. Cell. Mol. Life Sci. CMLS *61*, 930–944.

Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141–147.

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.

Gilinger, G., and Bellofatto, V. (2001). Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. Nucleic Acids Res. *29*, 1556–1564.

Gomez, E., and Pavitt, G.D. (2000). Identification of domains and residues within the epsilon subunit of eukaryotic translation initiation factor 2B (eIF2Bepsilon) required for guanine nucleotide exchange reveals a novel activation function promoted by eIF2B complex formation. Mol. Cell. Biol. *20*, 3965–3976.

Gomez, E., Mohammad, S.S., and Pavitt, G.D. (2002). Characterization of the minimal catalytic domain within eIF2B: the guanine-nucleotide exchange factor for translation initiation. EMBO J. *21*, 5292–5301.

Goumans, H., Thomas, A., Verhoeven, A., Voorma, H.O., and Benne, R. (1980). The role of eIF-4C in protein synthesis initiation complex formation. Biochim. Biophys. Acta *608*, 39–46.

Gray, N.K., and Hentze, M.W. (1994). Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs. EMBO J. *13*, 3882–3891.

Günzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L.-C., Chung, H.-M., Lee, P.-T., and Lee, M.G.-S. (2003). RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in Trypanosoma brucei. Eukaryot. Cell *2*, 542–551.

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

Hashem, Y., des Georges, A., Dhote, V., Langlois, R., Liao, H.Y., Grassucci, R.A., Hellen, C.U.T., Pestova, T.V., and Frank, J. (2013b). Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. Cell *153*, 1108–1119.

Henshaw, E.C., Guiney, D.G., and Hirsch, C.A. (1973). The ribosome cycle in mammalian protein synthesis. I. The place of monomeric ribosomes and ribosomal subunits in the cycle. J. Biol. Chem. *248*, 4367–4376.

Hershey, J.W., and Monro, R.E. (1966). A competitive inhibitor of the GTP reaction in protein synthesis. J. Mol. Biol. *18*, 68–76.

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. *advance online publication*.

Hinnebusch, A.G. (2017). Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation. Trends Biochem. Sci.

Holcik, M., and Sonenberg, N. (2005). Translational control in stress and apoptosis. Nat. Rev. Mol. Cell Biol. *6*, 318–327.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33–38, 27–28.

Hussain, T., Llácer, J.L., Fernández, 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.

Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M.-A., Adlem, E., Aert, R., et al. (2005). The genome of the kinetoplastid parasite, Leishmania major. Science *309*, 436–442.

Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. *11*, 113–127.

Jacobson, A. (2005). The end justifies the means. Nat. Struct. Mol. Biol. *12*, 474–475.

Jäger, A.V., De Gaudenzi, J.G., Cassola, A., D'Orso, I., and Frasch, A.C. (2007). mRNA maturation by two-step trans-splicing/polyadenylation processing in trypanosomes. Proc. Natl. Acad. Sci. U. S. A. *104*, 2035–2042.

Jensen, B.C., Sivam, D., Kifer, C.T., Myler, P.J., and Parsons, M. (2009). Widespread variation in transcript abundance within and across developmental stages of Trypanosoma brucei. BMC Genomics *10*, 482.

Johannes, G., Carter, M.S., Eisen, M.B., Brown, P.O., and Sarnow, P. (1999). Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray. Proc. Natl. Acad. Sci. U. S. A. *96*, 13118–13123.

Johnson, P.J., Kooter, J.M., and Borst, P. (1987). Inactivation of transcription by UV irradiation of T. brucei provides evidence for a multicistronic transcription unit including a VSG gene. Cell *51*, 273–281.

Karcher, A., Bu¨ttner, K., Märtens, B., Jansen, R.-P., and Hopfner, K.-P. (2005). X-Ray Structure of RLI, an Essential Twin Cassette ABC ATPase Involved in Ribosome Biogenesis and HIV Capsid Assembly. Structure *13*, 649–659.

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

Kerr, I.D. (2004). Sequence analysis of twin ATP binding cassette proteins involved in translational control, antibiotic resistance, and ribonuclease L inhibition. Biochem. Biophys. Res. Commun. *315*, 166–173.

Khare, S., Nagle, A.S., Biggart, A., Lai, Y.H., Liang, F., Davis, L.C., Barnes, S.W., Mathison, C.J.N., Myburgh, E., Gao, M.-Y., et al. (2016). Proteasome inhibition for treatment of leishmaniasis, Chagas disease and sleeping sickness. Nature *537*, 229–233.

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

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.

Khoshnevis, S., Hauer, F., Milón, P., Stark, H., and Ficner, R. (2012). Novel insights into the architecture and protein interaction network of yeast eIF3. RNA N. Y. N *18*, 2306–2319.

Kimball, S.R. (1999). Eukaryotic initiation factor eIF2. Int. J. Biochem. Cell Biol. *31*, 25–29.

Kiosze-Becker, K., Ori, A., Gerovac, M., Heuer, A., Nürenberg-Goloub, E., Rashid, U.J., Becker, T., Beckmann, R., Beck, M., and Tampé, R. (2016). Structure of the ribosome post-recycling complex probed by chemical cross-linking and mass spectrometry. Nat. Commun. *7*, 13248.

Kispal, G., Sipos, K., Lange, H., Fekete, Z., Bedekovics, T., Janáky, T., Bassler, J., Aguilar Netz, D.J., Balk, J., Rotte, C., et al. (2005). Biogenesis of cytosolic ribosomes requires the essential iron–sulphur protein Rli1p and mitochondria. EMBO J. *24*, 589–598.

Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., and Ban, N. (2012). Atomic structures of the eukaryotic ribosome. Trends Biochem. Sci. *37*, 189–198.

Kolupaeva, V.G., Unbehaun, A., Lomakin, I.B., Hellen, C.U.T., and Pestova, T.V. (2005). Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association. RNA *11*, 470–486.

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

Kozak, M. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. J. Biol. Chem. *266*, 19867–19870.

Kramer, S., and Carrington, M. (2011). Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids. Trends Parasitol. *27*, 23–30.

Krishnamoorthy, T., Pavitt, G.D., Zhang, F., Dever, T.E., and Hinnebusch, A.G. (2001). Tight Binding of the Phosphorylated α Subunit of Initiation Factor 2 (eIF2α) to the Regulatory Subunits of Guanine Nucleotide Exchange Factor eIF2B Is Required for Inhibition of Translation Initiation. Mol. Cell. Biol. *21*, 5018–5030.

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

Laurent, M., and Steinert, M. (1970). Electron microscopy of kinetoplastic DNA from Trypanosoma mega. Proc. Natl. Acad. Sci. U. S. A. *66*, 419–424.

Leidig, C., Thoms, M., Holdermann, I., Bradatsch, B., Berninghausen, O., Bange, G., Sinning, I., Hurt, E., and Beckmann, R. (2014). 60S ribosome biogenesis requires rotation of the 5S ribonucleoprotein particle. Nat. Commun. *5*, ncomms4491.

Li, K., Zhou, S., Guo, Q., Chen, X., Lai, D.-H., Lun, Z.-R., and Guo, X. (2017). The eIF3 complex of Trypanosoma brucei: composition conservation does not imply the conservation of structural assembly and subunits function. RNA N. Y. N *23*, 333–345.

Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. Nat. Rev. Mol. Cell Biol. *12*, 505–516.

Linder, P., Lasko, P.F., Ashburner, M., Leroy, P., Nielsen, P.J., Nishi, K., Schnier, J., and Slonimski, P.P. (1989). Birth of the D-E-A-D box. Nature *337*, 121–122.

Liu, F., Putnam, A., and Jankowsky, E. (2008). ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. Proc. Natl. Acad. Sci. U. S. A. *105*, 20209–20214.

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. U. S. A. *113*, 12174–12179.

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

Lloyd, M.A., Osborne, J.C., Safer, B., Powell, G.M., and Merrick, W.C. (1980). Characteristics of eukaryotic initiation factor 2 and its subunits. J. Biol. Chem. *255*, 1189–1193.

Lomakin, I.B., Kolupaeva, V.G., Marintchev, A., Wagner, G., and Pestova, T.V. (2003). Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. Genes Dev. *17*, 2786–2797.

Lukes, J., Mauricio, I.L., Schönian, G., Dujardin, J.-C., Soteriadou, K., Dedet, J.-P., Kuhls, K., Tintaya, K.W.Q., Jirků, M., Chocholová, E., et al. (2007). Evolutionary and geographical history of the Leishmania donovani complex with a revision of current taxonomy. Proc. Natl. Acad. Sci. U. S. A. *104*, 9375–9380.

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.

Lundgren, D.H., Hwang, S.-I., Wu, L., and Han, D.K. (2010). Role of spectral counting in quantitative proteomics. Expert Rev. Proteomics *7*, 39–53.

Maag, D., Fekete, C.A., Gryczynski, Z., and Lorsch, J.R. (2005). A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. Mol. Cell *17*, 265–275.

MacGregor, P., Savill, N.J., Hall, D., and Matthews, K.R. (2011). Transmission stages dominate trypanosome within-host dynamics during chronic infections. Cell Host Microbe *9*, 310–318.

MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J. Phys. Chem. B *102*, 3586–3616.

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. 0.

Mandal, R.K. (1984). The Organization and Transcription of Eukaryotic Ribosomal RNA Genes. In Progress in Nucleic Acid Research and Molecular Biology, W.E. Cohn, and K. Moldave, eds. (Academic Press), pp. 115–160.

Marintchev, A., Kolupaeva, V.G., Pestova, T.V., and Wagner, G. (2003). Mapping the binding interface between human eukaryotic initiation factors 1A and 5B: a new interaction between old partners. Proc. Natl. Acad. Sci. U. S. A. *100*, 1535–1540.

Martínez-Calvillo, Vizuet-de-Rueda, J.C., Florencio-Mart&#237, Nez, L.E., Manning-Cela, R.G., and Figueroa-Angulo, E.E. (2010). Gene Expression in Trypanosomatid Parasites. BioMed Res. Int. *2010*, e525241.

Martínez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K., and Myler, P.J. (2003). Transcription of Leishmania major Friedlin chromosome 1 initiates in both directions within a single region. Mol. Cell *11*, 1291–1299.

Meleppattu, S., Kamus-Elimeleh, D., Zinoviev, A., Cohen-Mor, S., Orr, I., and Shapira, M. (2015). The eIF3 complex of Leishmania—subunit composition and mode of recruitment to different cap-binding complexes. Nucleic Acids Res. *43*, 6222–6235.

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

Michaeli, S., and Agabian, N. (1990). A Trypanosoma brucei small RNP particle containing the 5S rRNA. Mol. Biochem. Parasitol. *41*, 7–15.

Nakamura, Y., and Ito, K. (2003). Making sense of mimic in translation termination. Trends Biochem. Sci. *28*, 99–105.

Naranda, T., Sirangelo, I., Fabbri, B.J., and Hershey, J.W. (1995). Mutations in the NKXD consensus element indicate that GTP binds to the gamma-subunit of translation initiation factor eIF2. FEBS Lett. *372*, 249–252.

Naveau, M., Lazennec-Schurdevin, C., Panvert, M., Mechulam, Y., and Schmitt, E. (2010). tRNA binding properties of eukaryotic translation initiation factor 2 from Encephalitozoon cuniculi. Biochemistry (Mosc.) *49*, 8680–8688.

Nielsen, P.J., and Trachsel, H. (1988). The mouse protein synthesis initiation factor 4A gene family includes two related functional genes which are differentially expressed. EMBO J. *7*, 2097–2105.

Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000). The structural basis of ribosome activity in peptide bond synthesis. Science *289*, 920–930.

Obayashi, E., Luna, R.E., Nagata, T., Martin-Marcos, P., Hiraishi, H., Singh, C.R., Erzberger, J.P., Zhang, F., Arthanari, H., Morris, J., et al. (2017). Molecular Landscape of the Ribosome Pre-initiation Complex during mRNA Scanning: Structural Role for eIF3c and Its Control by eIF5. Cell Rep. *18*, 2651–2663.

Palade, G.E. (1955). A SMALL PARTICULATE COMPONENT OF THE CYTOPLASM. J. Biophys. Biochem. Cytol. *1*, 59–68.

Parsons, M., Nelson, R.G., Watkins, K.P., and Agabian, N. (1984). Trypanosome mRNAs share a common 5' spliced leader sequence. Cell *38*, 309–316.

Parsyan, A., Shahbazian, D., Martineau, Y., Petroulakis, E., Alain, T., Larsson, O., Mathonnet, G., Tettweiler, G., Hellen, C.U., Pestova, T.V., et al. (2009). The helicase protein DHX29 promotes translation initiation, cell proliferation, and tumorigenesis. Proc. Natl. Acad. Sci. U. S. A. *106*, 22217–22222.

Parsyan, A., Svitkin, Y., Shahbazian, D., Gkogkas, C., Lasko, P., Merrick, W.C., and Sonenberg, N. (2011). mRNA helicases: the tacticians of translational control. Nat. Rev. Mol. Cell Biol. *12*, 235–245.

Paulin, F.E.M., Campbell, L.E., O'Brien, K., Loughlin, J., and Proud, C.G. (2001). Eukaryotic translation initiation factor 5 (eIF5) acts as a classical GTPase-activator protein. Curr. Biol. *11*, 55–59.

Perry, K.L., Watkins, K.P., and Agabian, N. (1987). Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. Proc. Natl. Acad. Sci. U. S. A. *84*, 8190–8194.

Pestova, T.V., and Kolupaeva, V.G. (2002). The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. Genes Dev. *16*, 2906–2922.

Pestova, T.V., Borukhov, S.I., and Hellen, C.U. (1998). Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. Nature *394*, 854–859.

Pestova, T.V., Lomakin, I.B., Lee, J.H., Choi, S.K., Dever, T.E., and Hellen, C.U.T. (2000). The joining of ribosomal subunits in eukaryotes requires eIF5B. Nature *403*, 332–335.

Peterson, D.T., Merrick, W.C., and Safer, B. (1979). Binding and release of radiolabeled eukaryotic initiation factors 2 and 3 during 80 S initiation complex formation. J. Biol. Chem. *254*, 2509–2516.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Phan, L., Zhang, X., Asano, K., Anderson, J., Vornlocher, H.P., Greenberg, J.R., Qin, J., and Hinnebusch, A.G. (1998). Identification of a translation initiation factor 3 (eIF3) core complex, conserved in yeast and mammals, that interacts with eIF5. Mol. Cell. Biol. *18*, 4935–4946.

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.

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. *20*, 624–636.

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.

Pisareva, V.P., and Pisarev, A.V. (2016). DHX29 reduces leaky scanning through an upstream AUG codon regardless of its nucleotide context. Nucleic Acids Res. *44*, 4252–4265.

Pisareva, V.P., Pisarev, A.V., Komar, A.A., Hellen, C.U.T., and Pestova, T.V. (2008). Translation initiation on mammalian mRNAs with structured 5′-UTRs requires DExH-box protein DHX29. Cell *135*, 1237–1250.

Preußer, C., Jaé, N., and Bindereif, A. (2012). mRNA splicing in trypanosomes. Int. J. Med. Microbiol. IJMM *302*, 221–224.

Rassi, A., Rassi, A., and Marin-Neto, J.A. (2010). Chagas disease. The Lancet *375*, 1388–1402.

Rezende, A.M., Assis, L.A., Nunes, E.C., da Costa Lima, T.D., Marchini, F.K., Freire, E.R., Reis, C.R., and de Melo Neto, O.P. (2014). The translation initiation complex eIF3 in trypanosomatids and other pathogenic excavates – identification of conserved and divergent features based on orthologue analysis. BMC Genomics *15*, 1175.

Riou, G., and Delain, E. (1969). Electron microscopy of the circular kinetoplastic DNA from Trypanosoma cruzi: occurrence of catenated forms. Proc. Natl. Acad. Sci. U. S. A. *62*, 210–217.

Rogers, G.W., Richter, N.J., and Merrick, W.C. (1999). Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. J. Biol. Chem. *274*, 12236–12244.

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.

Roy, F.L., Salehzada, T., Bisbal, C., Dougherty, J.P., and Peltz, S.W. (2005). A newly discovered function for RNase L in regulating translation termination. Nat. Struct. Mol. Biol. *12*, 505–512.

Rozen, F., Edery, I., Meerovitch, K., Dever, T.E., Merrick, W.C., and Sonenberg, N. (1990). Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. Mol. Cell. Biol. *10*, 1134–1144.

Saini, A.K., Nanda, J.S., Lorsch, J.R., and Hinnebusch, A.G. (2010). Regulatory elements in eIF1A control the fidelity of start codon selection by modulating tRNAiMet binding to the ribosome. Genes Dev. *24*, 97–110.

Saini, A.K., Nanda, J.S., Martin-Marcos, P., Dong, J., Zhang, F., Bhardwaj, M., Lorsch, J.R., and Hinnebusch, A.G. (2014). Eukaryotic translation initiation factor eIF5 promotes the accuracy of start codon recognition by regulating Pi release and conformational transitions of the preinitiation complex. Nucleic Acids Res. *42*, 9623–9640.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 129–143.

Salas-Marco, J., and Bedwell, D.M. (2004). GTP hydrolysis by eRF3 facilitates stop codon decoding during eukaryotic translation termination. Mol. Cell. Biol. *24*, 7769–7778.

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

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., et al. (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. Cell *102*, 615–623.

Schmitt, E., Naveau, M., and Mechulam, Y. (2010). Eukaryotic and archaeal translation initiation factor 2: A heterotrimeric tRNA carrier. FEBS Lett. *584*, 405–412.

Shalev-Benami, M., Zhang, Y., Matzov, D., Halfon, Y., Zackay, A., Rozenberg, H., Zimmerman, E., Bashan, A., Jaffe, C.L., Yonath, A., et al. (2016a). 2.8-Å Cryo-EM Structure of the Large Ribosomal Subunit from the Eukaryotic Parasite Leishmania. Cell Rep. *16*, 288–294.

Shalev-Benami, M., Zhang, Y., Matzov, D., Halfon, Y., Zackay, A., Rozenberg, H., Zimmerman, E., Bashan, A., Jaffe, C.L., Yonath, A., et al. (2016b). 2.8-Å Cryo-EM Structure of the Large Ribosomal Subunit from the Eukaryotic Parasite Leishmania. Cell Rep. *16*, 288–294.

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.

Simpson, L. (1987). The mitochondrial genome of kinetoplastid protozoa: genomic organization, transcription, replication, and evolution. Annu. Rev. Microbiol. *41*, 363–382.

Smith, M.D., Arake-Tacca, L., Nitido, A., Montabana, E., Park, A., and Cate, J.H. (2016). Assembly of eIF3 Mediated by Mutually Dependent Subunit Insertion. Struct. Lond. Engl. 1993 *24*, 886–896.

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.

Spahn, C.M., Beckmann, R., Eswar, N., Penczek, P.A., Sali, A., Blobel, G., and Frank, J. (2001). Structure of the 80S ribosome from Saccharomyces cerevisiae--tRNA-ribosome and subunit-subunit interactions. Cell *107*, 373–386.

Stoneley, M., and Willis, A.E. (2004). Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. Oncogene *23*, 3200–3207.

Strunk, B.S., Loucks, C.R., Su, M., Vashisth, H., Cheng, S., Schilling, J., Brooks, C.L., Karbstein, K., and Skiniotis, G. (2011). Ribosome Assembly Factors Prevent Premature Translation Initiation by 40S Assembly Intermediates. Science *333*, 1449–1453.

Sturm, N.R., and Simpson, L. (1990). Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. Cell *61*, 879–884.

Sun, C., Todorovic, A., Querol-Audí, J., Bai, Y., Villa, N., Snyder, M., Ashchyan, J., Lewis, C.S., Hartland, A., Gradia, S., et al. (2011). Functional reconstitution of human eukaryotic translation initiation factor 3 (eIF3). Proc. Natl. Acad. Sci. U. S. A. *108*, 20473–20478.

Sutton, R.E., and Boothroyd, J.C. (1986). Evidence for Trans splicing in trypanosomes. Cell *47*, 527–535.

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. Struct. Lond. Engl. 1993 *16*, 673–683.

Unbehaun, A., Borukhov, S.I., Hellen, C.U.T., and Pestova, T.V. (2004). Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. Genes Dev. *18*, 3078–3093.

Valásek, L., Phan, L., Schoenfeld, L.W., Valásková, V., and Hinnebusch, A.G. (2001). Related eIF3 subunits TIF32 and HCR1 interact with an RNA recognition motif in PRT1 required for eIF3 integrity and ribosome binding. EMBO J. *20*, 891–904.

Valášek, L., Mathew, A.A., Shin, B.-S., Nielsen, K.H., Szamecz, B., and Hinnebusch, A.G. (2003). The yeast eIF3 subunits TIF32/a, NIP1/c, and eIF5 make critical connections with the 40S ribosome in vivo. Genes Dev. *17*, 786–799.

Valášek, L.S., Zeman, J., Wagner, S., Beznosková, P., Pavlíková, Z., Mohammad, M.P., Hronová, V., Herrmannová, A., Hashem, Y., and Gunišová, S. (2017). Embraced by eIF3: structural and functional insights into the roles of eIF3 across the translation cycle. Nucleic Acids Res.

Wagner, S., Herrmannová, A., Šikrová, D., and Valášek, L.S. (2016). Human eIF3b and eIF3a serve as the nucleation core for the assembly of eIF3 into two interconnected modules: the yeast-like core and the octamer. Nucleic Acids Res. *44*, 10772–10788.

Weisser, M., Voigts-Hoffmann, F., Rabl, J., Leibundgut, M., and Ban, N. (2013). The crystal structure of the eukaryotic 40S ribosomal subunit in complex with eIF1 and eIF1A. Nat. Struct. Mol. Biol. *20*, 1015–1017.

Willyard, C. (2011). Putting sleeping sickness to bed. Nat. Med. *17*, 14–17.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. Nature *407*, 327–339.

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

Yarunin, A., Panse, V.G., Petfalski, E., Dez, C., Tollervey, D., and Hurt, E.C. (2005). Functional link between ribosome formation and biogenesis of iron–sulfur proteins. EMBO J. *24*, 580–588.

Yoffe, Y., ZUBEREK, J., LEWDOROWICZ, M., ZEIRA, Z., KEASAR, C., ORR-DAHAN, I., JANKOWSKA-ANYSZKA, M., STEPINSKI, J., DARZYNKIEWICZ, E., and SHAPIRA, M. (2004). Cap-binding activity of an eIF4E homolog from Leishmania. RNA *10*, 1764–1775.

Yoffe, Y., Léger, M., Zinoviev, A., Zuberek, J., Darzynkiewicz, E., Wagner, G., and Shapira, M. (2009). Evolutionary changes in the Leishmania eIF4F complex involve variations in the eIF4E–eIF4G interactions. Nucleic Acids Res. *37*, 3243–3253.

Yu, Y., Marintchev, A., Kolupaeva, V.G., Unbehaun, A., Veryasova, T., Lai, S.-C., Hong, P., Wagner, G., Hellen, C.U.T., and Pestova, T.V. (2009). Position of eukaryotic translation initiation factor eIF1A on the 40S ribosomal subunit mapped by directed hydroxyl radical probing. Nucleic Acids Res. *37*, 5167–5182.

Zamudio, J.R., Mittra, B., Campbell, D.A., and Sturm, N.R. (2009). Hypermethylated cap 4 maximizes Trypanosoma brucei translation. Mol. Microbiol. *72*, 1100–1110.

Zeiner, G.M., Sturm, N.R., and Campbell, D.A. (2003). The Leishmania tarentolae Spliced Leader Contains Determinants for Association with Polysomes. J. Biol. Chem. *278*, 38269–38275.

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. (2016a). Structures and stabilization of kinetoplastid-specific split rRNAs revealed by comparing leishmanial and human ribosomes. Nat. Commun. *7*, 13223.

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. (2016b). Structures and stabilization of kinetoplastid-specific split rRNAs revealed by comparing leishmanial and human ribosomes. Nat. Commun. *7*, 13223.

Zhao, Z., Fang, L.L., Johnsen, R., and Baillie, D.L. (2004). ATP-binding cassette protein E is involved in gene transcription and translation in Caenorhabditis elegans. Biochem. Biophys. Res. Commun. *323*, 104–111.

Zinoviev, A., and Shapira, M. (2012). Evolutionary Conservation and Diversification of the Translation Initiation Apparatus in Trypanosomatids. Int. J. Genomics *2012*, e813718.

Zinoviev, A., Léger, M., Wagner, G., and Shapira, M. (2011). A novel 4E-interacting protein in Leishmania is involved in stage-specific translation pathways. Nucleic Acids Res. *39*, 8404–8415.

UNIVERSITÉ DE STRASBOURG

**Jailson BRITO QUERIDO**

# Structural study of mRNA translation in kinetoplastids by cryo-electron microscopy

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

## ABSTRACT

Kinetoplastid is a group of flagellated protozoans highly relevant to public health, which threatens more than 400 million people world-wide. They possess unusual large rRNA expansion segments (ES) in the 40S, such as ES6$^{S}$, ES7$^{S}$ and ES9$^{S}$ and their location suggests an involvement in the initiation process. Furthermore, all mature mRNAs possess a conserved 5′ spliced-leader. Here, we purified from *T. cruzi* cell lysates native initiation complexes and native 40S subunits that we then analysed by cryo-EM and mass spectrometry. The structure of native initiation complexes reveals several kinetoplastid-specific aspects of translation, such as an intricate interaction network between eIF3 and ES6$^{S}$ and ES7$^{S}$. Furthermore, it reveals the role of DDX60 in translation initiation in kinetoplastids. The structure of native 40S subunits reveals the existence of an uncharacterized factor (termed ηF) bound at platform of the 40S. The binding site of ηF suggests a role in translational control. Moreover, we reported a novel kinetoplastid-specific ribosomal (r-) protein (KSRP) bound to the 40S subunit. Our work represents the first structural characterization of kinetoplastids-specific aspects of translation initiation.

## RÉSUMÉ

Les kinétoplastides sont un groupe de protozoaires flagellés hautement importants dans le domaine de santé public, et qui menace plus de 400 millions de personnes dans le monde entier. Ils possèdent des segments d'expansion d'ARNr (SE) inhabituellement plus larges dans les sous-unités 40S, tels que SE6$^{S}$, SE7$^{S}$ et SE9$^{S}$ et leur emplacement suggère une implication dans le processus d'initiation. De plus, tous les ARNm matures possèdent une coiffe commune de structure conservée appelée le Spliced-Leader d'épissage. Ici, nous avons purifié à partir de lysats de cellules de *T. cruzi* des complexes d'initiation natifs et des sous-unités de 40S natives que nous avons ensuite analysées par cryo-ME et pas spectrométrie de mass. La structure des complexes d'initiation natifs révèle certains des aspects spécifiques de la traduction aux kinétoplastides, tels qu'un réseau d'interaction complexe entre eIF3 et SE6$^{S}$ et SE7$^{S}$. En outre, notre structure met en évidence le rôle de DDX60 dans l'initiation de la traduction chez les kinétoplastides. La structure d'une sous-unité 40S native révèle l'existence d'un facteur non caractérisé (appelé ηF) lié à la plate-forme des 40S. Le site de liaison de ηF suggère un rôle dans le contrôle de la traduction. De plus, nous avons rapporté la structure d'une nouvelle protéine ribosomale (-r) spécifique des kinétoplastides (KSRP) liée à la sous-unité 40S. Notre travail pose les premières bases structurales des aspects spécifiques de l'initiation de la traduction chez les kinétoplastides.