Biochemical and structural study of MARS complexes from plasmodium

José Refugio Jaramillo Ponce

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

CNRS – IBMC – UPR 9002

THÈSE présentée par :

José Refugio JARAMILLO PONCE

soutenue le : **4 décembre 2020**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline : Sciences du vivant

Spécialité : Aspects Moléculaires et Cellulaires de la Biologie

Etude Biochimique et Structurale des Complexes MARS de *Plasmodium*

THÈSE dirigée par :

Dr. FRUGIER Magali Directeur de recherches CNRS, Université de Strasbourg, IBMC, France

RAPPORTEURS :

Dr Marc MIRANDE Directeur de recherches CNRS, Université Paris-Saclay, I2BC, France

Dr Thibaut CREPIN Directeur de recherches CNRS, Université Grenoble Alpes, IBS, France

---

AUTRES MEMBRES DU JURY :

Pr Hubert BECKER Professeur de l'Université de Strasbourg, IPCB, France

Remerciements

Je tiens à remercier ma directrice de thèse, Magali Frugier, pour tout son soutien, sa grande disponibilité et ses nombreux conseils tout le long de ces quatre années de thèse. J'ai été vraiment chanceux d'avoir une encadrante si passionnée, très impliquée dans les expériences et surtout avec une grande curiosité scientifique. J'ai tellement adoré toutes nos discussions, j'ai tellement aimé mon sujet de thèse et j'ai été tellement heureux de travailler avec toi. Merci pour m'avoir donné l'opportunité de devenir un chercheur. Ce n'est pas le métier le plus facile du monde mais il peut être énormément gratifiant. « Il faut être content de chaque succès car il n'y en a pas tous les jours », « il faut être persévérant et croire à nous » et « il faut oser imaginer et se permettre de rêver », ce sont quelques-uns de tes conseils que je garderai très précieusement dans ma mémoire. Merci pour tout.

Je remercie les membres de mon jury de thèse, le Dr Marc Mirande, le Dr Thibaut Crépin et le Pr Hubert Becker pour avoir accepté d'évaluer ma thèse. Merci pour tous vos conseils.

Je remercie tous les organismes et les institutions qui ont financé d'une façon ou d'autre le travail effectué lors de ces quatre années de thèse : le CONACYT du Mexique, l'UPR 9002 « Architecture et Réactivité de l'ARN » du CNRS et le LabEx NetRNA.

Je remercie très chaleureusement les membres de l'équipe « Biologie des ARNt et pathogénicité ». Anne Théobald-Dietrich, Joëlle Rudinger-Thirion et Caroline Paulus, merci pour tout votre soutien, tout votre aide et tous vos conseils. C'était un grand plaisir de travailler avec vous dans notre cher labo 443. Claude Sauter, Phillipe Bénas et Bernard Lorber, merci pour me faire découvrir le monde des cristaux et des rayons X. Vous savez que c'est quelque chose qui me passionne particulièrement. Merci l'équipe pour toutes les expériences que j'ai pu partager avec vous pendant ces années : toutes les discussions, tous les pots (de thèse, d'anniversaire, de publication et d'autres), les repas de Noël, les voyages et les congrès, les BBQ sur le campus ou chez Claude, les petits pains pour le café du matin. Je garderai un très bon souvenir de vous. Vous allez toutes me manquer.

Je remercie Pascale Romby, la directrice l'UPR 9002, pour être toujours à l'écoute des doctorants et pour nous inciter à développer de plus en plus notre curiosité scientifique. Je voudrais également remercier Danièle Werling, Leandra Hernandez et Anaïs Lamalle pour votre aide lors des questions administratives, en particulier pour les départs en mission.

Je ne peux pas oublier de remercier tous les amis que j'ai fait à Strasbourg.

Marta Cela, Raphaël De Wijn et Cédric Schelcher, des collègues de travail qui sont devenus des très chers amis. Vous m'avez tellement manqué après votre départ mais c'est parce que nous avons partagé tellement des bonnes choses : des discussions, des congrès, des balades comme celle au paradis de Schwarzwald et plein des soirées thématiques (cocktails, jeux de société, danse, raclette, nouvelle année, anniversaires et d'autres bien particulières).

Luc Thomes, mon cher ami et le tout premier que j'ai fait dans cette ville. Je suis tellement content d'avoir eu le courage d'aller à cette fête de Noël. Merci pour ton soutien à tout moment (même la veille de ma soutenance), pour tous nos discussions de la vie et pour toutes les expériences que nous avons partagé. Laura Antoine, ma chère amie, merci pour ton soutien, en particulier pendant les journées obscures de rédaction. Nous avons aussi partagé beaucoup des bons moments ensemble dont je vais toujours m'en souvenir. J'espère que l'on pourra à nouveau s'asseoir tous les trois sur une terrasse pour discuter de la vie et faire des expériences de résistance.

Kévin Rollet, maintenant tu es le grand doctorant, tu fais aussi partie de mes plus beaux souvenirs. Merci pour tout ton soutien, pour tous nos discussions et pour écouter mes dilemmes des cristaux mon cher roomie au Léman. Roberto, tu viens de commencer ta thèse mais tu es devenu si proche. Merci pour ton soutien, ton aide et pour réveiller ma curiosité scientifique avec nos discussions.

Je voudrais remercier toutes mes collègues, qui ont été ou qui sont toujours à l'institut : Alain Lescure, Mélanie Thami-Braye, Mireille Baltzinger, Javier Rol, Eva Renard, Léna Coudray, Juan Carlos Mellado, Lucas Herrgott, Emma Desgranges, Paula Lopez, Ronald, Justine, Léa …

Je voudrais remercier tous mes proches du Mexique. Mon père et ma mère, soyez fiers de moi car vous êtes déjà ma fierté. Ma sœur et frères, mes grands-parents et toute ma famille. Vous m'avez toujours encouragé à suivre mes rêves, même si cela nous a éloigné physiquement. Merci pour tout.

Finalement, je voudrais remercier ma chérie épouse Gabriela Garcia, qui a été avec moi depuis le début de ma thèse. Merci pour ton soutien inconditionnel, merci pour ta patience, en particulier lors de la rédaction de ce manuscrit, et surtout merci pour avoir accepté de partager l'avenir avec moi. Je te dédie très spécialement cette thèse.

TABLE OF CONTENTS

INTRODUCTION.

DISCUSSION

Material & Methods

INTRODUCTION

I. *Plasmodium*, a vampiric parasite

1. A brief overview of the parasite

1.1 Malaria

Malaria is a very ancient disease that has been sickening and killing people for thousand years (Goldsmith, 2011). It was one of the causes of death of Pharaoh Tutankhamun (Hawass et al., 2010), contributed to the fall of the Roman Empire (Sallares, 2002) and even Shakespeare mentioned it in some of his plays (Reiter, 2000). However, the origin of malaria remained unknown until the late 19th century when Alphonse Laveran discovered the parasite *Plasmodium* and Ronald Ross figured out its mode of transmission (Lalchhandama, 2014).

Plasmodium is a genus of protozoan parasites that cause malaria in humans and some animals. These parasites are transmitted through the bites of infected *Anopheles* mosquitoes and feed from hepatocytes and red-blood cells of the host. Five species can infect humans: *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale* and *P. knowlesi*. Of these, *P. falciparum* and *P. vivax* are the most prevalent, and *P. falciparum* is the most dangerous in terms of deaths and complications (WHO, 2018).

The symptoms of malaria appear several days after the infective mosquito bite. These include fever, headache, chills, nausea and muscle pains. If not treated, malaria can progress to a severe illness, often leading to death (Ruíz López del Prado et al., 2014). Three syndromes occurring separately or in combination dominate most malaria deaths: severe anemia, respiratory distress and cerebral malaria (Cowman et al., 2016); the last one is characterized by a coma caused by the presence of infected red-blood cells in the cerebral micro-circulation (Idro et al., 2010).

Malaria infection can be effectively treated if diagnosed promptly. Important medications include quinoline derivatives such as chloroquine (Parhizgar & Tahghighi, 2017), antifolate drugs like sulfadoxine/pyrimethamine (Desai et al., 2018) and the Artemisinin-based Combination Therapy (ACT), which is the best available treatment for *P. falciparum* malaria (WHO, 2018). However, the development of resistance to these drugs represents a major threat in the control of malaria (Cowell & Winzeler, 2019). Modern intervention programs have contributed to reduce remarkably the malaria burden but the disease remains one of the most severe health issues in the world. In 2018, there were 228 million cases and about 405 000 deaths, most of them children under 5 years old (WHO, 2018).

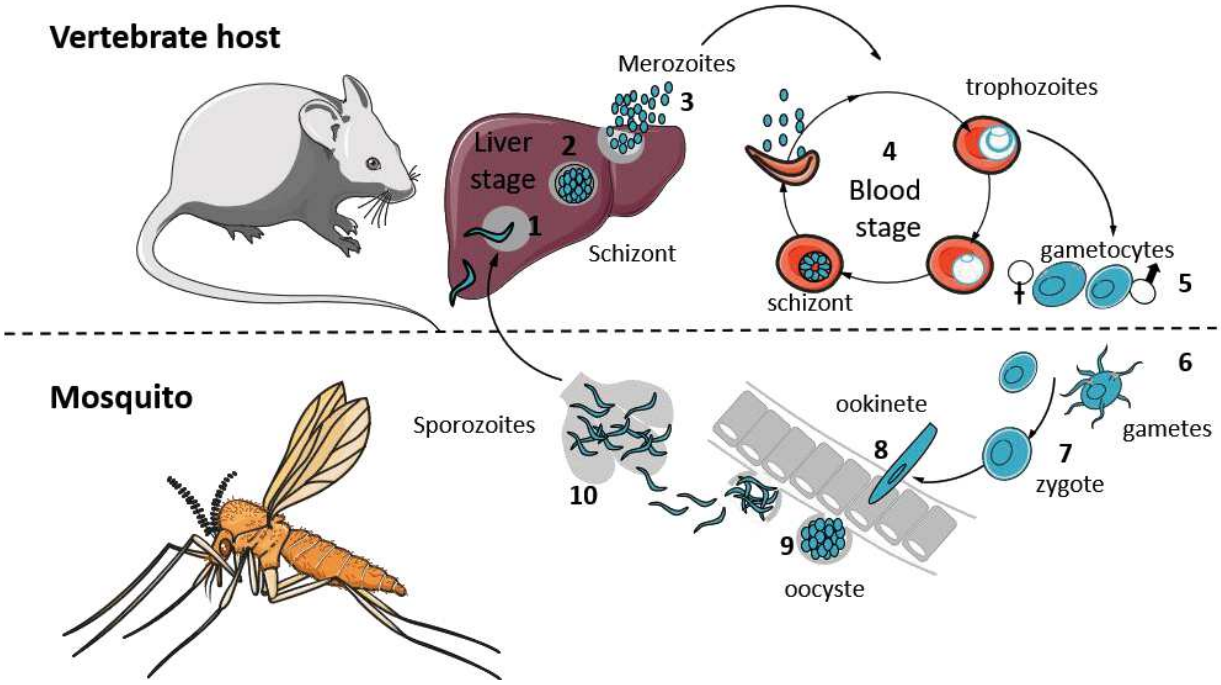


Figure 1: Life cycle of *Plasmodium berghei*. The development of *Plasmodium* is divided between its mosquito vector, the female *Anopheles* (sexual reproduction), and the mouse (asexual reproduction). In the vertebrate host, the life cycle has two main stages: the liver cycle and the erythrocytic cycle. Each stage of development of *Plasmodium* is indicated and explained in the text.

1.2. The life cycle of Plasmodium

During the course of its life, *Plasmodium* transmogrifies into many different forms, which vary in both morphology and physiology. These transformations are necessary for the survival of the parasite, as it must escape from the immune system in two different species: the *Anopheles* mosquito and a vertebrate organism (Shah, 2010). Besides humans, *Plasmodium* can infect birds, reptiles and a diversity of mammals (e.g. apes, rodents, bats). Rodent malaria is of particular interest as it has been widely used as a model to study malaria pathology, host-pathogen interactions and anti-malarial drug efficacy (De Niz & Heussler, 2018). *Plasmodium* species that infect rodents include *P. berghei*, *P. chabaudi*, *P. yoelli* and *P. vinckei*.

A malaria infection begins when the motile and extracellular form of the parasite, named sporozoite (Frischknecht & Matuschewski, 2017), is injected into the host dermis through a mosquito bite (Figure 1). Sporozoites enter the bloodstream and reach the liver. There, they traverse multiple cells before invading a hepatocyte (Vaughan & Kappe, 2017). The interaction of the migrating sporozoite with the highly sulfated heparan sulfate proteoglycans of hepatocytes induces the proteolytic cleavage of the CSP (circumsporozoite protein), the sporozoites' main

surface protein. This activates the sporozoite for productive invasion and formation of a parasitophorous vacuole (Coppi et al., 2007) **(1)**. Within this vacuole, the sporozoites turn into schizonts and begin to replicate asexually **(2)**. Sometimes, the sporozoites differentiate into dormant forms called hypnozoites, which emerge several months or years later (Markus, 2020). Liver infection is a completely asymptomatic stage (Vaughan & Kappe, 2017).

Within 2 to 10 days, hepatic schizonts produce thousands of merozoites (Prudêncio et al., 2006), which are released into the bloodstream **(3)**. Then the blood stage starts **(4)**. Some merozoites escape the immune system and enter rapidly into the erythrocytes. Inside the red blood cell, the parasite resides in a vacuole, digests most of the hemoglobin and retrieves the amino acids necessary for its own protein synthesis. The merozoites become trophozoites, schizonts and between one to four days (depending on the *Plasmodium* species) they replicate into 16 to 32 merozoites, which burst from erythrocytes and infect new ones. These repeated cycles of infection rapidly expand the population of parasites (parasitemia) and lead to a severe anemia and other malaria-related symptoms. Every time the merozoites are exposed in the bloodstream, the host suffers another bout of fever and chills (Shah, 2010; Goldsmith, 2011).

After each cycle of erythrocytic infection, a fraction of the asexual parasites differentiates into male and female gametocytes **(5)**, the sexual forms of *Plasmodium* (Beri et al., 2018). They remain in the bloodstream and are picked up when a female mosquito bites the infected host. Inside the mosquito midgut, the gametocytes become gametes **(6)** and fuse to form a zygote **(7)**, which further develop into ookinete **(8)**. Ookinetes are motile and traverse the epithelium of the midgut. They transform and develop into oocysts **(9)** under the basal lamina surrounding the digestive organ of *Anopheles* mosquitoes. Asexual replication occurs in oocystes and thousands of sporozoites emerge in the hemolymph and travel to the mosquito's salivary glands **(10)**. The infected mosquito injects the sporozoites when feeding on the next vertebrate host and the cycle continues (Aly et al., 2009).

1.3. Parasite control: the vaccine challenge

A malaria vaccine has been long considered as a potential game changer in the fight against malaria (Penny et al., 2020). The ideal vaccine must confer lifelong complete protection with only a few doses (Cowman et al., 2016). Creating a malaria vaccine is not a trivial task, as many difficulties have to be overcome. Indeed, the life cycle of *Plasmodium* is very complex and the parasites are constantly metamorphosing and hiding. If that were not enough, malaria parasites are artists in escaping the host defenses, including a vaccine-induced response (Laurens, 2018).

Antigenic variation, alternative invasion pathways and smoke-screen diversion are just some of their evasion strategies (Casares & Richie, 2009; Rénia et al., 2016). Perhaps, the most important obstacle is the complexity of the parasite itself. Many aspects of the biology of *Plasmodium* remain unknown and those that have been unveiled are highly unusual.

Malaria infections do not imply an effective protective immunity against reinfection. This may be due to the ability of parasites to evade the immune response in an immunized host, for example by exploiting polymorphism or antigenic variations. The best-known example of polymorphism is that of *Pf*EMP1. *Pf*EMP1 is specific to *P. falciparum* and is the main factor that contributes to its virulence (Bernabeu et al., 2016; Gilson, 2017). To escape host antibodies, *P. falciparum* switches between approximately 60 different *var* (variable) genes, using an epigenetic mechanism that guarantees that only one *Pf*EMP1 antigen is expressed by each parasite at any time (Boddey & Cowman, 2013; Smith et al., 2013).

*Pf*EMP1 facilitates the binding of the infected erythrocyte to various ligands on the vascular endothelium (e.g. CD36). The ability to cytoadhere on the vasculature is important for the parasite survival because it prevents the passage of the infected erythrocyte through the spleen and thus prevents its destruction by the macrophages. Unfortunately, infected erythrocytes that stick to the vascular wall can clog the host vasculature and lead to severe malaria (Lee et al., 2019). Other adhesive proteins that show high polymorphism in *P. falciparum* are named RIFINs (*Plasmodium falciparum*-encoded repetitive interspersed families of polypeptides) and STEVOR (subtelomeric variant open reading frame) proteins (Wahlgren et al., 2017). Orthologous of these proteins in other *Plasmodium* species are collectively referred as the *pir* m family (Chan et al., 2014).

Three types of parasite stages have been extensively exploited for vaccine development: (1) pre-erythrocytic stages, (2) blood stages and (3) sexual stages (Frimpong et al., 2018). Pre-erythrocytic vaccines (PEV) target antigens from the sporozoite and liver stages. PEVs currently in development are based on the CSP and whole-cell attenuated sporozoites. Blood stage vaccines (BSV) aim to block the merozoite invasion of host erythrocytes by targeting surface proteins like the apical membrane antigen 1 (AMA1), the merozoite surface protein 1 (MSP1) and the reticulocyte homolog (Rh) proteins. These targets are highly immunogenic but present the inconvenient of being also highly polymorphic (Duffy & Gorres, 2020). Transmission-blocking vaccines (TBV) interrupt parasite transmission to mosquitoes by targeting pre-fertilization and post-fertilization proteins (Duffy & Gorres, 2020).

After decades of development, only the PEV vaccine RTS,S/AS01 has completed phase 3 studies (Greenwood & Doumbo, 2016). The RTS,S vaccine is composed of the repeated "NPNA" region of the *P. falciparum* CSP, fused to the Hepatitis B virus surface antigen, and is formulated with the AS01 adjuvant (Coelho et al., 2017). Vaccination with three doses of RTS,S reduced clinical malaria cases by 28% in young children and only provided a short duration of protection (Cowman et al., 2016). Currently, together with RTS,S/AS01, 20 other candidate vaccines are undergoing clinical trials (Frimpong et al., 2018). The Malaria Vaccine Technology Roadmap, published in 2006 and updated in 2013, aims to develop a vaccine with a protective efficacy of at least 75% by 2030 (Malaria Vaccine Funder Group, 2013).

2. A trafficker of molecules

Parasites of the genus *Plasmodium* have a specific internal organization: they include a unique mitochondrion, an apicoplast and organelles involved in the invasion of host cells called micronemes and rhoptries (Figure 2A). The apicoplast is a non-photosynthetic plastid harboring essential metabolic pathways such as synthesis of type II fatty acids (Shears et al., 2015), synthesis of isoprenoid precursors and part of the heme synthesis pathway (Lim & McFadden, 2010).

2.1. Living in the blood

One of the most striking features of *Plasmodium* is that erythrocytes are their principal host cells. Mature erythrocytes have no nucleus and no protein synthesis machinery (Cooke et al., 2004). They are "floating corpses filled with hemoglobin" (Gratzer, 1984). Erythrocytes can protect the parasite from the host's immune system (because of the lack of a major histocompatibility complex) but provide only limited cellular resources (Belachew, 2018). Nonetheless, the parasite manages to completely remodel the erythrocyte by exporting hundreds of effector proteins that assemble into molecular machineries for trafficking, harvesting of nutrients and evasion from host's defenses (Boddey & Cowman, 2013).

The blood stage merozoite is a polarized cell whose apical extremity contains micronemes and rhoptries, which are organelles implicated in the invasion of erythrocytes (Cowman et al., 2017). (Figure 2A). Invasion is a fast and dynamic process that has been filmed using video-microscopy and is described in Figure 2B (Dvorak et al., 1975; Gilson & Crabb, 2009; Weiss et al., 2015). It comprises two main stages: pre-invasion (steps **1**, **2** and **3**) and internalization (step **4** and **5**).

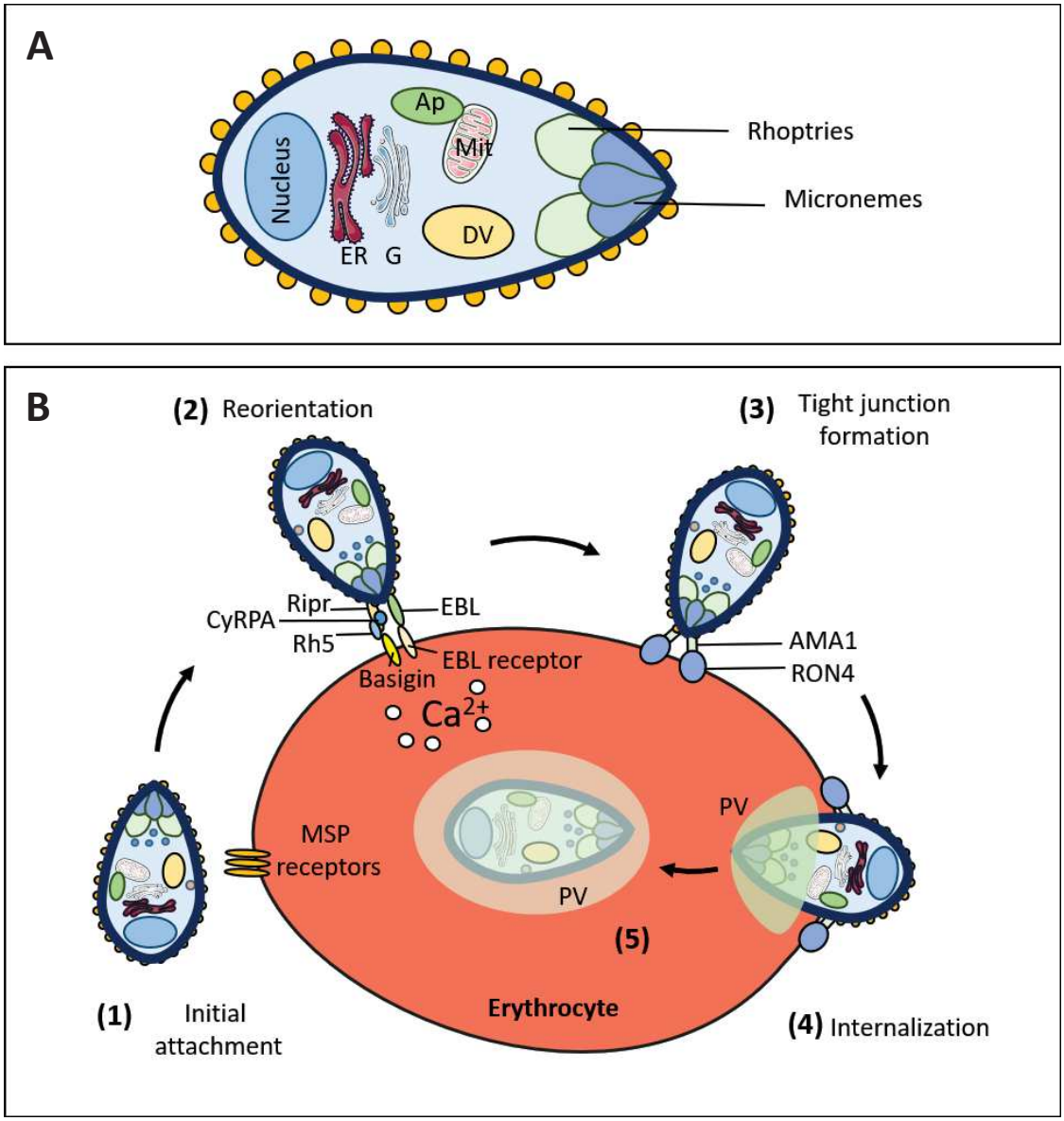


Figure 2: Invasion of erythrocytes. A. Schematic representation of a merozoite. The parasite contains several organelles: a nucleus, an endoplasmic reticulum (ER), a Golgi apparatus (G), an apicoplast (Ap), a unique mitochondrion (Mit), a digestive vacuole (DV) and apical organelles (rhoptries and micronemes) for erythrocytic invasion. **B. Stages of erythrocyte invasion.** Pre-invasion **(1 - 3)** begins when **(1)** a merozoite interacts with the erythrocyte surface. Initial attachment is mediated by interactions between the merozoite surface protein (MSP) and its receptor at the surface of the erythrocyte. **(2)** These interactions produce deformations in the erythrocyte membrane that facilitate reorientation of the merozoite in such a way that its apical end contacts directly the host membrane. New interactions are then established between receptors on the erythrocyte surface and adhesins released from the merozoite micronemes. Two families of adhesins are involved: the erythrocyte binding-like proteins (EBL) and the reticulocyte-binding protein homologs (Rh). Rh5, in complex with Ripr (Rh5 interacting protein) and CyRPA (cysteine-rich protective antigen), binds the receptor basigin on the erythrocyte surface and triggers an influx of calcium into the erythrocyte. **(3)** An irreversible tight junction is then established *via* the AMA1-RON4 (Apical Membrane Antigen 1-Rhoptry Neck protein 4) complex. **(4)** In the internalization step, the actomyosin motor propels the merozoite into the erythrocyte while the contents of rhoptries are released to form the parasitophorous vacuole (PV) (Satchwell, 2016). After 20-60 seconds, the merozoite is completely inside and enclosed by the PV (Weiss et al., 2015). Figure adapted from Cowman et al. (2016) and Cowman et al. (2017).

The blood stage is metabolically very active. It requires many nutrients that neither the parasite nor the host cell synthesize. Consequently, *Plasmodium* establishes a series of transport pathways that allow import of these nutrients through three layers of membranes (Figure 3). Indeed, as the merozoite invades the erythrocyte, it becomes surrounded by a parasitophorous vacuole (PV) and several parasite-derived structures appear within the cytosol of the erythrocyte (Martin, 2020). They are collectively referred as the "exomembrane system" and include the PV, the tubulovesicular network, the Maurer's clefts (MC), several electron-dense vesicles (EDV) and other mobile compartments (Sherling & van Ooij, 2016).

Remodeling of the host erythrocyte is marked by major changes in its permeability (Figure 3). New permeation pathways (NPPs) facilitate the uptake of low-molecular-weight nutrients from the blood plasma, including monosaccharides and other polyols, amino acids and small peptides, nucleosides, some vitamins and inorganic and organic ions (Martin, 2020). Additionally, NPPs provide a mean to remove toxic metabolites such as lactate generated from the parasite's high rate of glycolysis (de Koning-Ward et al., 2016).

Although *Plasmodium* has biosynthetic pathways for asparagine, glutamine, aspartate, glutamate, glycine and proline, most amino acids for protein synthesis are obtained from the digestion of host hemoglobin (Liu et al., 2006). Hemoglobin is taken up from the erythrocyte cytosol by endocytosis, the parasite digests most of it, and uses only a discrete fraction of the released amino acids for its own protein biosynthesis (Lew et al., 2003). Human hemoglobin does not contain isoleucine and the parasite needs to import this amino acid directly from the blood plasma. Indeed, cultures of *P. falciparum in vitro* require supplementation with isoleucine to support their growth (Geary et al., 1985). Some strains even need to be supplemented with methionine, as this amino acid is rare in hemoglobin (Liu et al., 2006). In this case, isoleucine and methionine enter the infected erythrocyte *via* the NPPs (Martin, 2020). However, due to the presence of NPPs the permeability of the infected cell is considerably increased and the parasite must consume more hemoglobin than it needs to preserve its osmotic stability (Lew et al., 2003). Conveniently, the NPPs contribute to excreting the excess of amino acids generated by the digestion of hemoglobin (Dhangadamajhi et al., 2010) and the heme is detoxified *via* its conversion into an inert biocrystal named hemozoin (Kapishnikov et al., 2017). Additionally, the parasite uses non-standard amino acids: selenocysteine is found in at least four proteins (Lobanov et al., 2006), hypusine in the translation initiation factor eIF5A (Kaiser et al., 2007) and formyl-methionine is probably used as the initiator amino acid in apicoplastic translation (Haider et al., 2015).

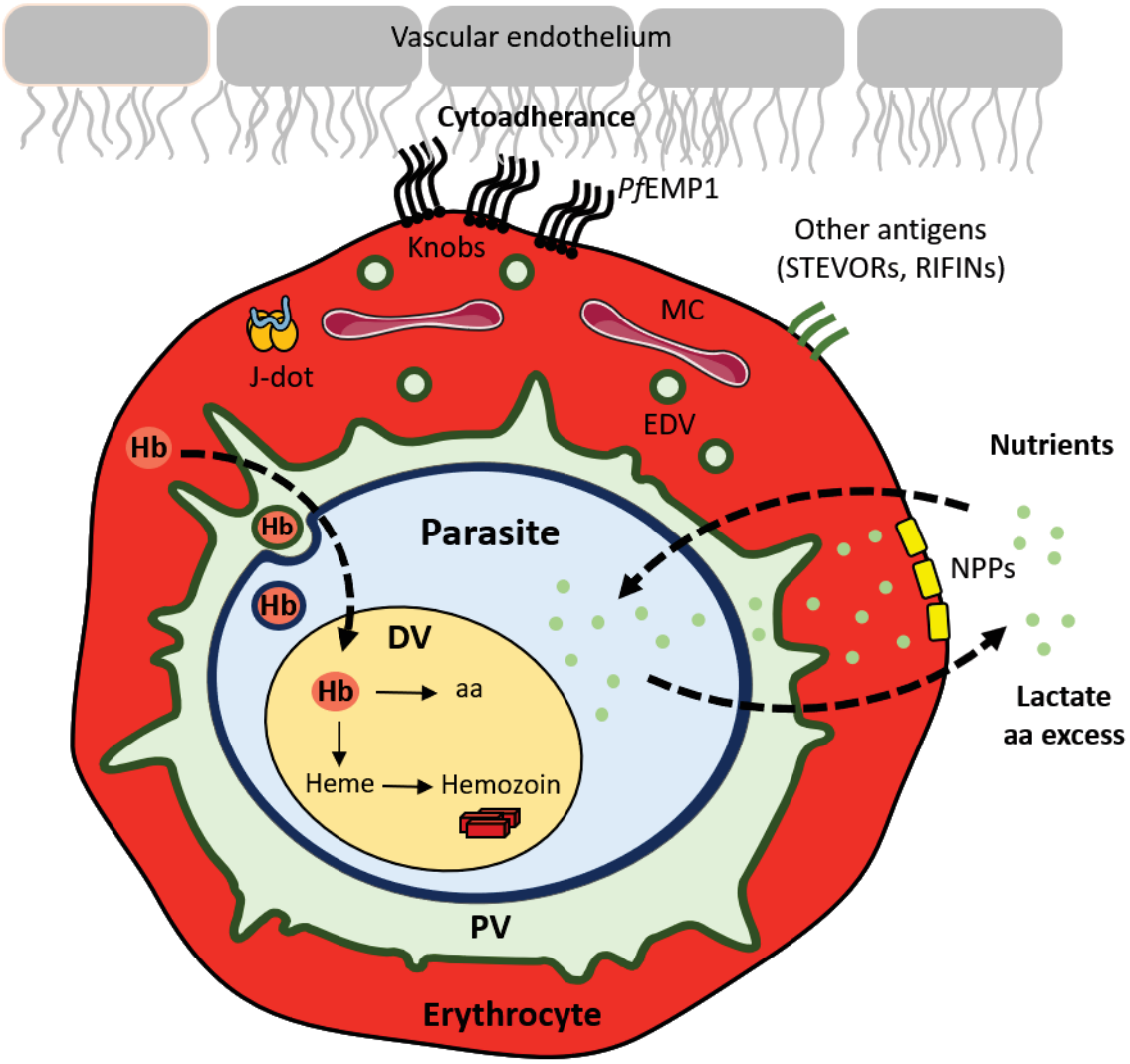


Figure 3: Remodeling of infected red blood cells and molecular exchanges. The erythrocyte (red) is infected by a *Plasmodium* parasite (light blue) that is surrounded by its parasitophorous vacuole (PV) (light green). The PV membrane forms finger-like protrusions into the host cytosol, which are known as the tubulovesicular network. The parasite exports hundreds of proteins into the host cytosol and beyond. Protein trafficking is mediated by electron-dense vesicles (EDVs) and chaperone complexes named J-dots. The Maurer's clefts (MC) are membranous structures originated from the PV membrane that function as sorting depot for proteins destined to the erythrocyte membrane. Exported proteins modify the erythrocyte membrane and membrane skeleton, by forming structures called knobs. These knobs are the platform for the presentation of the surface antigen *Pf*EMP1. Together with STEVORs and RIFIN proteins, *Pf*EMP1 is responsible for the parasite cytoadherence to the vascular endothelium, potentially obstructing blood flow. Erythrocyte remodeling also includes the formation of New Permeation Pathways (NPPs) that allow the uptake of nutrients from the blood plasma. Hemoglobin (Hb) is engulfed from the erythrocyte cytosol and metabolized in the digestive vacuole (DV), where it is hydrolyzed into amino acids (aa) and heme. Detoxification of heme occurs by conversion into the inert crystal hemozoin. Figure adapted from Martin (2020) and de Koning-Ward et al. (2016).

2.2. Trafficking of proteins

Protein trafficking is a highly sophisticated process in *Plasmodium*. The parasite directs proteins to several cellular compartments, including the mitochondrion, the apicoplast, the digestive vacuole and the invasion organelles (Deponte et al., 2012). Additionally, the parasite in the blood stage needs to export a large number of effector proteins to its own plasma membrane, the PV and PV membrane, the erythrocyte's cytosol and membrane and even beyond, into the host's blood plasma. Exported proteins consist of about 550 proteins, representing 10% of the proteome. They are collectively called the "exportome" (Matthews et al. 2019) or "secretome" (Kooij et al., 2006).

2.2.1. The classical pathway

Plasmodium proteins targeted to the endoplasmic reticulum (ER), the parasite membrane, PV, PV membrane and apical organelles contain a conventional hydrophobic signal at the N-terminal end of the nascent polypeptide chain (Lingelbach, 1993; Cooke et al. 2004) and trafficking of these proteins within the parasite involves a classical vesicle-mediated secretory pathway (Figure 4A). The initial step is the translocation of the protein across the ER membrane: ribosome nascent chains containing signal sequences bind to the SRP (signal recognition particle) and are guided to the translocation channel, Sec61/SPC25 at the ER membrane of the parasite (Panchal et al., 2014). Newly synthesized trans-membrane proteins or secreted proteins are then matured by a signal peptidase (SP) that removes the signal sequence and are transferred to their final location *via* the secretory pathway (Marapana et al., 2018).

2.2.2. The PTEX complex

Proteins meant to be secreted in the erythrocyte and beyond need to be discriminated from the others. Such proteins exported by *Plasmodium* are classified into two types. The first group is called PEXEL-containing proteins and comprises ~400 members characterized by the pentameric amino acid motif (RxLxE/Q/D) known as the *Plasmodium* EXport Element (Marti et al., 2004; Hiller et al., 2004). The second group is called PEXEL-negative proteins (PNEPs). PNEPs do not contain any conserved element, making them difficult to identify. They can display an internal transmembrane segment that functions as an input signal to the ER or have a standard N-terminal signal sequence (Spielmann et al., 2006; Heiber et al., 2013). An example of PNEP is the surface antigen family *Pf*EMP1 (de Koning-Ward et al., 2016).

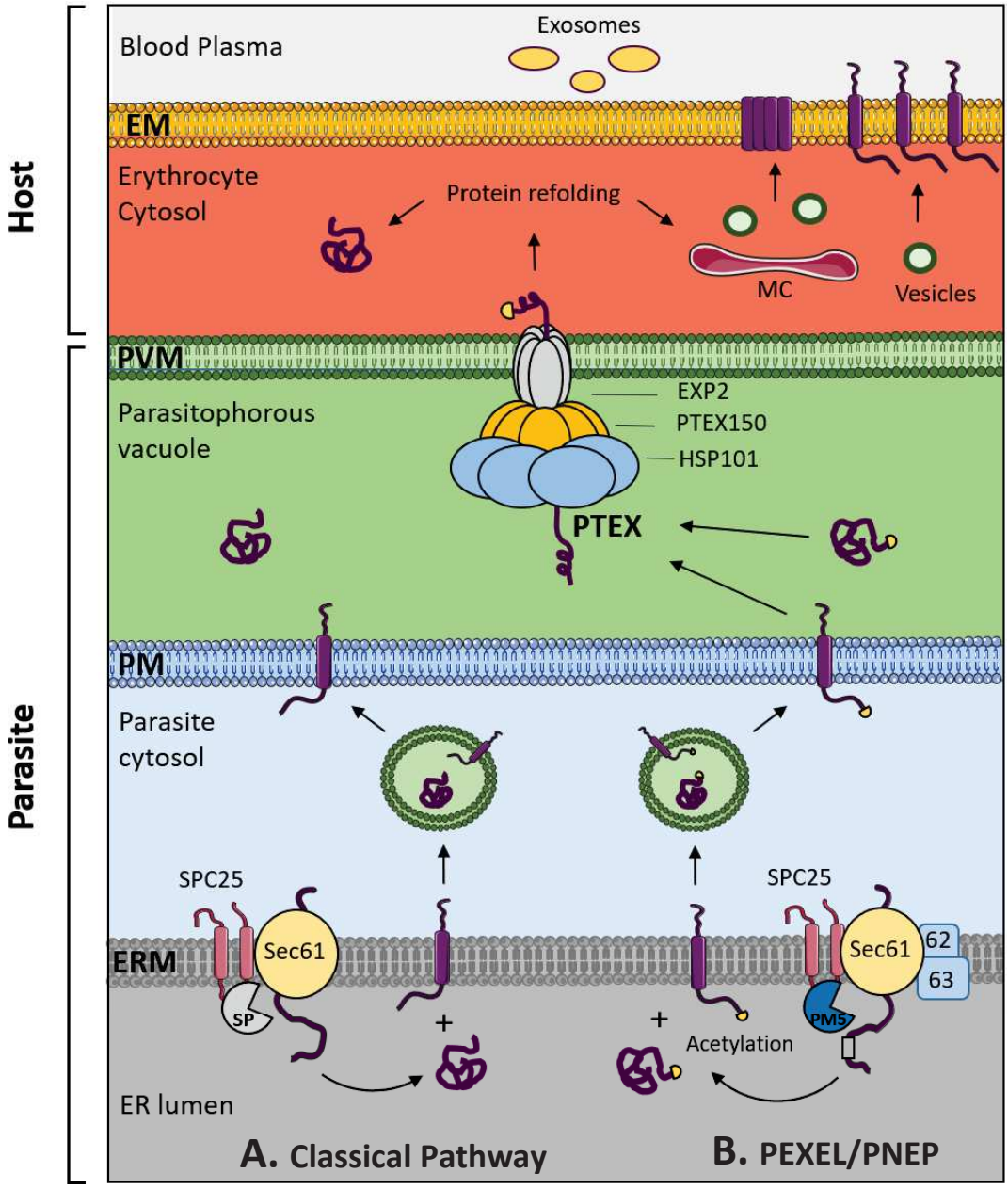


Figure 4. Protein export pathways in *Plasmodium*-infected erythrocytes. Protein export begins at the endoplasmic reticulum (ER), where proteins are translocated across the ER membrane (ERM). **A. The classical pathway.** Proteins targeted to the parasite membrane (PM), the parasitophorous vacuole (PV), the PV membrane (PVM) and some organelles in the parasite cytosol are transported via the classical secretory pathway. These proteins enter the ER *via* a Sec61/SPC25 complex and are matured by a signal peptidase (SP). **B. The PEXEL/PNEP pathway.** Proteins that are targeted to the erythrocyte cytosol, the erythrocyte membrane (EM) and the blood plasma display PEXEL signal or not (PNEP: PEXEL-negative exported proteins). They enter the ER *via* the Sec61/Sec62/Sec63/SPC25 complex, the PEXEL motif is cleaved by the protease plasmepsin V (PM5) and the N-terminal amino acid of the protein is acetylated. Membrane and soluble proteins are loaded into secretory vesicles, which travel to the parasite membrane and release their content into the parasite membrane (PM) or the PV, respectively. Membrane proteins inserted into the parasite membrane require extraction for further trafficking. Both PEXEL and PNEP proteins cross the PV membrane thanks to the *Plasmodium* translocon of exported proteins (PTEX). Proteins that reach the erythrocyte cytosol are refolded by chaperone complexes before being delivered to their final destination. Figure adapted from Matthews et al. (2019).

Unlike proteins that are transported by the classical route, PEXEL and PNEP proteins enter the ER using a separate Sec61/SPC5 import complex (Figure 4B). This specific complex is associated to Sec62, Sec63 and the protease plasmepsin V (PM5) (Marapana et al., 2018). The subunit Sec62 is important to translocate post-translationally proteins that do not display any obvious hydrophobic signal at the N-terminus (Marapana et al., 2018). PM5 cleaves the PEXEL motif and the new mature N-terminus is then acetylated (Boddey et al., 2009; Boddey & Cowman, 2013). Both membrane and soluble proteins are loaded into secretory vesicles and transported to the parasite membrane or the PV, respectively. Further trafficking of membrane proteins requires their extraction by a factor not yet identified. Alternatively, some membrane proteins such as *Pf*EMP1 are transported through the PV as soluble chaperoned complexes (Matthews et al., 2019).

To access the erythrocyte cytosol, PEXEL and PNEP proteins need to cross the PV membrane. A 1.6-MDa protein complex named *Plasmodium* Translocon of EXported proteins (PTEX) (de Koning-Ward et al., 2009) mediates the translocation of N-terminal acetylated proteins in an unfolded state (Gehde et al, 2009; Boddey et al., 2009) (Figure 4). The core of the PTEX complex is composed of three major proteins: the heat shock protein 101 (HSP101), the protein PTEX150 and the export protein 2 (EXP2). HSP101 is a Clp/B ATPase from the AAA+ superfamily (AhYoung et al., 2015) and is responsible for unfolding and translocating proteins (de Koning-Ward et al., 2009). PTEX150 has no homology beyond the *Plasmodium* species. It is tightly bound to HSP101 and has a structural role (de Koning-Ward et al., 2016). EXP2 is thought to include the membrane-spanning component of the PTEX complex, although it does not contain any canonical transmembrane segment (de Koning-Ward et al., 2016). Indeed, EXP2 is a protein localized at the PV membrane (de Koning-Ward et al., 2009), it forms high-order oligomers (Bullen et al., 2012) and leads to the formation of pores (Hakamada et al., 2017). Moreover, independently of HSP101, EXP2 is implicated in the formation of other kinds of channels. These channels are also localized in the PV membrane and they facilitate the passage of nutrients (Garten et al., 2018).

Additionally, the PTEX complex contains two auxiliary components, TRX2 and PTEX88, which are not essential for the survival of the parasite. TRX2 is a thioredoxin-like protein, whose substrate is still unknown and its deletion in *P. berghei* reduced the efficiency of protein export, the growth and the virulence of the parasite (Matthews et al., 2013; Matz et al., 2013). Like PTEX150, PTEX88 is a protein with no obvious homology outside *Plasmodium* species and it interacts with HSP101. However, this interaction is not exclusive since it also interacts with other chaperones present in the PV and with the Exported Protein-Interacting Complex (EPIC) located at the PV membrane. The function of PTEX88 may involve the delivery of cargos initially interacting with EPIC and other chaperones to the PTEX complex (Chisholm et al., 2018).

Ho et al. (2018) determined the cryo-EM structure of the PTEX complex extracted directly from *P. falciparum*-infected erythrocytes (Figure 5). It has an unusual asymmetrical structure where the three components HSP101:PTEX150:EXP2 are found in a ratio of 6:7:7 (Figure 5A). HSP101 is organized as a hexamer and is the motor that unfolds and translocates the proteins through the membrane. PTEX150 is an heptamer which acts as an adapter between HSP101 and EXP2. Only 20% of the structure of PTEX150 could be resolved, suggesting that it is a very flexible and mobile molecule. Finally, EXP2 is the component that anchors the complex to the membrane. The seven monomers of EXP2 oligomerize to form a funnel-shaped structure (Figure 5B and 5C). Indeed, the N-terminus of each molecule consists of amphipathic helices that twist around each other to form a pore across the PV membrane. The inner surface of the channel formed by EXP2 is coated with charged and polar residues, creating an aqueous pore, while the outer surface contains a majority of hydrophobic residues.

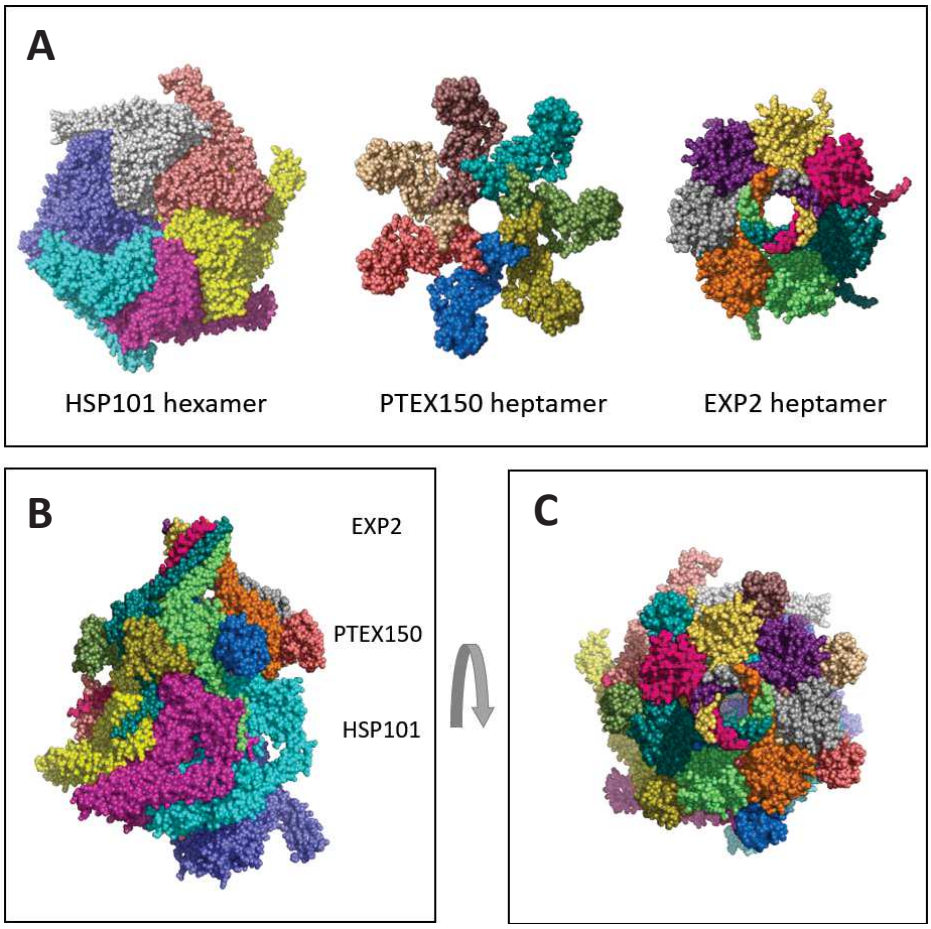


Figure 5. Architecture of the PTEX complex from *P. falciparum*. The cryo-EM structure of PTEX complex purified from infected erythrocytes was solved by Ho et al. (2018). **A. Disassembly of the PTEX structure.** PTEX is composed of three types of subunits, HSP101, PTEX150 and EXP2, which are distributed in different layers. The PTEX150 heptamer connects the HSP101 hexamer to the EXP2 heptamer. The seven N-terminal helices of the EXP2 heptamer twist together to form a pore in the PV membrane. **B. Side view** and **C. Top view of PTEX**. PDB IDs: 6E10 and 6E11.

2.2.3. At the end of the funnel

Once the cargo protein gains access to the erythrocyte cytosol, they are directed to one of the three possible final destinations, the erythrocyte cytoplasm, the Maurer's clefts or the erythrocyte membrane and membrane skeleton (Figure 4). Soluble proteins are refolded immediately by either host chaperones or exported parasite chaperones (Spillman et al., 2017). Maurer's clefts are sorting stations for a number of exported proteins. Some proteins are permanently localized in Maurer's clefts, while proteins destined to the erythrocyte membrane associate transiently with Maurer's clefts. For example, the virulence factor *Pf*EMP1, is transported first by chaperones through the Maurer's clefts and then to the host membrane using chaperone or vesicle mediated mechanisms (de Koning-Ward et al., 2016). In fact, only a few trafficked proteins do not pass through the Maurer's cleft. They correspond to proteins targeted to the erythrocyte membrane and are transported through a vesicle-independent route involving chaperone-associated transport complexes known as J-dots. Once at the membrane, some proteins are packed into exosomes-like vesicles and released into de bloodstream, providing a mechanism for infected erythrocytes to communicate and modify the host immune response (Mantel et al., 2013; Regev-Rudzki et al., 2013).

2.3. Expanding the *Plasmodium* transportome ?

2.3.1. Inventory of the *Plasmodium* exportome

As mentioned previously, the erythrocyte provides very little to the parasite. Thus, in order to survive, the parasite depends on many different transporters, some of which are still poorly characterized. *Plasmodium* possesses a repertoire of transporters facilitating the up-take of nutrients from the host and the excretion of metabolic waste. The full set of transporters, including channels, carriers and pumps, encoded by the genome of an organism is referred as the "transportome" (Martin, 2020), although the term "permeome" has also been used (Martin et al., 2005; Kooij et al., 2006). Given the multitude of cellular compartments observed in blood stage *Plasmodium*, one would expect a large repertoire of transporters. However, the transportome of the parasite represents only 2.5% of its genes, which is significantly less than other organisms (e.g. *A. thaliana* 3.64%, *H. sapiens* 4.32%, *S. cerevisiae* 5.4% and *E. coli* 14.4%).

The *P. falciparum* transportome currently consists of 19 channels, 69 carriers and 29 pumps, a total of 117 transporters (Martin, 2020). Advances in genetic manipulation of *Plasmodium*

allowed the determination of gene disruption phenotypes for every gene in the transportome. In the blood stage of *P. falciparum* and *P. berghei*, approximately 80 transporters are essential (56%) and 10 required for normal growth (7%) (Martin, 2020). This substantial proportion of essential genes might indicate little redundancy in the function of the transporters and could be exploited for drug target (Staines et al., 2017). Additionally, a number of transporters are certainly exposed at the surface of the parasite or of the infected erythrocyte and could be used as potential vaccine targets (Panda & Mahapatra, 2017).

The list of *Plasmodium* transporters might increase as a significant proportion of the genome is still awaiting annotation. Moreover, the study of Martin et al. (2005) leading to the identification of more than 100 transporters in *P. falciparum* only considered transporters with seven or more transmembrane domains as search criteria. However, many transporters might contain less than seven transmembrane domains and other might exhibit non-canonical transmembrane segments difficult to identify. For example, the protein EXP2 of the PTEX complex displays only one non-conventional transmembrane segment that is sufficient to anchor the PETEX complex to the PV membrane (Figure 5).

2.3.2. Looking for more transporters

The characterization of a new transporter requires the determination of its localization within the cell. Yet, each method has advantages and especially disadvantages. The numerous membranes and cellular compartments in *Plasmodium* make difficult the characterization of a specific transporter *in vivo*. Therefore, heterologous expression of the transporter is often used to study its activity without confounding effects due to the presence of other parasite proteins (Staines, 2017). The most successful systems are *Xenopus oocytes* (frog's eggs) and yeast, but other systems such as baculovirus (Kim et al., 2019), bacteria (Razakantoanina et al., 2008) and cell-free systems (Nozawa et al., 2020) have also been used. It is worth to keep in mind that a protein expressed in a heterologous system may differ from the native protein *in vivo*. Factors that may change the activity of the protein include the variation in post-transcriptional modifications, the formation of disulfide bonds, the availability of protein partners and/or chaperones and the composition of the lipid bilayer. The characterization of a transporter in a heterologous system must thus be complemented by assays of the transport process *in vivo* (Martin, 2020).

In vivo, the localization of a transporter is typically done by using antibodies against the native protein that allow investigation of the unaltered protein under endogenous expression. However,

generation of high-quality antibodies is difficult and the possibility of cross-reaction with other parasite and host proteins has led to erroneous results more than once. A confident localization should be reproduced using more than one antibody and validated with other localization methods and/or the measurement of the transport activity by physiological or biochemical assays. A number of transporters have been tagged with short epitopes (e.g. HA) or fluorescent proteins (e.g. GFP or mCherry) to investigate their localization. These tags are introduced by transfection with plasmids or homologous recombination. However, care must be taken as the tag may alter the folding, oligomerization, trafficking or stability of the protein, leading to unreliable localization. In addition to these methods, proteomic analyses of parasite's plasma membrane and organelles provide additional data to validate many transporters (Lamarque et al., 2008; Swearingen et al., 2016; Siau et al., 2016; Swearingen & Lindner, 2018; Nilsson Bark et al., 2018, Boucher et al., 2018). Martin (2020) gives some examples of discrepancies in localization data obtained for *Plasmodium* transporters.

2.4. A unique tRNA import system

Our laboratory has evidenced a new transport system that allows the import of exogenous transfer RNAs (tRNAs) inside the parasite *in vitro* (Bour et al., 2016). This exchange is mediated by a protein named tRip (tRNA import protein) that has been extensively characterized by the team. Stable recombinant *Pf*tRip was successfully expressed in bacteria and purified in the presence of detergent yielding high quality samples. *In vitro*, full-length *Pf*tRip(1-402) and the C-terminal domain *Pf*tRip(214-402) bind specifically tRNAs by recognizing the characteristic 3D structure of these molecules. In solution, *Pf*tRip is a dimer and unpublished data suggest that it has tendency to form higher-order oligomers when tRNA is limiting, suggesting tRNA-dependent pore-forming activity.

Transcriptomic and proteomic data available in PlasmoDB (https://plasmodb.org) indicate that tRip is expressed in all parasite stages. This was confirmed by immunodetecting tRip in multiple stages in both the mosquito and the vertebrate host (Bour et al., 2016). tRip localizes at the surface of the parasite. The subcellular localization of tRip in sporozoites was investigated by immunofluorescence using an affinity purified antibody raised against the C-terminal tRNA binding domain of *Pf*tRip. In native conditions, the localization is unchanged, suggesting that the C-terminal tRNA binding domain is exposed outside the parasite (Figure 6A). Further biochemical experiments such as protease protection assays and detergent-based extractions indicated that tRip is also in an integral membrane protein in the blood stage (Figure 6B).

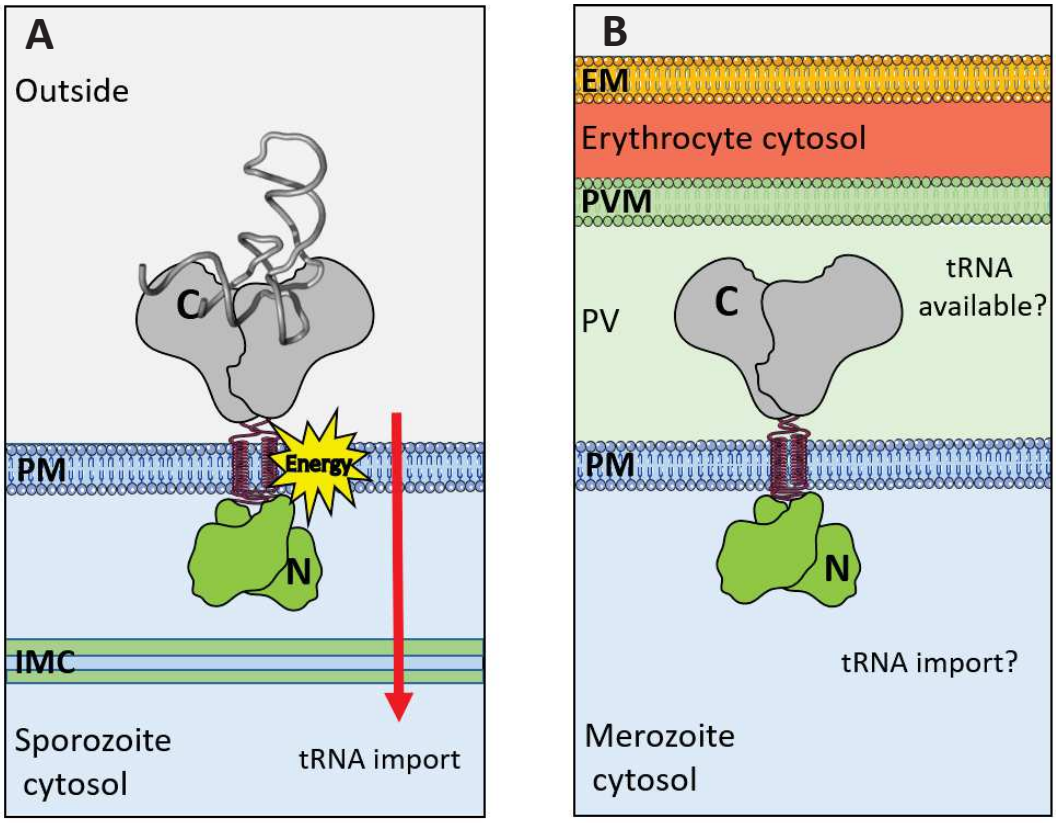


Figure 6: Localization of the tRNA import protein (tRip) in two different parasite stages. tRip is anchored in the parasite plasma membrane, the N-terminal domain (green) being inside the parasite and the C-terminal tRNA binding domain (grey) being outside. *In vitro*, the recombinant tRip forms a dimer (Bour et al., 2016). **A. tRip in sporozoites.** In sporozoites, tRip localizes at the plasma membrane (PM) and mediates the import of full-length exogenous tRNAs by an active process (*in vitro*). An inner membrane complex (IMC) present in invasive forms (e.g. sporozoites, merozoites) separates tRip from the parasite cytosol but does not prevent tRNA from reaching the cytosol. **B. tRip in merozoites.** In blood-stage parasites, tRip is an integral membrane protein in the PM but is separated from the erythrocyte cytosol by the parasitophorous vacuole (PV) and its membrane (PVM). It is not clear if tRNA import occurs within the blood-stage parasites. tRip does not contact directly the host cytosol and erythrocytes are poor in tRNAs.

Import of different exogenous tRNAs was evidenced in sporozoites. The sporozoite stage was chosen to test this activity because it is an extracellular form of the parasite that can be purified directly from mosquito salivary glands and that can be kept alive for about 12 hours *in vitro*. In addition, the sporozoite is a parasite stage (like ookinete) that does not form a PV or forms only a transient PV when it passes through host cells (transmigration of skin cells and liver cells before invasion and multiplication at the liver stage). After few minutes of incubation, exogenous tRNAs were detected by FISH (fluorescence in situ hybridization) inside live parasites. Likewise, exogenous radiolabeled tRNA remained undamaged after RNase treatment only when incubated with alive sporozoites (Bour et al., 2016). These results suggest that sporozoites import full-length tRNAs by an active process.

The implication of tRip in this process was evidenced by two approaches. First, by treating the sporozoites with an antibody against tRip before its incubation with exogenous tRNAs. Antibody binding on endogenous tRip prevented tRNA import. The second approach was to generate a tRip-KO mutant. As expected, the tRip-KO parasite was no longer able to import exogenous tRNA, confirming the tRip-dependency of the process. The absence of tRip was not lethal for the parasite, but slowed down its development in the blood and reduced its protein synthesis. One explanation would be that exogenous tRNAs might support the protein synthesis of the parasite either by participating directly in protein translation or acting as regulatory RNAs. Indeed, *Plasmodium* is the eukaryote with the smaller set of tRNA genes (Gardner et al., 2002) and does not contain an RNA interference pathway (Baum et al., 2009). The fact that tRip is important for the development of the parasite in the blood stage is quite unexpected as mature erythrocytes lack translation machinery, although some *Plasmodium* species (especially *P. berghei*) prefer to invade reticulocytes (Cromer et al., 2006) containing very high levels of tRNAs (Smith & McNamara, 1972; Kabanova et al., 2009). Moreover, tRip might have a function other than tRNA import. Indeed, tRip shares homology to proteins such as Arc1p and AIMP1, which are scaffold proteins participating in the assembly of multi-aminoacyl-tRNA synthetase (MARS) complexes.

Finally, if trafficking of tRNA within cellular compartments has been extensively characterized in other eukaryotic cells (Chatterjee et al., 2018; Hopper & Nostramo, 2019), *Plasmodium* is the first example of a cell importing exogenous tRNAs. The biological relevance of imported tRNAs remains unclear and require further investigation.

3. Translating in Javanais[1]

3.1. Bias in amino acids composition of *Plasmodium* proteins

The nuclear genome of *P. falciparum* is one of the most AT-rich genomes sequenced to date, with an overall (A+T) composition of 80.6% which raises to about 90% in non-coding regions. All *Plasmodium* genomes sequenced to date exhibit the same high AT-content, with the exception of *P. vivax* (59.4%) (Carlton et al., 2008). Rodent malaria species such as *P. berghei* and *P. yoelli* have an AT content above 77% (Hamilton et al., 2017). AT-content in the genome of *Plasmodium* species infecting birds is event higher than in *P. falciparum* (Videvall, 2018). For comparison, the AT content of the genome in *Homo sapiens, S. cerevisiae* and *Toxoplasma gondii* are 58.9%, 61.5% and 47.7%, respectively (Hamilton et al., 2017).

[1] Javanais is a type of French slang where the extra syllable ⟨av⟩ is infixed inside a word after every consonant that is followed by a vowel, in order to render it incomprehensible.

An extreme AT content has certainly consequences on gene expression and both transcription and translation machineries would need special adaptations to deal with these unusual sequences. For instance, the AT-richness of *Plasmodium* increases the probability of finding extended tracts of As and Ts in both inter- and intra-genic regions. Indeed, more than 60% of *P. falciparum* transcripts carry poly A tracks (Djuranovic et al., 2018). In most eukaryotes, poly A tracks act as negative regulators of gene expression, stalling the ribosome, causing frameshifting and activating the mRNA surveillance mechanisms (Arthur et al., 2015; Koutmou et al., 2015; Tournu et al., 2019). However, in *P. falciparum*, proteins are efficiently and accurately translated (Djuranovic et al., 2020). This suggest that issues with poly A tracks in *Plasmodium* are resolved by adaptations in protein synthesis and mRNA quality control systems (Erath et al., 2019).

The high-AT content of *Plasmodium* genome necessarily affects the composition of the proteins. Amino acids encoded by AT-rich codons such as lysine (K), asparagine (N), methionine (M), isoleucine (I), tyrosine (Y) and phenylalanine (F) should occur more frequently in *Plasmodium* proteins. This is the case for *P. falciparum*, where N, K, I and L are the most represented amino acids in the proteome (Bastien et al., 2004). In contrast, proteins from organisms with higher GC content (e.g. *M. tuberculosis*, 65.9% GC) are enriched in amino acids coded by GC-rich codons such as glycine (G), alanine (A) and arginine (R) (Singer & Hickey, 2000). Protein composition is certainly affected by the nucleotide bias, but other factors such as selective constraints, adaptive changes and genetic drift also play important roles in sequence evolution (Singer & Hickey, 2000).

3.2. Insertions everywhere

Identification of *Plasmodium* proteins by sequence homology is always difficult. In the first draft of the genome, more than 60% of genes did not have sufficient homology to be functionally assigned (Gardner et al., 2002). Although advances in homology matching have improved the genome annotation, today about 30% of genes are of “unknown function” (Sexton et al., 2019). This difficulty may be a reflection of the greater evolutionary distance between the parasite and model organisms, intensified by the AT richness of the genome and the presence of numerous insertions in proteins.

Indeed, the number of protein-coding genes in *Plasmodium* is similar to that of *S. cerevisiae*, but the genome of the parasite is considerably larger. This difference is reflected on the size of *Plasmodium* proteins, which can be up to 50% longer than in yeast (Aravind et al., 2003). Multiple sequence alignment reveal that this difference is in part due to the presence of long insertions separating well-conserved blocks adjacent in the homologous proteins (Figure 7A). These

insertions are commonly low-complexity regions (LCRs), which are characterized by a highly recurrent amino acid usage (Pizzi & Frontali, 2001). The length of *Plasmodium* LCRs varies from small insertions (< 10 amino acids) to long insertions (> 100 amino acids) (Aravind et al., 2003). The composition of LCRs is strongly influenced, although not exclusively, by the extreme AT content of the parasite genome. In *P. falciparum*, LCRs are mostly composed of asparagine (N) and lysine (K), which are encoded by AT-rich codons. Other amino acids such as glutamic acid (E) and aspartic acid (D) are also enriched in LCRs, although less frequently (Musto, 1995; DePristo et al., 2006). In the case of *P. vivax*, whose genome is richer in GC, LCRs are instead composed of alanine (A) repeats (Dalby, 2009).

Almost 90% of *P. falciparum* proteins contain at least one LCR, including many highly conserved housekeeping genes (Aravind et al., 2003). They appear in regions separating different protein domains and also inside well-conserved globular folds. LCRs are believed "to encode non-globular domains that are extruded from the protein core and do not impair the functional folding of the protein" (Pizzi & Frontali, 2001). When compared to homologous proteins of known structure, LCRs of *P. falciparum* match the loops between secondary structural elements and are oriented towards external surfaces (Aravind et al., 2003). Zilversmit et al. (2010) identified three families of LCRs in *P. falciparum* (Figure 7B): **(1)** the heterogeneous family characterized by aperiodic regions containing a reduced alphabet of amino acids, **(2)** poly-N stretches of different length and **(3)** the high-GC family composed of long heterogeneous repeats showing numerous insertions and deletions (indels).

Poly-N LCRs are particularly abundant in *P. falciparum* and are present in all protein families from all developmental stages, although they seem to be underrepresented in surface antigens where E repeats are prevalent (Singh et al., 2004). In other organisms, proteins with large N repeats have tendency to form insoluble aggregates, particularly at high temperatures (Halfmann et al., 2011). Given that protein aggregation is often toxic to cells, it is remarkable that the parasite maintains a proteome with so many potentially toxic LCRs, especially when fever is a feature of malaria (Davies et al., 2017). In *Plasmodium*, it has been reported that chaperones are particularly efficient at suppressing the aggregation of these proteins (Muralidharan et al., 2012) and this ability would neutralize the negative selective pressure against the expansion of poly-N LCRs, allowing the propagation of these insertions and further evolution into new domains with novel functions (Muralidharan & Goldberg, 2013). About 10% of *P. falciparum* proteins contain poly-N LCRs and the corresponding subproteome is enriched in regulatory proteins, such as transcription factors and RNA binding domains (Pallarès et al., 2018). It has been proposed that such LCRs might be involved in the recruitment of multiple binding partners. Even though the

unique study that deleted a poly-N LCR present in the protein Rpn6 (an essential proteosome component) did not show any difference in the expression profile, the protein half-life, the cellular localization, the function and the protein-protein interactions (Muralidharan et al., 2011). Interestingly, long N stretches are less abundant in other *Plasmodium* species, despite their richness in AT (Muralidharan & Goldberg, 2013).

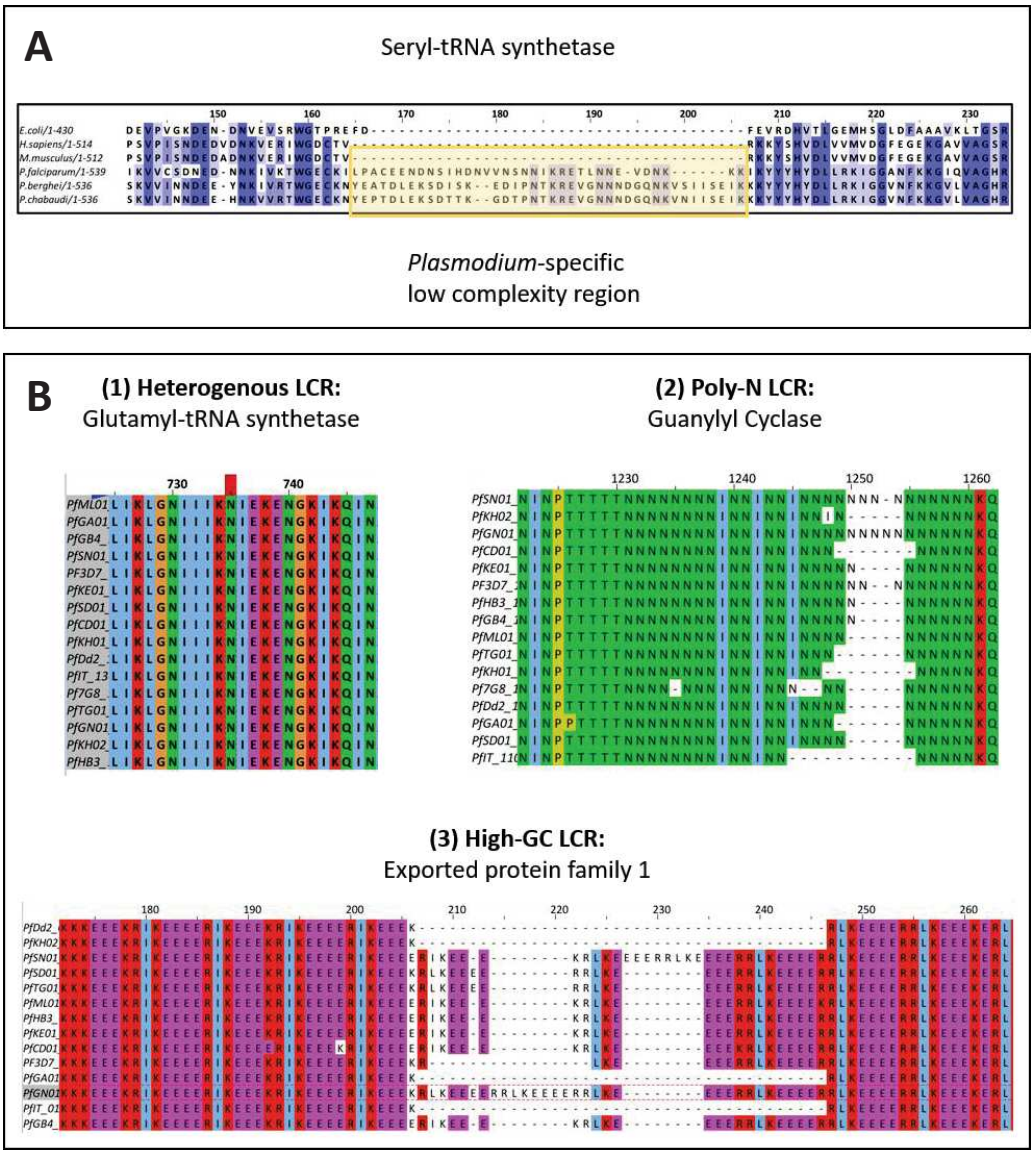


Figure 7. Low complexity regions in *Plasmodium*. A. Highlighting insertions in *Plasmodium* proteins. Sequence alignment of *Plasmodium* seryl-tRNA synthetase with homologous proteins from other species reveal recurrent occurrence of insertions (light yellow) in the three Plasmodium proteins, which are characterized by a biased amino acid composition. Note that LCRs vary in size and sequence between the different *Plasmodium* species. Residues are colored by percentage of identity. **B. Different types of LCRs in *P. falciparum*. (1)** Heterogenous LCRs are non-repetitive sequences containing a reduced alphabet of amino acids. **(2)** Poly-N LCRs are repetitive stretches of asparagine (N) residues. **(3)** High-GC LCRs are periodic repeats characterized by the presence of insertions and deletions (indels). These family is termed High-GC because most recombination breakpoints occur at regions with low AT-content. Sequence alignments are colored according to the Clustal X code. Figure adapted from Zilversmit et al. (2010).

3.3. Origin and function of LCRs

It is predicted that most LCRs in *P. falciparum* are intrinsically disordered, suggesting that they are probably highly dynamic and exist as an ensemble of different conformations. However, they might also adopt a structured conformation upon interaction with an appropriate binding partner. Although poorly investigated, there are few examples where LCR perform a relevant function in protein interactions. For example, it has been shown that two enzymes of the *de novo* pyrimidine synthesis pathway interact *via* a LCR to form a complex that has a catalytic advantage over the individual enzymes (Imprasittichail et al., 2014). Another example is the presence of an heterogenous LCR (29-31 aa) in the cytosolic aspartyl-tRNA synthetase (DRS) from *P. falciparum* that is important for dimerization of the enzyme and thus its catalytic activity (Bour et al., 2009).

Despite their high divergence and potential toxicity, LCRs are maintained across *Plasmodium* species (Aravind et al., 2003). It has been proposed that LCRs confer a selective advantage to the parasite and I will present a selection of theories that try to explain the origin and function of *P. falciparum* LCRs: (1) the rapid adaptation/smoke screen concept, (2) the non-adaptive model (3) the cryptic introns and (4) the tRNA sponges. Additionally, it has been proposed that LCRs may participate in protein-protein interactions, protein-nucleic acid interactions and in dictating subcellular localization (Davies et al., 2017).

3.3.1. LCRs as immune evasion tool

Since many surface antigens such as CSP (sporozoites) and MSP1 (merozoites) contain LCRs that are highly immunogenic, it has been proposed that these regions are a source of structural polymorphism allowing the parasite to escape from the immune system (MacRaild et al. 2016). Additionally, LCRs in surface antigens are numerous and immunologically cross-reactive. They may act as a smoke screen to divert the immune system towards the production of low-affinity antibodies against them at the expense of generating high affinity antibodies against other essential epitopes in the parasite (Kemp et al. 1987; Ridley, 1991; Rich et al., 1997; Hughes, 2004). High-GC LCRs may contribute to this function as they seem to be more susceptible to recombination (presence of numerous indels). Some major antigens such as MSP1 and MSP2 contain High-GC LCRs, but there is no evidence showing that this family of LCRs is preferentially found in surface antigens (Zilversmit et al., 2010). Interestingly, genes encoding surface antigens (e.g. *Pf*EMP1) have tendency to cluster in the subtelomeric regions of chromosomes (Gardner et al., 2002), where the frequency of recombination is higher. This might contribute to their antigenic variations.

3.3.2. LCRs as a neutral evolution feature

The non-adaptative model proposed by DePristo et al. (2006) states that the unusual number and size of LCRs in *P. falciparum* are simply a consequence of its extreme genomic AT content and high rate of recombination. Molecular mechanisms such as replication slippage and unequal crossover recombination might contribute to the continuous expansion of LCRs. This explains well the origin of poly-N LCRs, but not the others. Heterogeneous LCRs are characterized by a slower rate of evolution (lack of indels polymorphism) and they do not expand rapidly. High-GC LCRs do evolve rapidly, but AT content is not the driving force, as most recombination breakpoints occur at low-AT (high GC) regions (Zilversmit et al., 2010). The model of DePristo et al. (2006) would explain why LCRs are so abundant in *Plasmodium* proteins and does not exclude that some LCRs might be beneficial, for example, to surface antigens. However, it is difficult to reconcile the composition and variable lengths of LCRs in different *Plasmodium* species with conventional sequence-dependent functions (DePristo et al., 2006).

3.3.3. LCRs as cryptic introns

Xue & Forsdyke (2003) proposed that LCRs might act as introns to allow optimum folding in RNA. According to them, the amino acid composition in LCRs would be the consequence of forces operating at the nucleic acid level. To support this hypothesis, the authors highlighted that the first and second positions of the codons used in LCRs are indeed AT-rich, but that the third position is often a G or a C. Thus, the first two positions would explain the bias in amino acid composition of LCRs, but the nature of the third position would contribute to balance the GC content of the RNA and promote better folding and stability. Since the third position is usually independent of the encoded amino acid, it should be able to perform nucleic acid level functions If this is the case, Zilversmit et al. (2010) pointed out that LCR coding sequences in *Plasmodium* genes should co-localize with introns in homologous genes from other organisms.

3.3.4. LCRs as tRNA sponges

The rate of translation of a protein depends on the concentration of available aminoacyl-tRNAs (Komar, 2016). Given the high occurrence of repetitive residues in LCRs, particularly N, the parasite requires large amounts of the corresponding N-tRNA^{N} to efficiently translate these regions. In many organisms, the most used codons are decoded by the most abundant tRNAs, whose genes are usually present in multiple copies in the genome (Kanaya et al., 2001; Rocha, 2004). Thus, one would expect the *P. falciparum* genome to contain multiple copies of tRNAs that

recognize N codons and other codons highly used in LCRs. However, there is only one gene copy for each tRNA and therefore the tRNAs that decode amino acids in LCRs are not more abundant than the others. Frugier et al. (2010) proposed that LCRs containing repetitive amino acids (tRNA sponges) reduce the translation rate of the ribosome because the pool of available aminoacyl-tRNA is depleted faster than it is recycled. In this way, the already synthesized N-terminal domains would have more time to fold properly (Figure 8). Indeed, the translation rate is inversely proportional to the folding efficiency of proteins (Yu et al., 2015). LCRs in *Plasmodium* would be intrinsic chaperones replacing the codon preferences and mRNA structures that generally affect translation efficiency of multidomain proteins. This theory is supported by the following observations: (1) the amount of $tRNA^N$ is comparable to other tRNAs in the blood stage of *P. falciparum* and (2) asparaginylation is no more efficient than other aminoacylation systems of the parasite (Filisetti et al., 2013). Alternatively, it has been proposed that heterogeneous LCRs (Figure 7B) also play this role by separating the functional domains to be synthesized, their size compensating for the diversity of the codons used. In other words, the more complex the sequence, the longer it must be to allow co-translational folding of proteins containing several structural domains (Frugier et al., 2010).

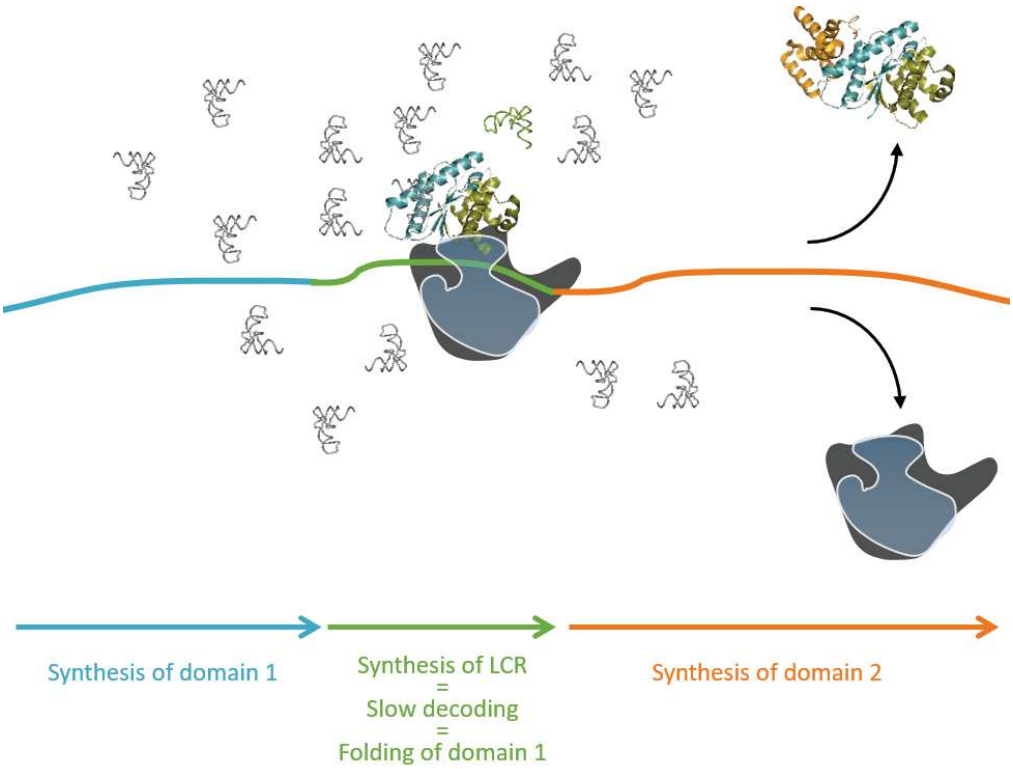


Figure 8: LCRs act as tRNA sponges. Summary of the concept described in Frugier et al. (2010). During the translation of a protein containing several structural domains, the limiting concentration of $N\text{-}tRNA^N$ in *Plasmodium* would slow down the ribosome when decoding mRNA sequences corresponding to asparagine-rich LCRs (in green). This process would give some time for the first domain (blue) to fold independently before the synthesis of the second domain of the protein (orange).

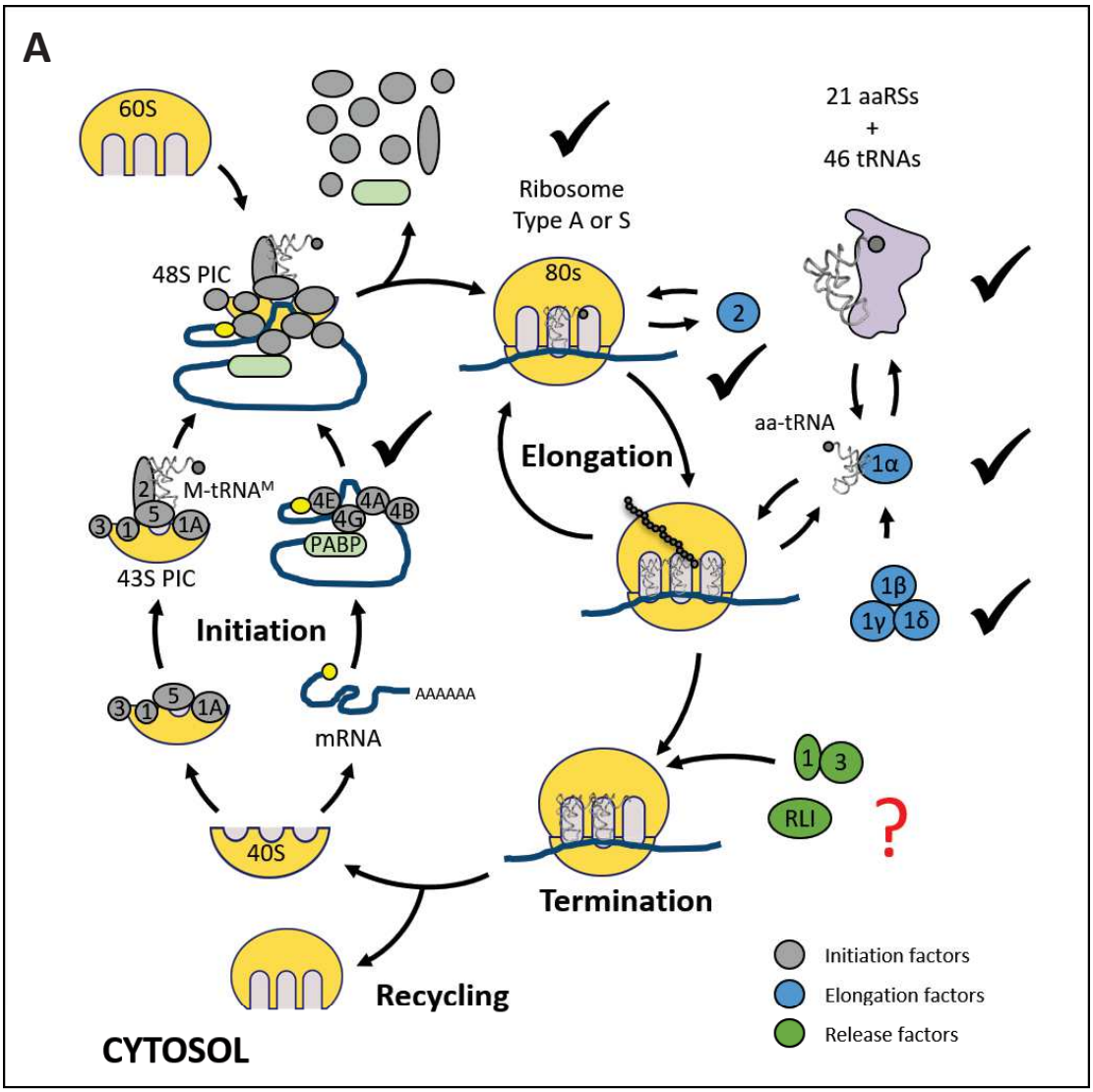


Figure 9. Translational machineries in *Plasmodium*. Protein translation is a four-step process that includes initiation, elongation of the polypeptide chain, termination and recycling of the ribosome. **A. Simplified cytosolic translation.** Like in other eukaryotes, *Plasmodium* ribosomes (80S) are composed of a 40S small subunit and a 60S large subunit. The initiation step is the more complex. It begins when the small subunit recruits eIF1, eIF1A, eIF3, eIF5 and eIF2 complexed with the initiator M-tRNA^M to form the 43S pre-initiation complex (PIC). In parallel, the messenger RNA (mRNA) binds the complex eIF4F (eIF4B:eIF4E:eIF4G:eIF4A) and several poly-A binding proteins (PABPs) to form a circular ribonucleoprotein. This structure is assembled into the 43S PIC to form a larger 48S complex that scans for the start codon. Upon recognition, the initiation factors are released and the large subunit is recruited (Melnikov et al., 2012). In the elongation step, the 40S subunit matches the codons in mRNA with the correct aminoacyl-tRNA (aa-tRNA) while the 60S subunit transfers the amino acid to the peptide growing chain. Three elongation factors are implicated: eEF1α delivers the aa-tRNA to the ribosome powered by GTP, eEF2 allows ribosome translocation to the next codon and eEF1βγδ recharges eEF1α with GTP to deliver the next aa-tRNA (Andersen et al., 2003). Synthesis of aa-tRNAs is performed by aminoacyl-tRNA synthetases (aaRSs) and there is generally one distinct enzyme for each amino acid (Ibba & Söll, 2000). Once all amino acids have been added, several release factors (in green, eRF1 and eRF3) recognize the stop codons (UAA, UAG or UGA) and the protein translation is terminated (Adio et al., 2017). Finally, the RLI protein (also called ABCE1) helps to split the two subunits of the ribosome in order to recycle them for another round of translation (Becker et al., 2012; Hellen, 2018). Check marks: factors experimentally characterized in *Plasmodium*. Question marks: further research is required (*Figure continues in page 26*).

3.3.5. LCRs, a puzzle for the production of recombinant proteins

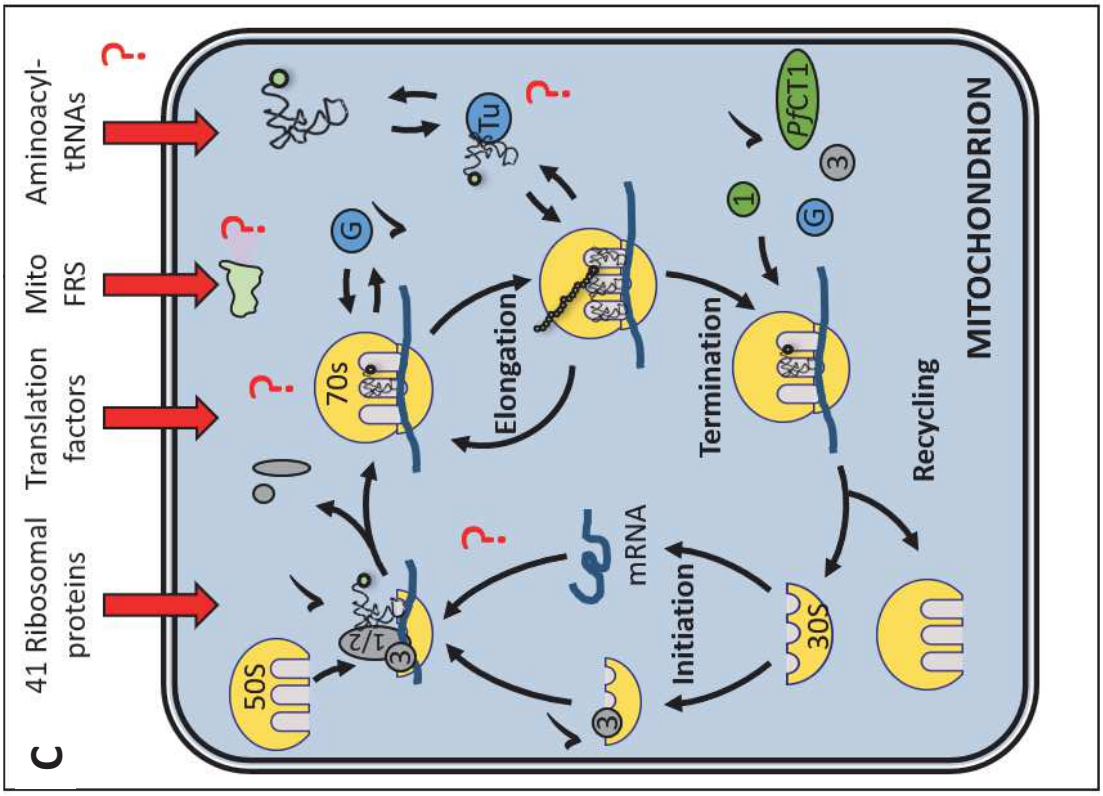
Expressing *Plasmodium* proteins in heterologous systems is a challenging task. The high AT content of *Plasmodium* DNA makes difficult its manipulation into expression vectors and the nucleotide and amino acids biases are not always tolerated by standard eukaryotic or prokaryotic expression systems. For too many years, those that aspired to express *Plasmodium* proteins were routinely frustrated by insoluble pellets and small yields. Advances in DNA synthesis (and its cost reduction) have improved our ability to express *Plasmodium* proteins. Now, researchers can easily optimize both AT content and codon usage to suit the specific expression system that they are using. Despite this, good expression and solubility is not always a guarantee, as *Plasmodium* proteins not only have an unusual composition but also a strange architecture (Tham et al., 2017).

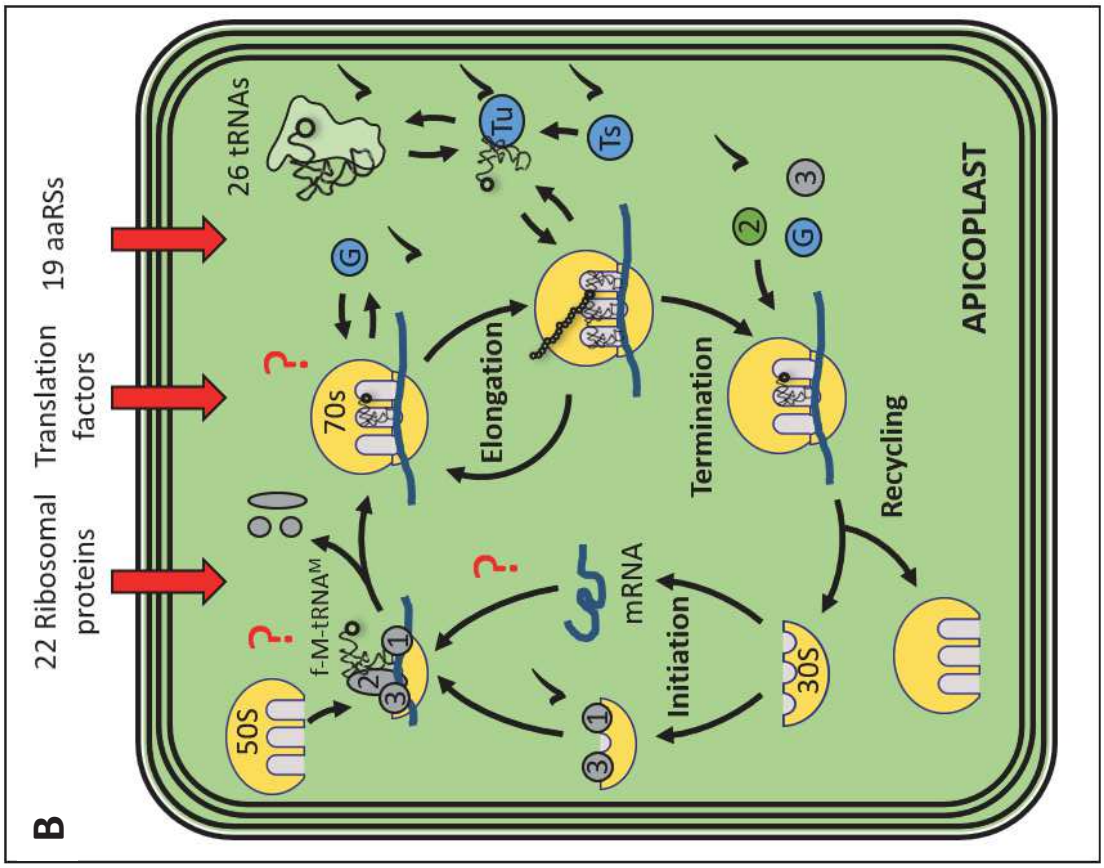
4. Three translation-active compartments

Plasmodium is a unicellular eukaryote that belongs to the phylum *Apicomplexa*, a large assemblage of parasitic organisms including other well-studied parasites such as *Toxoplasma*, *Babesia*, *Theileria*, *Eimeria* and *Cryptosporidium* (Aravind et al., 2003). These parasites possess 3 different compartments where translation occurs: the cytosol, the mitochondrion and a relic plastid called apicoplast (Jackson et al., 2011). The current version of the *P. falciparum* genome (2015-06-18) comprises 23.33 Mb encoding 5712 predicted genes distributed in 14 linear chromosomes of different sizes. Compared to other free-living eukaryotic microbes, *Plasmodium* encodes fewer enzymes and transporters, but a large proportion of genes are devoted to immune evasion and host-parasite interactions (Gardner et al., 2002). In addition to the 550 exported proteins, approximately 500 nuclear-encoded proteins are predicted to be targeted to the apicoplast (Gardner et al., 2002) and 300 to the mitochondrion (Ke & Mather, 2017).

4.1. Cytoplasmic translation machinery in *Plasmodium*

The core components of the protein synthesis machinery are highly conserved in *Plasmodium*. These include the ribosome, translation factors, aminoacyl-tRNA synthetases (aaRSs) and transfer RNAs (tRNAs) (Figure 9A). Despite the conservation of this essential process, the translation machinery of *Plasmodium* shows several unique features that can be exploited for the development of new antimalarial drugs. Additionally, targeting housekeeping pathways such as protein translation has advantages over other targets because they don't change with the stage of the parasite (Khan, 2016).





Legend in next page.

Figure 9 (*Continuation*). Translational machineries in *Plasmodium*.

B. Translation machinery in the apicoplast. The apicoplast genome encodes a complete set of tRNAs, some ribosomal components and the elongation factor EF-Tu. All the other components of translation apparatus are nuclear-encoded and targeted to the organelle. It includes a complete set of enzymes for tRNA aminoacylation (19 aaRSs + 1 amidotransferase), most translation factors and several ribosomal proteins. Translation initiates with the formation of a complex involving the 30S small subunit and the initiation factors IF1 and IF3. The complex binds the mRNA and IF2 carrying the formyl-M-tRNA^{M} initiator, which allows the incorporation of the 50S large subunit. The initiation factors are released and the elongation of the protein chain proceeds with the repeated action of the elongation factors EF-Tu and EF-G, with EF-Ts recharging EF-Tu with GTP. Protein synthesis is terminated by the release factor RF2 that recognizes the stop codons (UAA or UGA) and hydrolyzes the peptide chain from tRNA. RF2, together with IF3 and EF-G, contribute to the dissociation of the ribosomal subunits and their recycling for the next round of translation.

C. Translation machinery in the mitochondrion. The small mitochondrion genome encodes only three proteins and some ribosomal RNA fragments, thus most components of the translation machinery are nuclear-encoded and imported into the organelle. One aaRS (FRS) is imported to the mitochondrion but its role is unclear (Sharma & Sharma, 2015). Translation initiation involves only two factors, IF3 and an IF2/IF1 hybrid. Elongation is also performed with a reduced set of factors as there is no obvious mitochondrial EF-Ts encoded in the nuclear genome. However, termination and ribosome recycling involve an additional release factor (*Pf*CT1) besides RF1, EF-G and IF-3. Check marks: factors experimentally characterized in *Plasmodium*. Question marks: further research required. Adapted from Habib et al. (2016).

4.1.1. Ribosomes: Are two better than one?

Unlike many eukaryotes, *Plasmodium* does not possess long repeated arrangements of ribosomal RNA (rRNA) genes. Instead, *Plasmodium* genomes contain only 4 to 8 single-copy units of rRNA genes (Gardner et al., 2002). Thus, while most other organisms have optimized ribosome production, it is still not known how the malaria parasite produces enough ribosomes. Moreover, *Plasmodium* species have two structurally distinct and stage-specific ribosomes (Gunderson et al., 1987; McCutchan et al., 1988; Waters et al., 1989). The difference in sequence and expression profile over the life cycle has classified them as type A (for Asexual stage specific) and type S (for Sporozoite specific). *P. vivax* is an exception with a third rRNA type O (Li et al., 1997; Van Spaendonk et al., 2001). Thus, type A ribosomes are present at the liver and blood stages and type S ribosomes are found both in the mosquito stages and in vertebrate hepatocytes (at the beginning of the liver stage) (Zhu et al., 1990). Types A and S rRNAs are not expressed in an exclusive manner, but rather as a dynamic and heterogeneous population in which one subtype is the most dominant at a particular stage of the life cycle (Li et al., 1997).

Plasmodium ribosomes must deal with the translation of exceptionally high AU content and long poly-A stretches in mRNAs. Ribosome profiling and biochemical assays suggest an increased or modified fidelity such that parasite ribosomes do not stall or frameshift on poly-A tracks (Djuranovic et al., 2020). The mechanism of this altered fidelity may result not only from the modification of the ribosomal RNA sequence, but also from changes in the protein components of ribosomes.

The cryo-EM structure of the cytoplasmic 80S ribosome from blood stage *P. falciparum* has been solved at 3.2 Å resolution (Wong et al., 2014). It includes 74 ribosomal proteins but, for example, *Pf*RACK1 is absent (Sun et al., 2015). Thus, modified rRNA sequence, ribosomal structure, ribosomal proteins would help *Plasmodium* to adapt to the translation of the mRNA poly-A/U tracks coding for homopolymeric repeats.

4.1.2. Translation factors

Initiation of cap-dependent translation seems to be conserved in *Plasmodium* as interactions between *P. falciparum* eIF4G, eIF4E and PABP have been demonstrated *in vitro* (Shaw et al., 2007; Tuteja & Pradhan, 2009; Tuteja & Pradhan, 2010; Vembar et al., 2016). In contrast, eIF4B has not been yet identified in the *P. falciparum* genome; however, the sequence of this molecule is poorly conserved in eukaryotes (Jackson et al., 2011).

The search for the elongation factor complex lead to the identification of interactions between the native proteins of the parasite eEF1β, eEF1δ and eEF1γ, also associated with eEF1α (Takebe et al., 2007). Since the accuracy and efficiency of translation of AU-rich mRNA may be determined by ribosome stalling, it would be useful to further characterize translation elongation in malaria parasites (Vembar et al., 2016). Additionally, a new multi-stage antimalarial drug has been found to target the translation elongation factor eEF2, which is responsible for the GTP-dependent translocation of the ribosome along the mRNA (Baragaña et al., 2015).

There are at least three genes encoding putative release factors in the nuclear genome of *Plasmodium*. However, the corresponding proteins have not been characterized yet and any aspects of release factors are yet to be studied, including their localization and their mechanisms of action (Patankar et al., 2013).

4.1.3. tRNAs

The *Plasmodium* nuclear genome contains a total of 46 tRNA genes, encoding 45 tRNA isoacceptors (2 different genes encode the initiator and elongator $tRNA^M$) (Jackson et al., 2011). These include a selenocysteinyl-tRNA ($tRNA^U$) that places selenocysteine at an internal stop codon (UGA) influenced by a *cis*-disposed element in the mRNA (Gardner et al., 2002; Mourier et al., 2005). *Plasmodium* cytosolic tRNAs are similar to eukaryotic tRNAs and adopt the canonical L-shaped structure (Figure 10). Despite the AT richness of the genome, they have a balanced content of purines and pyrimidines (about 56% of G+C content) (Pütz et al., 2010).

Strikingly, there is only one gene copy per tRNA isoacceptor, which makes *Plasmodium* the eukaryotic cell with the smallest set of tRNA genes (Gardner et al., 2002). Indeed, in most eukaryotes, tRNA genes are present in multiple copies and the abundance of tRNA isoacceptors is correlated with the codon usage of the organism (Moriyama & Powel, 1997; Duret, 2000). It remains unclear how the parasite does to accurately decode 61 codons with only 45 cytoplasmic tRNA isoacceptors, although a model in which tRNA modifications modulate the translation efficiency of codon-biased proteins has been proposed (Ng et al., 2018).

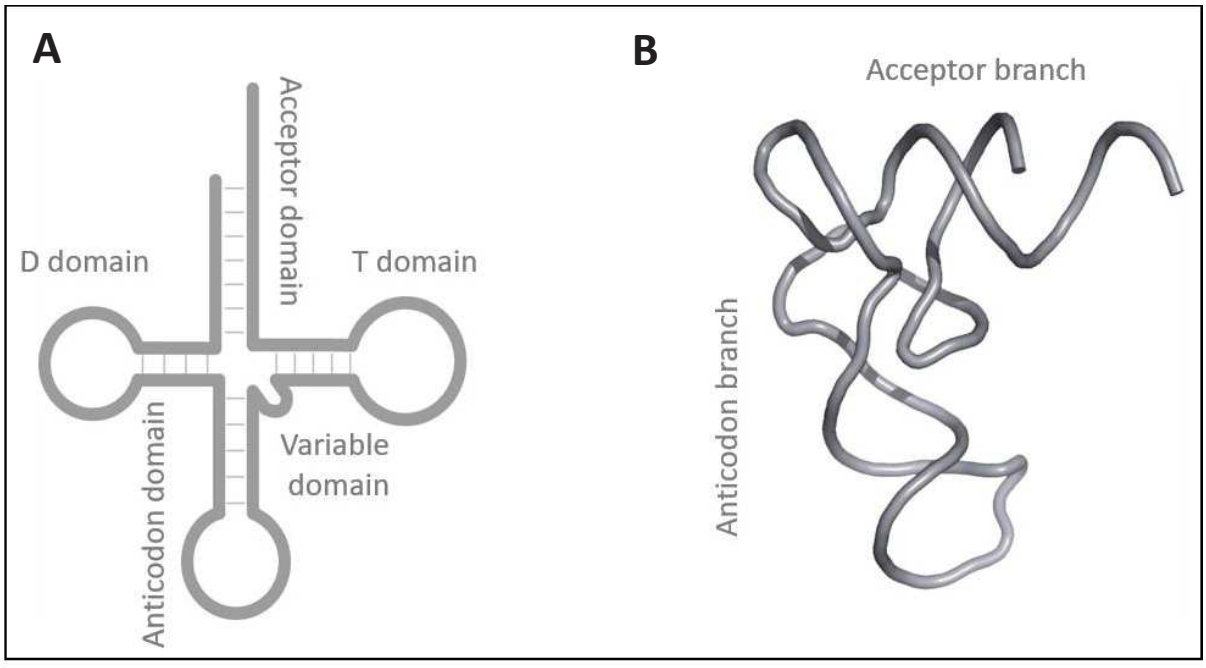


Figure 10: Secondary and tertiary structures of tRNAs. A. Classic cloverleaf folding. The names of the 5 tRNA domains are indicated. **B. Characteristic L-shaped structure of tRNAs.** The acceptor and anticodon branches are indicated.

4.2. Translation machinery in organelles

4.2.1. Translation in apicoplast

The apicoplast is a non-photosynthetic organelle of red-algal origin maintained by all apicomplexans (except *Cryptosporidium*). It was probably acquired by an event of secondary endosymbiosis between two ancestral eukaryotes, which resulted in the four membranes that now surround the organelle. Several processes including the metabolism of heme and iron, the biosynthesis of fatty acids and biosynthesis of isoprenoid precursors occur within this organelle (McFadden & Yeh., 2017).

Apicoplast proteins are translated by a synthesis machinery of prokaryotic origin (Figure 9B). Many bacterial translation inhibitors target the apicoplast and have been used to prevent and treat infections by apicomplexans, although their effect is slow. Three classes of antibiotics - tetracyclines, lincosamides and macrolides - are currently approved for use in malaria infections (Goodman et al., 2016). Parasites treated with these drugs show a "delayed death" effect, where they complete one cycle of infection normally, but fail to complete a second one even if drug treatment is stopped in the first cycle. The parasites transmit an apicoplast to their progeny, but this apicoplast is defective and causes the death of the daughter cell (Goodman et al., 2016). Cultures of *P. falciparum* treated with the doxycline antibiotic can be rescued from the delayed death effect by supplementing the grow media with isopenthyl-pyrophosphate (IPP), the isoprenoid product of the apicoplast (Yeh & DeRisi, 2011).

The 35 kb circular genome of the apicoplast is AT-rich (86.9%) and includes 68 genes encoding the large and small subunit ribosomal RNAs (rRNAs), a minimal but complete set of tRNAs (35), several ribosome proteins (18), the three subunits of a bacteria-like RNA polymerase, the translation elongation factor EF-Tu, an iron-sulfur assembly pathway protein (SufB), a Clp protease and chaperones (Wilson et al., 1996; Milton & Nelson, 2016). Detection of the EF-Tu gene product by western-blot and immunofluorescence demonstrated that translation does occur within the apicoplast (Chaubey et al., 2005). However, most genes involved in the apicoplast functions have been transferred to the nuclear genome and are first translated in the cytoplasm and then transported into the organelle (Garcia et al., 2008).

The apicoplast genome contains 35 genes encoding 26 tRNA isoacceptors. Similar to cytoplasmic versions, these tRNAs contain all elements required to adopt the canonical cloverleaf structure (Pütz et al., 2010) (Figure 10). Interestingly, the initiator $tRNA^M$ shows a 11-nucleotides long

variable region, which distinguish it from elongator $tRNA^M$ and cytoplasmic $tRNA^M$. Contrary to their cytoplasmic counterparts, apicoplast tRNAs have a lower content of GC of about 26%. The identification of apicoplast-targeted methionyl-tRNA-formyltransferase (MFT) and peptide deformylase (PDF) suggest that formylated methionyl-$tRNA^M$ functions as initiator of apicoplast translation. However, the IF2 recruiting this tRNA remains unidentified (Haider et al., 2015).

The apicoplast ribosome is composed of 23S rRNA, 16S rRNA and a total of 40 ribosomal proteins (Gupta et al., 2014). The rRNA genes are present in two copies (Wilson et al., 1996). From the 16 small subunit ribosomal proteins, 10 are encoded in the apicoplast and 6 are nuclear-encoded. In the case of the 24 large subunit proteins, 8 are apicoplast-encoded and 16 nuclear-encoded (Gupta et al., 2014). As the apicoplast ribosome lacks 5S rRNA, some ribosomal proteins that normally interact with this rRNA are absent as well (e.g. L5, L25). Protein S13 is also missing, which is surprising as this protein plays an important role in translocation and is essential for translation in bacteria (Cukras & Green, 2005). Although some models have been proposed, there is no structure reported on this reduced ribosome yet.

All translation factors required for initiation, elongation and termination are nuclear-encoded, with the exception of EF-Tu. Apicoplast genes are transcribed as mono- or poly-cistrons and do not contain a Shine-Dalgarno (SD) sequence for ribosome binding. They probably use another mechanism to position the ribosome on the initiation codon as is observed in chloroplasts (Marin-Navarro et al., 2007).

The initiation factors IF1 and IF3 have been identified and characterized in *P. falciparum*. Apart from its interaction with the ribosome, *Pf*IF1 contains an OB-fold that acts as nucleic acid chaperone, allowing the melting of nucleic acid secondary structures (Haider et al., 2015). In the same study, two IF2 candidates were identified, but none of them was localized at the apicoplast (Haider et al., 2015).

EF-Ts was identified, it interacts with EF-Tu and its nucleotide exchange activity has been confirmed (Biswas et al., 2011). EF-G was also identified and indirect evidence of its activity was provided by the inhibition of *P. falciparum* growth in presence of fusidic acid, an antibiotic that blocks specifically EF-G on the ribosome (Johnson et al., 2011).

Finally, apicoplast ORFs have either UAA or UGA as stop codons, with UAA being the most frequently used. These codons are recognized by a single nuclear-encoded RF2 that localizes to the apicoplast (Vaishya et al., 2016).

4.2.2. Translation machinery in the mitochondrion

The 6-kb mitochondrial genome of *Plasmodium* is one of the smallest known to date. It only encodes three proteins, the cytochrome *b* and the subunits I and III of cytochrome oxidase (Cox1 and Cox3) (Vaidya et al., 1989). Translation of the proteins encoded by the mitochondrial genome is essential for the parasite and they have been validated as drug targets (Goodman, 2017). The antimalarial atovaquone targets the mitochondrial cytochrome *bc*1 complex and point mutations in mitochondrial *cytb* gene correlate with resistance to this drug (Afonso et al., 2010).

Since the mitochondrial genome does not contain any aaRS or tRNA genes, it suggests that the mitochondrion relies on the import of all aaRSs and tRNAs or on the import of all aminoacyl-tRNAs from the cytosol in order to translate its three ORFs (Figure 9C). Import of aminoacylated tRNAs (reviewed in Schneider, 2011) was established in organisms such as yeast (Martin et al., 1977), *Leishmania* (Simpson et al., 1989), *Trypanosoma* (Hancock & Hadjuk, 1990) and *Toxoplasma* (Esseiva et al., 2004). Moreover, there is only one nuclear-encoded aaRS that is potentially targeted to the mitochondria. This FRS, that is specific to *Plasmodium* species, was shown to co-localize with the mitochondria and it has been proposed that it could act as a local sensor for phenylalanine levels in the organelle, regulating this amino acid by consuming it *via* the aminoacylation reaction (Sharma & Sharma, 2015).

The mitochondrial ribosome of *Plasmodium* remains largely uncharacterized. The *P. falciparum* mitochondrial genome contains 27 rRNAs fragments ranging from 23 to 190 nucleotides that contain highly conserved portions of large and small subunit rRNAs. It has been proposed that they may associate into functional ribosomes (Feagin et al., 1997). As the mitochondrial genome is polycistronically transcribed, these small RNA are by definition expressed (Ji et al., 1996). However, these rRNA fragments are not encoded in linear order. Instead, they are intermixed with one another and the protein genes, and are coded on both DNA strands. Mapping of these fragments on the ribosome of *T. thermophilus* and *H. marismortui* revealed that they cluster on the interface between the two subunits (Feagin et al. 2012). Although the rRNA is highly reduced, it retains the peptidyl transferase center and the peptide exit tunnel where most antibiotics bind (Gupta et al., 2014). A total of 41 ribosomal proteins have been detected, 14 for the small subunit and 27 for the large subunit, all of them are nuclear-encoded (Gupta et al., 2014).

Plasmodium mitochondrial genes do not contain SD sequence and all translation factors are encoded by the nuclear genome (Figure 9C). Similar to other mitochondria, translation initiation requires IF2 and IF3, but no IF1. *Pf*IF2a and *Pf*IF3 were both localized at the mitochondrion and

*Pf*IF2a was able to interact with initiator tRNA charged with either formylated or unformylated methionine *in vitro* (Haider et al., 2015). *Pf*IF2a also contains a 29-amino acids extension that mimics the function of IF1, explaining the absence of this factor (Haider et al., 2015).

The mitochondrial EF-Tu was annotated in the nuclear genome, but its final localization has not been demonstrated yet. The presence of EF-G in the mitochondrion was confirmed by Johnson et al., (2011). There is no mitochondrial EF-Ts candidate, but it was proposed that its absence might be tolerated and that the slow recycling of EF-Tu·GDP to EF-Tu·GTP may suffice for the translation of the three mitochondrial genes (Habib et al., 2016).

UAA is the stop codon in the three mitochondrial ORFs and it is recognized by RF1. Additionally, a second non-canonical RF, called *Pf*CT1, is imported to the mitochondrion where it mediates the release of prematurely terminated proteins (Vaishya et al., 2016).

4.3. *Plasmodium* aminoacyl-tRNA synthetases as targets for anti-malarial drugs

The nuclear genome codes for 37 aaRS genes and neither the apicoplast nor the mitochondrion encode any aaRS (Bhatt et al. 2009). As phenylalanyl-tRNA synthetase (FRS) is composed of two subunits encoded by different genes, the 37 nuclear genes encode a total of 36 enzymes. Several localization experiments and bioinformatics studies revealed that 16 aaRS are found exclusively in the cytoplasm, 15 are targeted to the apicoplast, 4 are dual-targeted to the cytoplasm and apicoplast and only 1 would be exported to the mitochondrion.

The crystal structure of several cytoplasmic aaRSs has been solved (Bhatt et al., 2011; Koh et al., 2013; S. Khan, Garg, Sharma et al., 2013; S. Khan, Garg, Camacho et al., 2013; Jain et al., 2014; Jain et al., 2016; Sonoiki et al., 2016) and several inhibitors have been characterized. Cladosporin, a secondary metabolite from fungi, inhibits *P. falciparum* growth in the blood and liver stages by targeting selectively its cytoplasmic lysyl-tRNA synthetase (KRS) (Hoepfner et al., 2012). Febrifugine, the active principle of a traditional Chinese herbal remedy for malaria, and its derivatives are strong inhibitors of cytoplasmic prolyl-tRNA synthetase (PRS), but show some toxicity for human cells (Keller et al., 2012; Herman et al., 2015).

Interestingly, although the parasite regulates excess cellular heme during the digestion of hemoglobin, some is still encountered in malaria parasites treated with chloroquine. It has been shown that in *P. falciparum*, heme binds to the monomeric cytosolic arginyl-tRNA synthetase

(RRS), and induces a dimeric form of this enzyme (Jain et al., 2016). As a dimer, the enzymatic activity of RRS is inhibited since it can no longer bind its homologous $tRNA^R$. Treatment with chloroquine would thus lead to reduced levels of charged $tRNA^R$, suggesting reduced protein synthesis.

Among *Plasmodium* aaRSs, FRS is unique because the parasite contains four genes that encode three enzymes, one for each compartment. The cytoplasmic FRS is an $(\alpha\beta)_2$ heterotetramer while the mitochondrial and apicoplast enzymes are α-monomers (Sharma & Sharma, 2015). Cytoplasmic FRS also contains a predicted nuclear localization signal (Bhatt et al., 2009), which is not unusual as some tRNAs may need to be aminoacylated prior to export in the cytoplasm (Lund & Dahlberg, 1998). Additionally, as many eukaryotic FRS (Perona & Hadd, 2012), the cytosolic enzyme contains DNA binding domains, which is coherent with a possible nuclear localization (Bhatt et al., 2009). Several compounds, including a series of bicyclic azetidines, are reported as potential drugs targeting the cytoplasmic FRS (Kato et al., 2016).

Apicoplast aaRSs in general and the four dual-targeted aaRSs in particular are attractive targets for drug development. Dual-targeted aaRSs, referred as ATGC enzymes by Yogavel et al. (2018), correspond to alanyl-, threonyl-, glycyl- and cysteinyl-tRNA synthetases (ARS, TRS, GRS and CRS, respectively). A single transcript for each gene is alternatively spliced to generate the two isoforms that are targeted to either the cytosol or to the apicoplast (Jackson et al., 2012; Pham et al., 2014). Targeting any of these enzymes will arrest translation in both compartments simultaneously. The natural macrolide borrelidin clears malaria parasites from mice by targeting TRS, but lacks specificity over the human enzyme (Novoa et al., 2014). However, some borrelidin analogues have been synthesized and show less toxicity (Sugawara et al., 2013). Mupirocin is a well-known inhibitor of bacterial isoleucyl-tRNA synthetase (IRS) that kills blood-stage *P. falciparum* by targeting the apicoplast IRS and shows delayed death effect (Jackson et al., 2012; Istvan et al., 2011).

Finally, *Plasmodium* does not code for an apicoplastic glutaminyl-tRNA synthetase (QRS) and produce glutaminyl-tRNA *via* a two-step indirect aminoacylation pathway in this organelle. First, a non-discriminating glutamyl-tRNA synthetase (ERS) misacylates $tRNA^Q$ with glutamate. Then, a tRNA-dependent glutamyl-tRNA amidotransferase (E-AdT) convert the E-$tRNA^Q$ into Q-$tRNA^Q$ (Mailu et al., 2013). E-AdT is a heterodimeric enzyme composed of two nuclear-encoded subunits, GatA and GatB. This complex is essential for the parasite in the blood stage (Mailu et al., 2015) and thus could also be a good target for anti-malarial development.

II. aaRSs in translation and beyond

1. The specificity of the aminoacylation reaction

In addition to codon-anticodon recognition, the fidelity of protein translation depends on the accuracy of aminoacyl-tRNAs synthesis. The aminoacyl-tRNA synthetases (aaRS) are the enzymes that ensure the proper attachment of an amino acid to its corresponding tRNA. Generally, there are 20 different aaRSs, one for each canonical amino acid and its corresponding tRNA isoacceptors (Ibba & Söll, 2000). The aaRSs are modular enzymes composed of several domains that have distinct roles in the aminoacylation reaction (Delarue & Moras, 1993; Alexander & Schimmel, 2001). All aaRSs contain a catalytic domain (CD) that binds ATP, amino acid and the 3'-end of the tRNA. Most of them contain an anticodon-binding domain (ABD) appended to their N- or C-terminus that contributes significantly to the efficiency and specificity of the aminoacylation. Some aaRSs contains editing domains embedded or appended to their CD, which perform proofreading functions and ensure the accuracy of aminoacylation (Yadavalli & Ibba, 2012). Additionally, eukaryotic aaRSs often contain appended domains that are implicated in functions non related with tRNA aminoacylation.

The aminoacyltation reaction occurs in two highly specific steps, resulting in the 3'-esterification of tRNA with the appropriate amino acid. In the first step (Figure 11A), the amino acid is activated with ATP:Mg^{2+}, leading to the formation of a stable aaRS-aminoacyl-adenylate complex and the release of pyrophosphate (PPi). This occurs by nucleophilic attack on the α-phosphate of ATP by the carboxyl group of the amino acid. In the second step (Figure 11B), the 3'-terminal adenosine of the enzyme-bound tRNA reacts with the aminoacyl-adenylate, leading to its esterification with the amino acid and the release of AMP (Ibba & Söll, 2000). Later, the aminoacyl-tRNA, also referred as "charged tRNA", is detached from the aaRS and transported to the ribosome.

The aaRSs must perform their task with high accuracy, as every mistake will result in a misplaced amino acid in the newly synthesized protein. These enzymes make about one mistake in 10,000 (Goodsell, 2001). Accurate aminoacylation depends on the specific recognition of both the tRNA and the amino acid. An aaRS must be able to recognize all isoacceptor tRNAs for the corresponding amino acid and only them. This is directed by the presence of tRNA identity elements in the molecule (Giegé et al., 1998). Some nucleotides act as positive elements (determinants) and promote a productive interaction between the tRNA and the aaRS, others serve as negative elements (antideterminants) that prevent mischarging of non-cognate tRNAs.

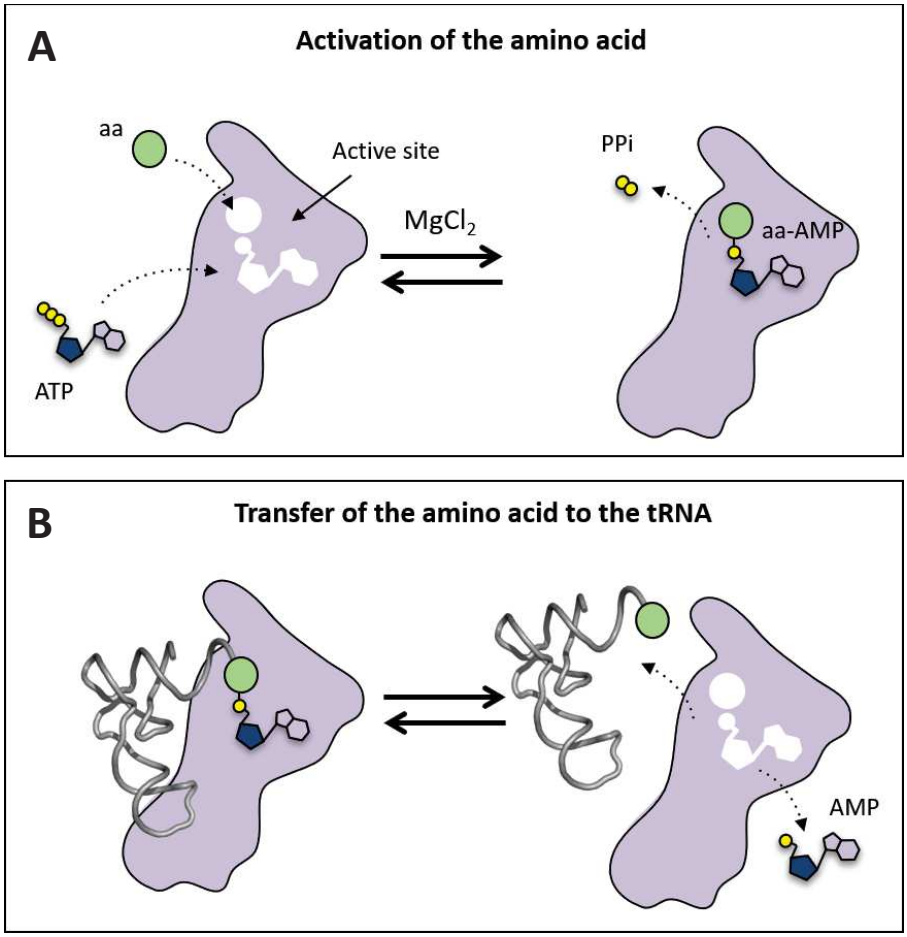


Figure 11. The aminoacylation reaction. A. Amino acid activation. The amino acid (aa) is activated in the active site of the enzyme with ATP in the presence of $MgCl_2$, which results in the formation of an aminoacyl-adenylate (aa-AMP) intermediate and the release of pyrophosphate (PPi). **B. amino acid transfer.** The enzyme binds tRNA and the amino acid moeity of aa-AMP is transferred to its 3' extremity, forming the product aa-tRNA while releasing AMP. Figure adapted from Rajendran et al. (2018).

Discrimination of the correct amino acid can be difficult for some aaRSs. Isoleucine is a classic example. This amino acid is recognized by an isoleucine-shaped hole in the active site of isoleucyl-tRNA synthetase (IRS), which is too small to fit larger amino acids such as methionine and phenylalanine, and too hydrophobic to bind polar amino acids of similar size. However, valine differs only by one single methyl group and fits nicely into this pocket, binding in place of isoleucine in about 1 case in 150 (Goodsell, 2001). The IRS solves this problem with a second active site that perform an editing reaction. Isoleucine does not fit the editing active site, but valine does (Fukai et al., 2000). A similar editing site is used to distinguish between phenylalanine and tyrosine, which only differ by a hydroxyl group (Roy et al., 2004). Besides the mechanism described above, aaRSs employ many others proofreading pathways that can occur either after activation prior to aminoacyl transfer (pre-transfer editing) or after transfer (post-transfer editing) (Yadavalli & Ibba, 2012).

2. Classification and architecture and of aaRSs

2.1. Two classes of synthetases

Based on differences in the structure of their CD, aaRSs are classified in two groups: class I and class II (Eriani et al., 1990; Cusack et al., 1990; Cusack, 1993) (Table 1). They are further divided into different subclasses based on phylogenetic analysis, comparison of structural and mechanical characteristics and domain organization (Rubio-Gomez & Ibba, 2020). Both Class I and II are divided into three subclasses (Cusack, 1995; Ribas de Pouplana & Schimmel, 2001), although some authors divide class I aaRSs into five (Perona & Hadd, 2012).

Table 1: Classification of aaRSs into classes and subclasses. Classification according to Perona & Hadd (2016). The typical quaternary structure is indicated as monomer (α), homodimer (α_2) and heterotetramers [$(\alpha\beta)_2$]. Note that KRS is present in both classes and GRS is placed in two different class II subclasses depending on its quaternary structure.

Class	Subclass	Name	Abbreviation	Quaternary structure
I	Ia	Isoleucyl-tRNA synthetase	IRS	α
		Leucyl-tRNA synthetase	LRS	α
		Methionyl-tRNA synthetase	MRS	α, α_2
		Valyl-tRNA synthetase	VRS	α
	Ib	Cysteinyl-tRNA synthetase	CRS	α, α_2
		Glutamyl-tRNA synthetase	ERS	α
		Glutaminyl-tRNA synthetase	QRS	α
	Ic	Tyrosyl-tRNA synthetase	YRS	α, α_2
		Tryptophanyl-tRNA synthetase	WRS	α_2
	Id	Arginyl-tRNA synthetase	RRS	α
	Ie	Lysyl-tRNA synthetase	KRS	α
II	IIa	Seryl-tRNA synthetase	SRS	α_2
		Prolyl-tRNA synthetase	PRS	α_2
		Threonyl-tRNA synthetase	TRS	α_2
		Glycyl-tRNA synthetase	GRS	α_2
		Histidyl-tRNA synthetase	HRS	α_2
	IIb	Aspartyl-tRNA synthetase	DRS	α_2
		Asparaginyl-tRNA synthetase	NRS	α_2
		Lysyl-tRNA synthetase	KRS	α_2
	IIc	Phenylalanyl-tRNA synthetase	FRS	α, $(\alpha\beta)_2$
		Glycyl-tRNA synthetase	GRS	$(\alpha\beta)_2$
		Alanyl-tRNA synthetase	ARS	α_2

Class I aaRSs are generally monomeric enzymes. They mostly charge bulkier amino acids (Rajendran et al., 2018), bind the acceptor stem of tRNA from the minor groove side (Ribas de Pouplana & Schimmel, 2001) and perform the aminoacylation on the 2'-OH of the ribose of adenosine 76 (Eriani et al., 1990). Binding the minor groove of tRNA is usually correlated with aminoacylation of the 2'-OH, although dimeric YRS and WRS are exceptions, as they bind two tRNAs from the major groove side and still perform aminoacylation on the 2'-OH (Yaremchuk et al., 2002). Interesting relations between aaRSs emerge when considering the grouping into subclasses (Rubio-Gomez & Ibba, 2020). The members of each subclass usually recognize chemically related amino acids. For instance, subclass Ia (IRS, LRS, MRS and VRS) recognizes hydrophobic amino acids and subclass Ic (YRS and WRS) aromatic amino acids. The subclass Ib (CRS, ERS and QRS) usually needs to bind their cognate tRNA before binding ATP and the cognate amino acid (Sekine et al., 2003). RRS is sometimes assigned to subclass Id due to its structural dissimilarity with other subclasses (Perona & Hadd, 2012; Rajendran et al., 2018). Likewise, KRS is similar to subclass Ib, but occupies a separate subclass (Ie) because it is the only aaRS present in both classes (Perona & Hadd, 2012). Class I KRS is mainly found in archaea and some bacteria while class II KRS is found in eukaryotes and most bacteria (Perona & Hadd, 2012).

Most class II aaRSs are dimeric enzymes, although some examples of monomeric (α), tetrameric (α_4) and heterotetrameric $(\alpha\beta)_2$ are known (Perona & Hadd, 2012). All class II aaRSs bind the acceptor stem of the tRNA from the major groove and perform the aminoacylation on the 3'-OH of the terminal adenosine of tRNA (except FRS) (Spritnzl & Cramer, 1975; Ruff et al., 1991). Subclass IIa aaRS (SRS, PRS, TRS, GRS and HRS) charge small polar amino acids, whereas subclass IIb (DRS, NRS and KRS-II) recognize charged and large polar amino acids.

2.2. Universal architecture of aaRSs

2.2.1. Modular organization of Class I aaRSs

The CD of class I enzymes contains a Rossmann fold constituted of alternate α-helices and β-sheets that allow the binding of ATP in its extended conformation (Figure 12A) (Rossmann et al., 1974; Brick et al., 1989; Moras, 1992). It can be divided in two halves, each one containing a signature sequence (Delarue & Moras, 1993). The "HIGH" signature is typically located in the loop after the first β-strand. The two histidines (H) participate in the stabilization of the ATP phosphate chain during the transition state (Schmitt, 1995). The first H and the G are almost invariant in all class I aaRSs, while the second H is sometimes substituted by N (Moras, 1992; Chaliotis et al., 2017). The "KMSKS" signature is located in a mobile loop after the fifth β-strand

and is also involved in ATP stabilization. The two basic K residues neutralize the negative charges carried by ATP, facilitating the nucleophilic attack on the α-phosphate (Schmitt, 1994). Despite the conserved function of this loop, the KMSKS motif is markedly degenerated within the class I enzymes (Schmitt, 1995; O'Donoghue & Luthey-Schulten, 2003). In prokaryotes, however, it seems to be more conserved (Chaliotis et al., 2017).

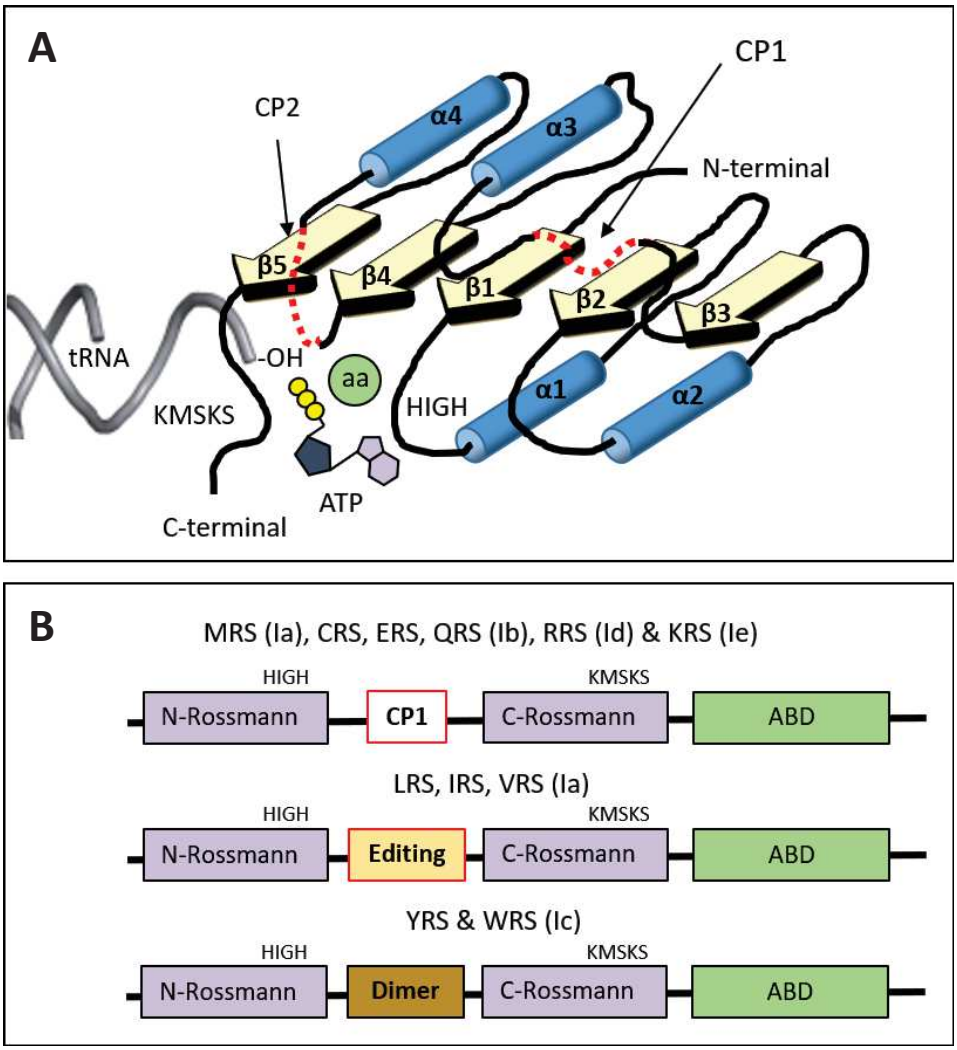


Figure 12: Typical architecture of class I aaRSs. A. The class I Rossmann fold. The catalytic domain of class I aaRSs is a Rossmann fold composed of five parallel β-strand sandwiched between several α-helices. It contains two highly conserved motifs involved in ATP binding. The motif "HIGH" stabilizes the ATP in the transition state while the motif "KMSKS" neutralizes its negative charges to facilitate the reaction with the amino acid. **B. Schematic representation of class I enzymes.** The Rossmann fold is always divided in two halves by the insertion of the connective polypeptide 1 (CP1). The structure and function of this insertion is variable among the different enzymes. In most cases (MRS, CRS, ERS, QRS, RRS and KRS-I), CP1 adopts a mixed α/β structure that contributes to the activity of the enzyme. In LRS, IRS and VRS, CP1 provides an editing domain contributing to the specificity of these enzymes. In YRS and WRS, CP1 is an interaction domain that allows the dimerization of these enzymes. Class I aaRSs possess an anticodon-binding domain (ABD) fused to the C-terminal extremity of the Rossmann fold. The topology of this domain is variable but is generally composed of β-barrels in subclass Ib and of α-helices in all the other enzymes. Figures adapted from Delarue & Moras (1993) and Perona & Hadd (2012).

Class I aaRSs cradle the tRNA by gripping the anticodon loop and placing the acceptor stem in the active site. This is usually accompanied of some structural changes in the tRNA molecule. For instance, the 3'-end is kinked into a tight hairpin inside the active site and, in some cases, the three bases of the anticodon are widely spread apart for better recognition (Rould et al., 1991). Most class I aaRSs contain an ABD located at their C-terminal extremity (Figure 12B). These domains contribute significantly to tRNA discrimination, but their structures are in general divergent, even within subclasses (Perona & Hadd, 2012). The ABD forms a topologically identical α-helix cage in KRS and some ERS (Nureki et al., 1995; Terada et al., 2002) and a pair of β-barrels in QRS and most ERSs (Rould et al., 1989). Subclass Ia aaRSs along with RRS and CRS possess a distinct structurally conserved α-helical ABD (Perona & Hadd, 2012). RRS contains an additional RNA binding domain at its N terminus (Add) that binds the D-loop of $tRNA^R$ (Cavarelli et al. 1998).

Large insertions called connective peptide 1 and 2 (CP1 and CP2) can exist within the Rossmann fold. CP1 separates the fold in two halves and is located between the 3rd and 4th β-strands while CP2 is found in the second half after the 4th β-strand (Delarue & Moras, 1993) (Figure 12A). The CP1 adopts a mixed α/β fold that binds the 3'-single-stranded end of tRNA in the monomeric enzymes (Perona & Hadd, 2012). Some class I aaRSs have an enlarged CP1 incorporating either a post-transfer editing site for hydrolysis of mischarged tRNA (e.g. LRS, IRS, VRS) (Yadavalli & Ibba, 2012) or a dimerization domain (e.g. YRS, WRS) (Perona & Hadd, 2012) (Figure 12B).

2.2.2. Modular organization of Class II aaRSs

The CD of these enzymes binds ATP in a bent conformation and is composed of seven antiparallel β-strands flanked by several α-helices (Figure 13A) (Delarue & Moras, 1993). Three degenerated motifs can be identified within the CD. Motif 1 is made up of an α-helix followed by a distorted β-strand. It includes a conserved proline (P) that is involved in homo-dimerization (Moras, 1992). Motif 2 contains a flexible loop of variable length located between two adjacent β-strands. It is characterized by conserved an arginine (R) involved in the stabilization of ATP α-phosphate during the transition state (Cavarelli et al., 1994). Motif 3 contains the central β-strand followed by a hydrophobic helix and includes also a conserved R that binds the ATP γ-phosphate in a bent conformation (Cavarelli et al., 1994).

The subclass IIa aaRSs are all characterized by a C-terminal ABD consisting of a five-stranded mixed β-sheet surrounded by α-helices. Only SRS lacks this ABD and uses a long antiparallel coiled coil tRNA binding domain (tRBD) to bind its cognate tRNAs. The subclass IIb possess an ABD at their N-termini, which form a β-barrel of the OB fold variety (Perona & Hadd, 2012). The

subclass IIc includes FRS, ARS and GRS, which are the largest and more complex aaRSs. ARS is unusual, because the C-terminal C-Ala domain, which bridges aminoacylation and editing functions and interacts with the outer corner of the tRNA L-shape, also form part of the dimerization interface (Naganuma et al., 2009). GRS is also intriguing because its divergent quaternary structure and abnormal charging properties. It forms α_2 dimers in eukaryotes and archaea, and $(\alpha\beta)_2$ heterotetramers in bacteria. In addition, GRSs only aminoacylate tRNAs from their own domains of life and do not function across species (Qin et al., 2014).

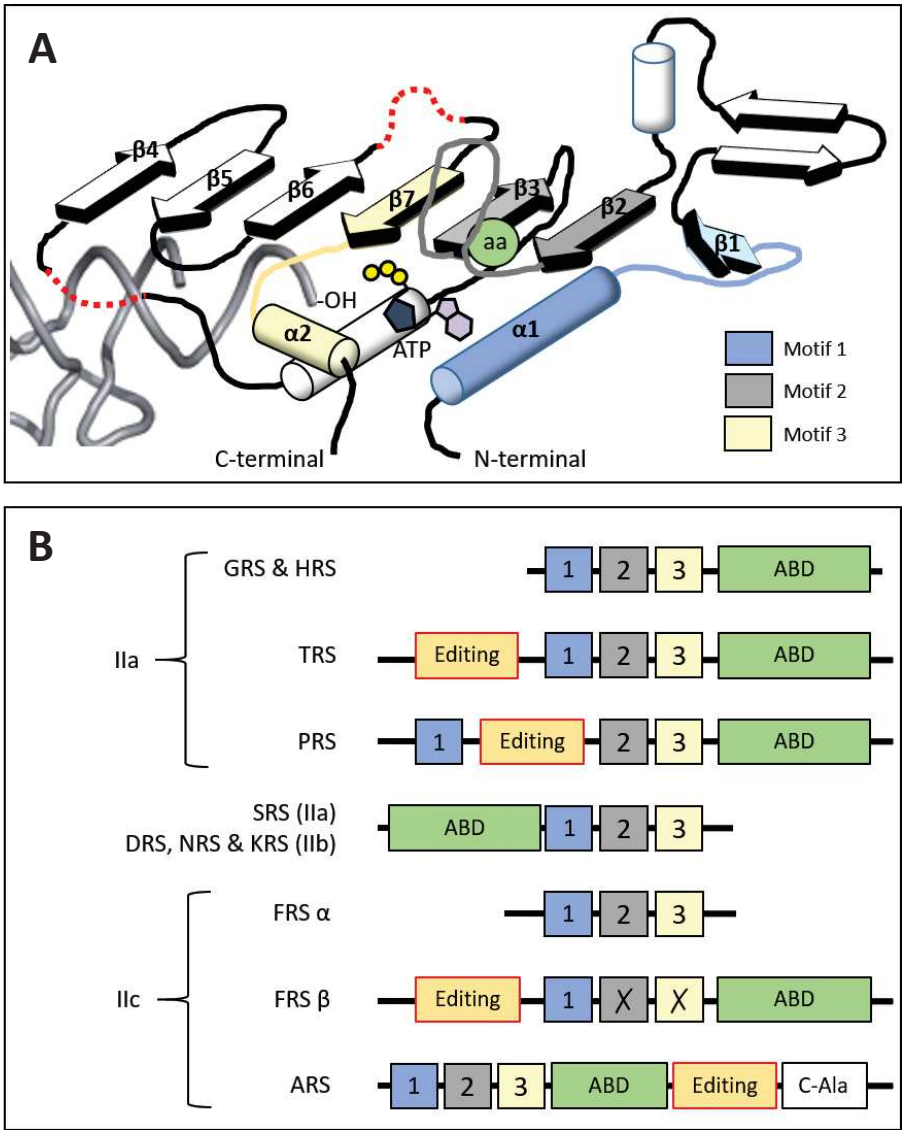


Figure 13. Architecture of class II aaRSs. A. The catalytic domain of class II aaRSs. It is composed of 7 antiparallel β-strands and α-helices and contains 3 more or less conserved motifs. Motif 1 allows dimerization of the aaRS and Motifs 2 and 3 are implicated in binding of ATP in a bent conformation. Insertions CP1 and CP2 are indicated by red dotted lines. **B. Modular organization of class II aaRSs.** Class II aaRSs are very heterogenous but all of them are dimers and include an anticodon-binding domain (ABD). Most of them contain editing domains located in different parts of the protein. FRS is composed of 2 different subunits that form a functional enzyme. ARS has an additional domain C-Ala contributing to both dimerization and tRNA binding. Figures adapted from Delarue & Moras (1993) and Perona & Hadd (2012).

FRS is often composed of two subunits, the α subunit containing the canonical class II CD and the β subunit lacking essential catalytic residues. These two chains form a $(\alpha\beta)_2$ heterotetramer containing two active sites that charge two tRNAs simultaneously. A total of 10 structural domains can be identified in the two FRS subunits. The subunit β contains notably an RNA-binding domain (RBD) similar to the spliceosome protein U1A that functions as ABD. The subunit α contains a coiled coil tRBD (similar to SRS) that contributes to stabilize the interaction with the tRNA (Mosyak et al., 1995; Goldgur et al., 1997; Perona & Hadd, 2012). Beyond the subclasses, four class II aaRS (PRS, TRS, FRS and ARS) display an editing domain, located in different places depending on the enzyme (Figure 13B).

3. aaRSs, much more than translators

Eukaryotic aaRSs have evolved with the addition of new domains. Interestingly, the incorporation of new domains is correlated with the complexity of the organism (Pang et al., 2014). Compared with their prokaryotic and protozoan counterparts, metazoan aaRSs have numerous additional domains appended at either the N- or C-terminus (Mirande, 1991; Wolf et al. 1999). Most of them are involved in the aminoacylation function, either by binding to tRNA (Francin et al., 2002; Crepin et al., 2004) or by forming complexes with other aaRSs (Havrylenko & Mirande, 2015), AIMPs (aaRS-interacting multifunctional proteins) or the elongation factor 1 (Bec et al., 1994; Negrutskii et al., 1999; Sang Lee et al., 2002). Another category of these additional domains is not at all involved in the canonical function of synthetases and contributes to alternative functions only. These functions include the metabolism of glucose and amino acids, the regulation of cell growth, control of angiogenesis, regulation of inflammatory responses, control of cell death, regulation of the immune response and more, many of them being essential for the cell, particularly in higher eukaryotes (Guo & Schimmel, 2013).

Table 2 recapitulates the additional domains appended to human aaRSs (Guo et al. 2010) and some of them are described in the next paragraphs. Among the 23 proteins involved in human aminoacylation, only the ARS is deprived of any additional domain. Other aaRSs contain well-known domains like, EMAPII, WHEP, Leucine-Zipper or GST but also domains that have no sequence similarity with other common structural modules. These specific domains are named UNE-X, where X corresponds to the aaRS to which it is appended (Guo & Yang, 2014). To date, we have no functional information about the CRS UNE-C1 and UNE-C2 found at proximity of the CP1 domain and at the C-terminus, respectively, and neither about the N-terminal UNE-T of TRS.

Table 2. Additional domains in human aaRSs and AIMPs. Additional domains in human aaRSs and AIMPs according to Guo et al. (2010) and Guo & Yang (2014). The position of each type of domain is indicated. N: domain appended at the N-terminus, int: internal domain, C: domain appended at the C-terminus. Slashes are used to indicate several additional domains of a certain type in the same protein.

Protein		Appended domain					
		UNE	α-helix	EMAPII	WHEP	L-Zipper	GST
ClassI aaRSs	MRS				C		N
	VRS						N
	LRS	C					
	IRS	C/C					
	CRS	int/C					N
	RRS					N	
	QRS	N					
	YRS			C			
	WRS				N		
	EPRS				int		N
ClassII aaRSs	SRS	C					
	TRS	N					
	ARS						
	GRS				N		
	HRS				N		
	DRS		N				
	NRS		N				
	KRS		N				
	FRS	(α) N					
AIMPs	1			C		N	
	2					N	C
	3						N

3.1. Not only Glue for tRNAs

3.1.1. tRNA binding UNEs

QRS contains an N-terminal extension of ~200 aa (UNE-Q) and resembles the two adjacent domains determining the specificity of tRNA in the GatB subunit of GatCAB (YqeY domain). In yeast, when the QRS lacks this domain, cells show growth defects and the enzyme has a reduced affinity for $tRNA^{Q}$ (Grant et al., 2012). FRS contains a UNE-F domain at the N-terminus of the α subunit. This domain interacts with the D, T loops and the anticodon stem of the tRNA and its deletion abolishes the aminoacylation activity of FRS (Finarov et al., 2010). UNE-F folds into a known structure that includes 3 DNA-binding fold domains found in many DNA-binding proteins and in double-stranded RNA adenosine deaminase, suggesting that human FRS might have non-canonical functions involving dsDNA/dsRNA binding such as in transcriptional regulations.

3.1.2. Amphipathic helices (DRS, KRS and NRS)

Amphipathic helices are the simplest extension domain in aaRSs (Guo & Yang, 2014). Ranging from 20 to 40 amino acids, these helices contain charged residues on one side and hydrophobic residues on the other (Figure 14A). When positively charged (K and R), the hydrophilic side is ideal to bind negatively charged nucleic acids in a non-specific manner, but it is also interesting to note that the interactions between aspartate (D) and glutamate (E) with RNA have been shown to provide very favorable free binding energies (Lustig et al., 1997). These helices are present in class IIb aaRSs and some aaRS-related proteins such as AIMP1 and AIMP2. Human KRS (Francin & Mirande, 2003), yeast cytosolic DRS (Frugier et al., 2000) *Brugia malayi* cytosolic NRS (Crépin et al., 2011) have a K-rich N-terminal polypeptide extension that promotes tRNA binding and enhances aminoacylation. The lack of specificity of this motif suggest its implication in functions involving the recognition of other types of RNA. This is the case for the yeast DRS which binds its own mRNA and inhibits its expression when $tRNA^D$ is low in the cytosol (Frugier et al., 2005). It has also been proposed that the function of KRS (the mitochondrial form) in HIV packaging depends on this N-terminal helix, presumably because of this RNA binding property (Cen et al., 2004; Kaminska et al., 2007). Interestingly, this same N-terminal helix also interacts with phospholipids and proteins, especially with the transmembrane region of 67LR laminin receptor. The interaction inhibits the ubiquitin-dependent degradation of 67LR thereby enhancing laminin-induced cancer cell migration (D. G. Kim et al., 2012).

3.1.3. EMAPII domains

EMAPII (Endothelial-Monocyte-Activating Polypeptide II) was initially identified in tumor cells as a secreted cytokine (Kao et al., 1992) derived from the protein AIMP1, a component of the metazoan MARS complex (Quevillon et al., 1997). The crystal structure of EMAPII (Y. Kim et al., 2000; Renault et al., 2001) revealed high similarity to bacterial tRNA binding proteins such as *Aquifex aoelicus* Trbp111 (Swairjo et al., 2000) (Figure 14B). The core of these proteins is an oligonucleotide/oligosaccharide-binding (OB) fold composed of five β strands that form an open barrel capped by a short helix α1. Two β strand regions can be distinguished, β1-β2-β3 and β1-β4-β5, both sharing the twisted β1 strand. Homodimerization of Trbp111 is necessary to create a binding site that recognizes the elbow structure of the tRNA molecule (Swairjo et al., 2000). Contrary to Trbp111, EMAPII is a monomeric protein (Quevillon et al., 1997), it contains a C-terminal extension that mimics the dimerization interface of Trbp111, providing a tRNA binding site and preventing dimerization of EMAPII (Renault et al., 2001).

EMAPII is another appended domain that facilitates tRNA binding that is found only at the C-terminus of human YRS and AIMP1. However, in both proteins, EMAPII domains have cytokine activities. This depends on the presence of a cytokine peptide buried in the structure of EMAPII. In the case of AIMP1 EMAPII, there is an heptapeptide "RIGRIVT" localized on the first β strand (Kao et al., 1994; Mirande et al., 2017). As for YRS, under specific conditions, it is secreted and cleaved in two cytokine-active fragments, the free EMAPII domain and the remainder of the protein known as mini-YRS (Wakasugi & Schimmel, 1999). Mini-YRS contains a tripeptide "ELR" cytokine motif that is exposed upon cleavage of the domain EMAPII (S. W. Lee et al., 2004). Interestingly, the motif "ELR" is present only in higher eukaryotes and correlates with the presence of EMAPII (Guo & Yang, 2014). EMAPII domains seems to be restricted to aaRS-related proteins. Besides metazoan AIMP1 and YRS, EMAPII-like domains are also found in other prokaryotic and eukaryotic aaRSs, including *Pyrococcus abyssi* MRS (Crepin et al., 2002), rice MRS (Kaminska et al., 1999) and *Entamoeba hystolytica* KRS and MRS (Castro de Moura et al., 2011).

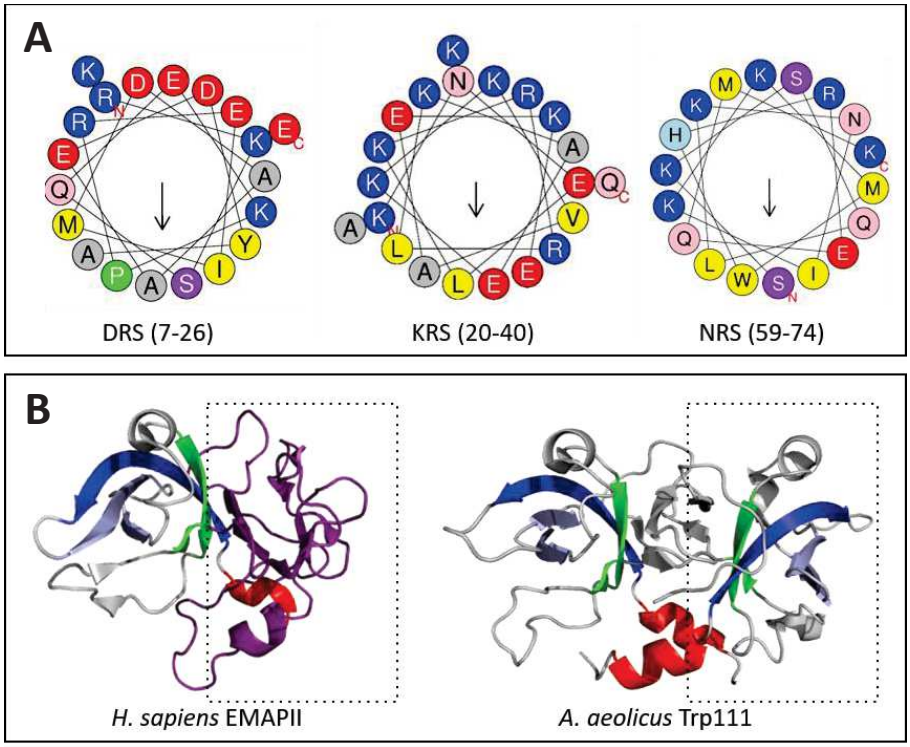


Figure 14. tRNA binding domains. A. Amphipathic N-terminal helices. These helices are found in eukaryotic DRS, KRS and NRS. They are represented using helical wheel projections (HeliQuest). Charged residues are concentrated on one side of the helix while hydrophobic residues are predominant on the other side. **B. EMAPII domain and its prokaryotic homologous Trbp111.** The crystal structure of *A. aeolicus* Trbp111 (1PYB) and the EMAPII domain of *H. sapiens* AIMP1 (1FL0) are shown in similar orientations. Both proteins contain an OB-fold formed by two groups of β-strands, β2-β3 (light blue) and β4-β5 (green), organized around a twisted β1 (dark blue) strand. A small α-helix (red) caps the N-terminus of the OB-fold. Trbp111 is a dimeric protein and the interface of interaction between the two monomers provides the site for tRNA binding. On the other hand, EMAPII is monomeric and contains an extension (purple) that mimics the dimerization interface of Trbp111, providing the interface for tRNA binding. The EMAPII extension and the mimicked monomer of Trbp111 are enclosed with dashed lines.

3.2. More than glue for aaRSs

3.2.1. Protein binding UNEs

IRS contains two large additions at the C-terminus (UNE-I_1 and UNE-I_2). UNE-I_2 exists only in vertebrates, interacts with the WHEP domains of EPRS and therefore may play a role in retaining IRS in the MARS complex (Rho et al., 1996; Rho et al., 1999).

LRS contains a unique domain UNE-L at the C-terminus, which allows interaction with Rag GTPase. Indeed, in human cells, LRS acts as a leucine sensor that activates the mTORC1 complex, a major regulator of cell growth and metabolism. Leucine binds the active site of LRS, promoting the interaction of the enzyme with Rag GTPase, which then activates mTORC1 (Han et al., 2012).

SRS contains a small motif of about 30–40 aa at its C-terminus. This UNE-S includes a nuclear localization signal (NLS) that allows SRS to enter the nucleus, regulate the VEGFA (Vascular Endothelial growth Factor A) through an unknown mechanism, and is essential for vascular development in zebrafish. However, UNE-S has only little effect on the aminoacylation activity of human SRS (Fukui et al., 2009; Xu et al., 2012).

3.2.2. WHEP domains

The WHEP domains exist only in aaRSs' genes. This domain was named in this way because it was initially identified in WRS, HRS and the bifunctional EPRS (Guo et al., 2010). In human, single WHEP modules are found at the N-terminus of WRS, HRS and GRS, at the C-terminus of MRS and as three tandem repeats linking ERS and PRS in the bifunctional enzyme (Guo et al., 2010). However, the distribution and the number of WHEP domains in EPRS may vary between 3 and 6 depending on species. This domain is a 50-amino acids long polypeptide that fold as a simple helix-turn-helix structure, with five conserved K and R residues forming a basic patch on one side of the structure (Figure 15A) (Rho et al. 1998; Cahuzac et al., 2000). This K and R-rich motif suggests that WHEP domains might be non-specific tRNA-binding motifs, though experiments testing this hypothesis have not come to a clear conclusion (Cerini et al., 1991; Wakasugi et al., 2002). However, it has been demonstrated that WHEP domains interact with proteins and other RNAs, then tRNAs. WHEP domains of EPRS interact with (1) the ribosomal protein L13a, (2) the protein NSAP1, (3) the GADPH and (4) the 3′ UTRs of a number of pro-inflammatory mRNAs to form the γ-interferon activated inhibitor of translation (GAIT) complex (Jia et al., 2008; Mukhopadhyay et al., 2009; Arif et al., 2011).

In addition, the WHEP domain of the human WRS regulates its angiostatic activity. Indeed, the crystal structure of human WRS showed that the orientation of the WHEP domain still allows entry of tryptophan and ATP in the active site and does not interfere with aminoacylation. However, it hides the residues involved in the interaction with the extracellular domain of VE-cadherin (Ilyin et al., 2000; Wakasugi et al., 2002; Tzima et al., 2003). Suppression of the WHEP domain, either by proteolysis or alternative splicing, produces a short version of the WRS able to interact with VE-cadherin, which is a surface protein involved in adhesion of endothelial cells.

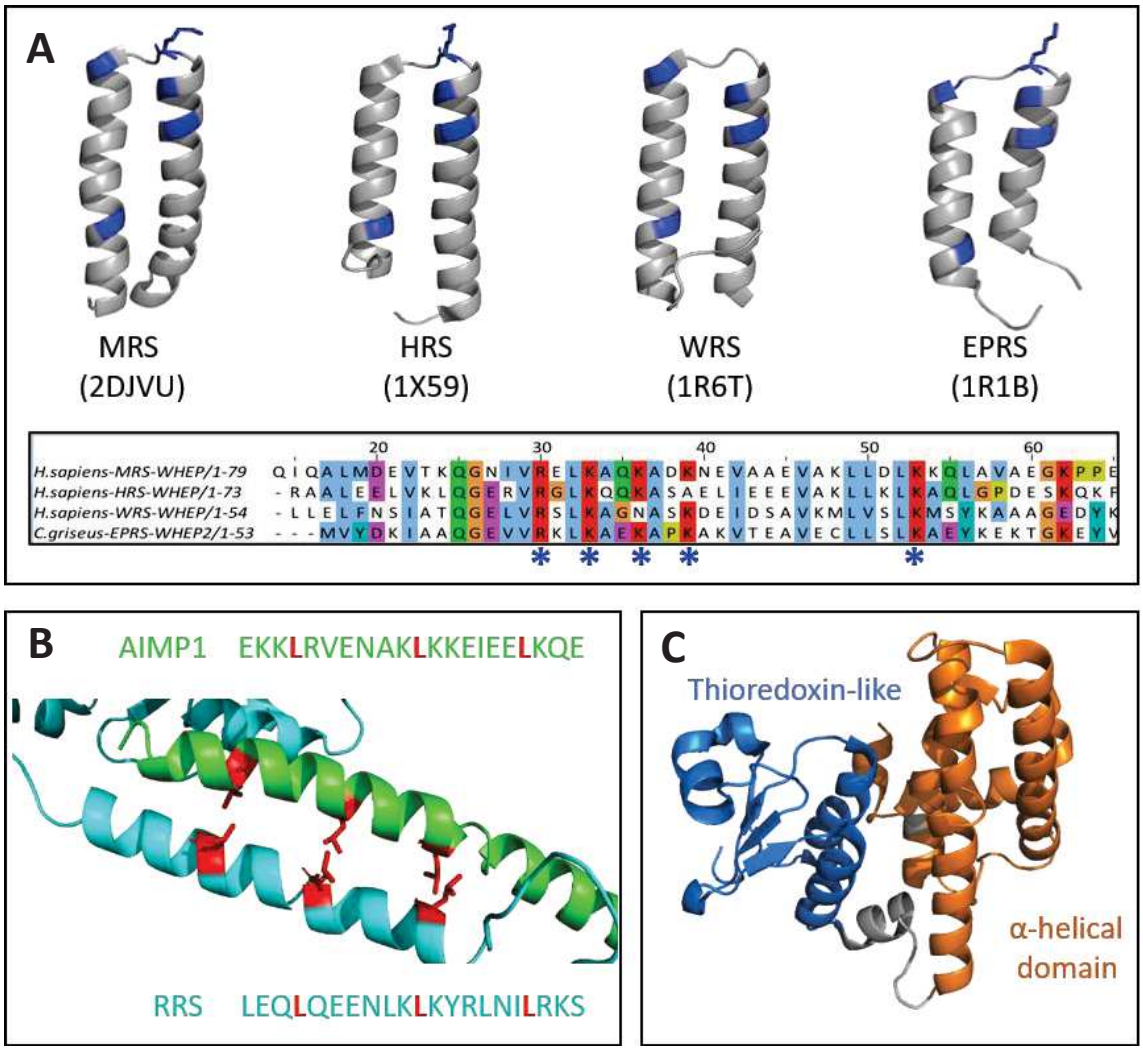


Figure 15. Three types of protein-protein interaction domains. A. WHEP domains. The 3D structure of the WHEP domains from human WRS, HRS, MRS and EPRS from *Cricetulus griseus* (second WHEP repeat). PDB accession numbers are indicated in parenthesis. Conserved K and R residues are colored in blue and marked with asterisks in the corresponding multi sequence alignment at the bottom of the figure. **B. Leucine zipper in the complex AIMP1:RRS.** Hydrophobic interactions in leucine zippers (LZ) are mediated by conserved leucines (L) appearing every fourth position in a repeat of seven amino acids. These leucines are represented as red sticks in the structure of human AIMP1:RRS complex (4R3Z) and are highlighted in the corresponding sequences. **C. Structure of a GST fold.** A glutathione S-transferase (GST) fold is composed of a N-terminal thioredoxin-like moiety (blue) and a C-terminal α-helical domain (orange). The GST of the elongation factor 1B-γ from yeast (1NHY) is shown here.

3.2.3. Leucine zipper

The leucine zipper (LZ) is a long helical domain that usually has several leucine residues with side chains aligned on the same side of the helix (Figure 15B) (Struhl, 1989; Buckland & Wild, 1989). This creates a hydrophobic backbone that connects with its partner to form a coiled coil zipper. These domains are found in many proteins (Rose & Meier, 2004), but in the context of metazoan aaRSs, they are exclusively dedicated to the assembly of the MARS complex. AIMP1, AIMP2 and RRS contain LZs at their N-terminal end (Guo et al., 2010). The LZ of RRS interacts with the LZ of AIMP1, which in turn interacts with the LZ of AIMP2, forming a sub-complex (Robinson et al., 2000; Ahn et al., 2003) (Figure 16).

3.2.4. Glutathione-S-transferase domains

GST domains, as structural modules, are commonly used for protein assembly and protein folding regulation, and many of them have no known enzyme activity. The Glutathione-S-transferase (GST) fold is composed of two parts (Figure 15C): (1) a N-terminal domain adopting a topology similar to that of the thioredoxin fold, consisting of four β-sheets with three flanking α-helices, and (2) a C-terminal domain containing an all-α-helical core structure composed of five or six amphipathic α-helices (Dirr et al., 1994; Sheehan et al., 2001).

Apart from a few exceptions, non-catalytic GST domains are fused to proteins involved in translation: the subunit γ of eukaryotic elongation factor 1 (eEF1-γ) (Koonin et al., 1994), AIMP2, AIMP3 and four class I aaRSs (EPRS, MRS, VRS and CRS). All aaRSs containing GST domains are found in complexes (Figure 16). EPRS, MRS, AIMP2 and AIMP3 are part of the MARS complex (Quevillon et al., 1999; Cho et al., 2015); VRS interacts with the complex eEF1-α/eEF1-β/eEF1-δ/eEf1-γ, (Bec et al., 1989; Bec et al., 1994); a GST-containing CRS is produced by alternative splicing in human and interacts with eEF1-γ (J.E. Kim et al., 2000). In all cases, binding the MARS complex or the translation elongation factors would facilitate tRNA channeling to the ribosome.

4. Exploration of MARS complexes

The most straightforward way to regulate the alternative functions of aaRSs is, perhaps, the assembly of multi aminoacyl-tRNA synthetase (MARS) complexes (Havrylenko & Mirande, 2015). Sequestration inside the complex confines the activity of the enzyme to their aminoacylation function, while they can perform alternative roles upon release. Additionally, the assembly of aaRSs in complexes seems to enhance their aminoacylation activity, particularly when aaRS-interacting multifunctional proteins (AIMP) are incorporated (Cestari et al., 2013), and may facilitate channeling of tRNA to the ribosome. There are three accessory proteins associated to the human MARS complex, which are called AIMP1, AIMP2 and AIMP3. In protozoans, the name of the unique AIMP varies a lot.

4.1. The human MARS complex

The size of the human MARS complex is estimated to be approximately 1.5 MDa and discrepancies between electron microscopy and SAXS studies suggest that it has a significant conformational flexibility (Norcum & Boisset, 2002; Dias et al., 2013). It is composed of nine aaRSs: MRS, DRS, KRS, RRS, LRS, QRS, IRS and EPRS (Mirande et al., 1985), and three AIMPs: AIMP1 (Quevillon et al., 1997), AIMP2 (Quevillon et al., 1999) and AIMP3 (Quevillon & Mirande, 1996) (Figure 16).

Four proteins - MRS, AIMP3, EPRS and AIMP2 - contain GST domains and form a heterotetrameric complex that function as scaffold for other MARS components (Cho et al., 2015). The WHEP domains of EPRS interact with a unique C-terminal extension (UNE-I) in IRS (Rho et al., 1998), which then binds the N-terminal region of LRS (K. Khan et al., 2020). AIMP2 also contains a N-terminal Leucine Zipper (LZ) that allows interaction with both a dimer of KRS (Quevillon et al., 1999; Ofir-Birin et al., 2013) and a portion of the N-terminal LZ of AIMP1 (Ahn et al., 2003). A second portion of the LZ of AIMP1 allows the recruitment of RRS by interacting with its N-terminal LZ (Fu et al., 2014). QRS integrates the complex by interacting with both RRS and AIMP1 *via* its catalytic domain (T. Kim et al., 2000; Fu et al., 2014). The flexible C-terminal EMAPII domain of AIMP1 does not participate in complex assembly and can be removed without disruption of the complex (Shalak et al. 2001). Finally, homodimerization of DRS, PRS and possibly AIMP2 allow the formation of a bisymmetric complex (Mirande, 2017; Hyeon et al., 2019; Cho et al., 2019).

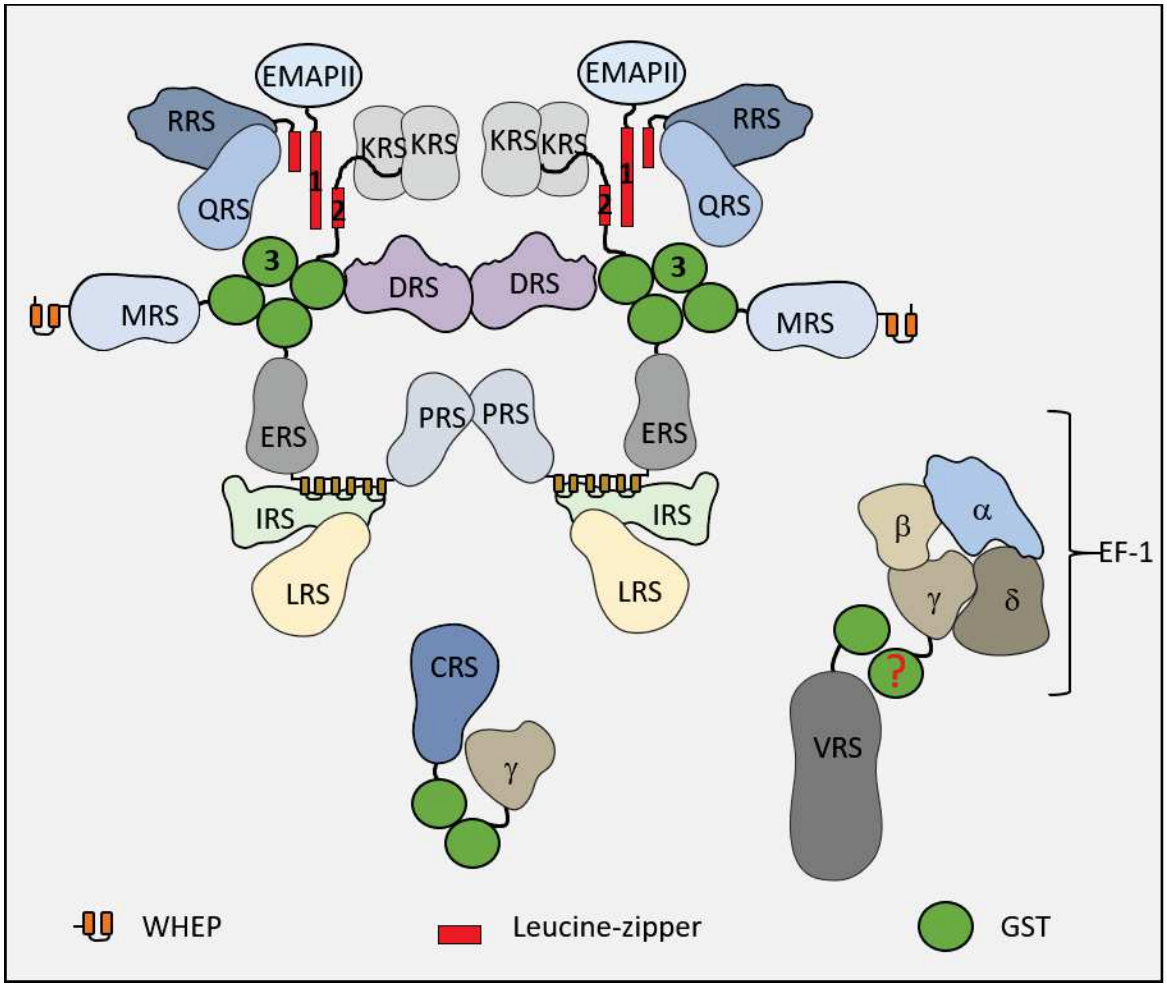


Figure 16: Architecture of human MARS complex. AIMP2 is the component with the largest number of binding partners and is essential for the assembly of the complex (J. Y. Kim et al., 2002). The components in the MARS complex can be grouped in two subcomplexes based on their association with AIMP2 (Robinson et al., 2000; Kaminska et al. 2009). The subcomplex I contains MRS, AIMP3, EPRS, IRS, LRS, KRS and DRS and the subcomplex II is composed of AIMP1, QRS and RRS. Two other small complexes, organized around the γ subunit of the elongation factor 1 (EF1γ), have been also identified. EF1α, EF1β, EF1γ and EF1δ constitute the EF1α GTP exchange factor (Bec et al. 1994; J.E. Kim et al., 2000). This complex promotes the exchange of the bound GDP for GTP to regenerate active EF1α-GTP. Figure adapted from Mirande (2017), Cho et al. (2019), Hyeon et al. (2019) and Khan et al. (2020).

4.2. The yeast MARS complex

In yeast *S. cerevisiae*, MRS and ERS form a complex with the protein Arc1p, an AIMP homologous to human AIMP1/AIMP2 (Simos et al., 1996). Arc1p possess a N-terminal GST domain (Simader, Hothorn & Suck, 2006) and a C-terminal tRNA binding domain similar to an EMAPII-like domain (Giessen et al., 2015). Contrary to classical GSTs, which are dimeric enzymes, Arc1p behaves as a monomer in solution (Golinelli-Cohen & Mirande, 2007; Koehler et al., 2013). Genetic studies suggested that the N-terminal domain of Arc1p is necessary and sufficient to bind simultaneously

the N-terminal GST domains of MRS and ERS (Galani et al., 2001) (Figure 17A). The crystal structure of two binary subcomplexes containing the N-terminal domain of Arc1p in complex with either the N-terminal domain of MRS or the N-terminal domain of ERS suggest a specific mode of assembly where Arc1p and MRS interact like a canonical GST dimer while Arc1p while ERS interact using a novel interface (Simader, Hothorn, Köhler et al., 2006).

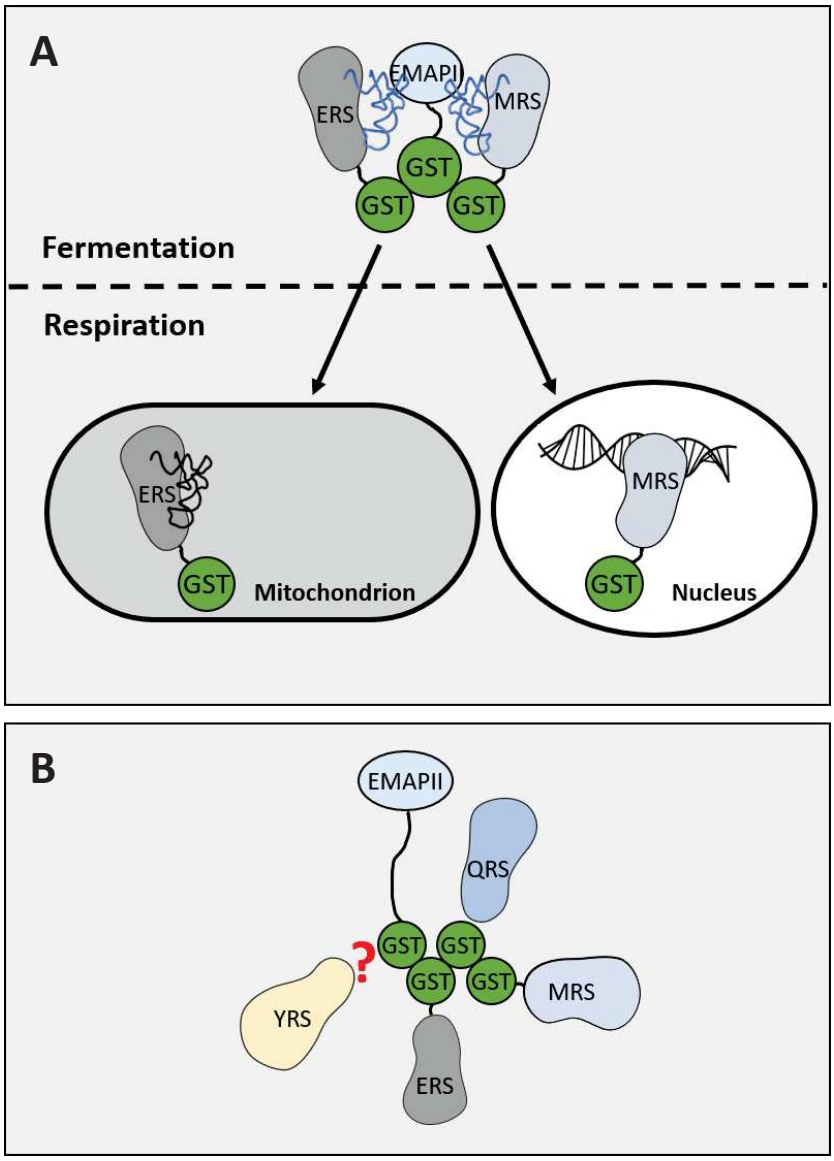


Figure 17: Organization of two protozoan MARS complexes. A. *Saccharomyces cerevisiae* MARS complex. The complex is composed of two aaRSs – ERS and MRS – and the AIMP Arc1p. Assembly occurs through interaction of GST domains appended to the N-terminus of these proteins. In fermenting yeast, Arc1p binds simultaneously ERS and MRS and confine them in the cytoplasm. Upon change to respiration, both ERS and MRS are released and targeted to mitochondria and nucleus, respectively. There, they perform functions to support the respiratory metabolism. **B. *Toxoplasma gondii* MARS complex.** The complex is composed of one AIMP (Tg-p43) and four aaRS (ERS, QRS, MRS and YRS). Except for YRS, complex assembly occurs through interaction of GST domains at the N-terminus of each partner.

Arc1p has two functions in the complex. First, it enhances the tRNA aminoacylation of both ERS and MRS (Simos et al., 1998; Golinelli-Cohen & Mirande, 2007; Graindorge et al., 2005). This function can be replaced by the human protein AIMP1, even if this protein does not interact neither with MRS nor ERS. This suggests that physical interaction with the aaRSs is not necessary to enhance tRNA aminoacylation. It has been suggested that Arc1p and AIMP1 may be involved in sequestering tRNA in order to increase its local concentration in the cytoplasm (Golinelli-Cohen et al., 2004). Second, Arc1p confines both the MRS and ERS in the cytoplasm. As the yeast switches from fermentation to respiration, the expression of Arc1p is down-regulated, causing the release of both MRS and ERS from the complex (Frechin et al., 2014); ERS is then partially relocated to the mitochondria where it synthesizes glutaminyl-tRNA^{Gln} *via* the GatFAB-dependent transamidation pathway and thus boosts the translation of the mitochondria-encoded F_1F_0ATP synthase complex (Frechin et al., 2009); MRS is translocated to the nucleus where it regulates the transcription of some nuclear-encoded oxidative phosphorylation genes, which are exported to the mitochondria (Frechin et al., 2014).

4.3. The *Toxoplasma gondii* MARS complex

Toxoplama gondii is an intracellular parasite closely related to *Plasmodium*. Similar to yeast, it contains a single AIMP called Tg-p43, which also contains a N-terminal GST domain and a C-terminal EMAPII-like domain. Using Tg-p43 as bait, a cytoplasmic complex containing MRS, ERS, QRS and YRS was identified (van Rooyen et al., 2014). With the exception of YRS, all these aaRSs contain GST domains appended to their N-terminus (Figure 17B). Deletion of Tg-p43 was not lethal for the parasite and it did not affect its pathogenicity neither. *In vitro*, the GST domain of Tg-p43 was sufficient to form a complex. The sample was significantly heterogeneous and initial electron microscopy imaging of the complex suggested a large degree of flexibility of the particle around a central ring-like core.

III. Objectives of my PhD

Among the multiple exchanges occurring in *Plasmodium*, import of exogenous tRNAs in sporozoites is probably the most intriguing. Although the function of these imported tRNAs is unknown, the transporter involved in their internalization has been identified (Bour et al., 2016). *In vitro*, tRNA import is dependent on the protein tRip. *In vivo*, tRip is an integral membrane protein and its tRNA binding domain is exposed at the surface of the parasite. Deletion of tRip is not lethal for the parasite, but it reduces significantly the development of the parasite in the blood.

Apart from its involvement in importing exogenous tRNAs into the parasite, tRip is homologous to AIMPs. *Plasmodium* is no exception and Dr. D. Kapps, a previous PhD student in the lab, has identified three aaRSs that interact with tRip *in vivo*: glutamyl-tRNA synthetase (ERS), glutaminyl-tRNA synthetase (QRS) and methionyl-tRNA synthetase (MRS) (Figure 18). However, the membrane localization of tRip as well as the external localization of its tRNA binding domain (tRBD) raises many questions about the organization of this MARS complex and its function in the synthesis of the parasite's proteins.

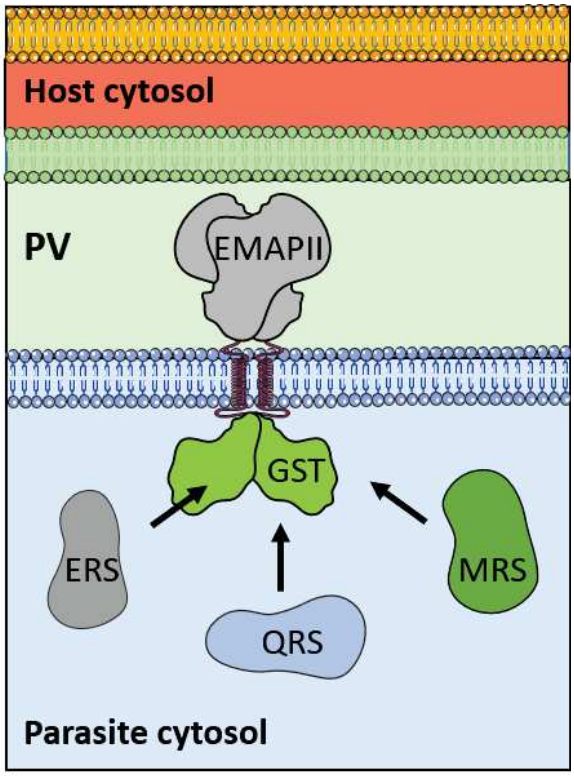


Figure 18: tRip interactome. Schematic representation of the proteins identified in tRip co-immunoprecipitation experiments. tRip is anchored in the parasite plasma membrane, the N-terminal GST domain (green) being inside the parasite and the C-terminal domain (tRNA binding domain homologous to EMAP-II, grey) being outside. *In vitro*, the recombinant tRip forms a dimer (Bour et al., 2016).

tRip is a protein with two functional domains: an EMAPII-like tRNA binding domain at the C-terminal end and a GST domain at the N-terminal end. The two domains are linked by a seemingly disordered linker. Using a purified antibody specifically directed against the tRBD domain, the three *Plasmodium* aaRSs were identified as specific tRip interactors. Analysis of the sequences of ERS, QRS and MRS revealed a particular modular architecture. Like the majority of *Plasmodium* proteins, ERS, QRS and MRS also contain LCRs. Moreover, these aaRSs encompass additional domains present at both ends. In the N-terminal part of the three aaRS, I have identified GST domains, suggesting that the interaction between tRip and the different aaRS is done through these GST domains. Two of these aaRSs are also characterized by the presence of additional domains at their C-terminal end. The MRS contains another EMAPII-like domain while the QRS contains a domain that has no obvious homology. The tRNA binding capacity of these C-terminal extensions was studied during this work.

I focused on the reconstitution of the *Plasmodium* MARS complex using GST domains to study their interactions and deduce the organization of the four proteins in the complex. Numerous constructions of each GST domain have been designed and produced. I tested the interaction capacity of the four GST domains by (i) pull-down, (ii) co-purification and analysis of the complex composition by mass spectrometry, (iii) stoichiometry determination and competition experiments, (iv) tRNA binding properties, (v) DLS/SLS (dynamic/static light scattering) measurements and (vi) SEC-SAXS (size-exclusion chromatography- coupled to small-angle X-ray scattering).

In addition, numerous crystallogenesis tests were performed on many different combinations of proteins to determine their structure by crystallography. Among all the crystallogenesis experiments I have done, only the GST domain of ERS has crystallized and gave us precious information to design point mutations and identify the interaction interfaces involved in the complex organization.

RESULTS

I. Identification of *Pb*tRip protein partners

(Unpublished results from the PhD work of Dr. Delphine Kapps)

During her PhD work, Dr. Delphine Kapps identified the *Pb*tRip interactome in *Plasmodium berghei* by co-immunoprecipitation (Co-IP) and mass spectrometry (unpublished results). Parasites were isolated from the blood of infected mice. Red blood cells and the parasitophorous vacuole were lysed to recover only intact parasites. Endogenous *Pb*tRip was immunoprecipitated with a specific antibody raised against the extracellular C-terminal tRNA binding domain (tRBD) of its *P. falciparum* homologue (Figure 19A) and in the presence of triton. Since this domain is localized outside the parasite, binding of the antibody does not interfere with the interactions that occur with the N-terminal intracellular domain of *Pb*tRip (GST domain).

Proteins co-purified with tRip represent potential partners and were identified by mass spectrometry. This analysis was performed in three biological replicates, each using the tRip-KO parasite as a negative control. After subtraction of background interactions (tRip-KO samples), only four proteins were significantly enriched (Figure 19B). Apart from *Pb*tRip, one candidate, the *Pb*ERS, was particularly abundant, since it was identified with similar number of spectral counts than *Pb*tRip. The two other potential partners were less abundant and corresponded to two others cytosolic aaRSs (*Pb*QRS and *Pb*MRS).

The composition of the *P. berghei* MARS complex is comparable to that observed in *Toxoplasma gondii*, which contains the AIMP Tg-p43 and 4 aaRSs: ERS, QRS, MRS and YRS (van Rooyen et al., 2014). However, no trace of *P. berghei* YRS was observed in the Co-IP data. Sequence comparisons revealed that *Toxoplasma* YRS contains a N-terminal extension (absent in *Plasmodium* YRS), which might facilitate the incorporation of YRS into the *T. gondii* MARS complex.

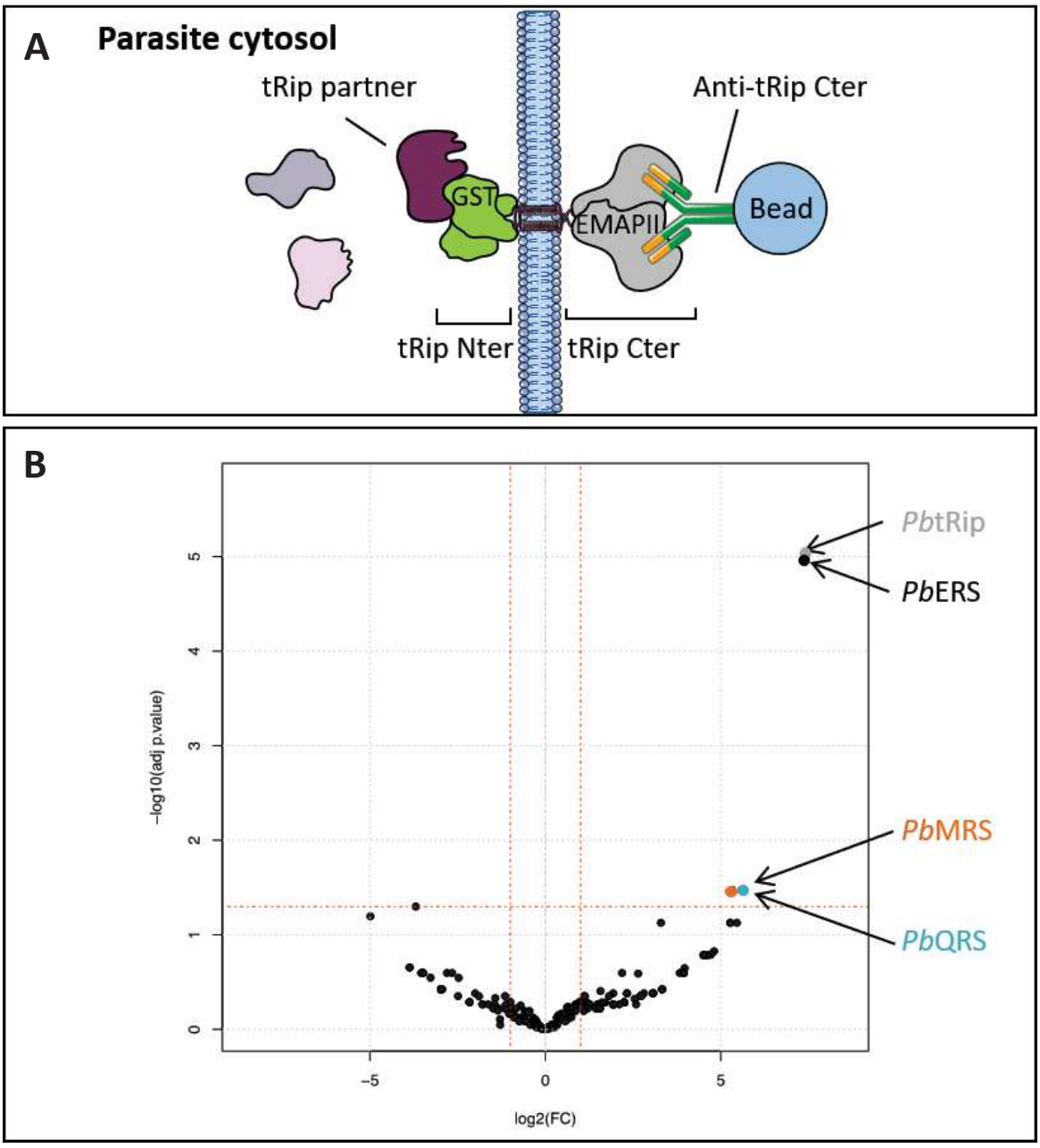


Figure 19. Identification of the *Pb*tRip interactome in the *P. berghei* blood stage. A. Design of the co-immunoprecipitation experiment. *Pb*tRip and its partners were extracted from blood-stage parasites using a purified and specific antibody raised against the extracellular C-terminal tRNA binding domain of *P. falciparum* tRip (Bour et al., 2016). **B. Volcano plot visualization of the *Pb*tRip interactome.** CoIP and mass spectrometry analysis (LC–MS/MS) of *Pb*tRip partners were performed in three biological replicates. A total of 229 *P. berghei* proteins were identified in the 6 compared samples (3 WT and 3 KO). As expected, *Pb*tRip has been identified only in the WT samples. After subtraction of background interactions and protein frequency assessment, we could identify only 3 proteins considered statistically significant out of the 229 proteins, with an adjusted p-value < 0.05, a minimum of 5 spectral counts and a LFC of at least 1. The horizontal dashed red line indicates p-value = 0.05 with points above having p-value < 0.05. Vertical dashed red lines show LFC = 1.

II. Bioinformatics

1. Bioinformatic analysis pipeline

Sequences of *Pb*tRip, *Pb*ERS, *Pb*QRS and *Pb*MRS were analyzed using different bioinformatics tools. First, similar proteins from model organisms were identified by BLAST searches using *P. berghei* sequences. In addition to the identification of similar proteins, BLAST analysis also provided an overview of the evolutionary context of the proteins. Identified proteins were then aligned with other sequences from different *Plasmodium* species to identify conserved functional domains and specific additional domains that characterize each enzyme. Two strategies of multi-sequence alignment (MSA) were used to analyze the proteins of the *Pb*MARS complex. (1) Sequences from different *Plasmodium* species were aligned to assess their level of conservation. With *Plasmodium* proteins, the first thing that can be noted is the presence of gaps in the aligned sequences. These gaps correspond to the position of low-complexity regions (LCRs), whose length and composition is variable depending on the *Plasmodium* species and even strains, but their localization is conserved within homologous proteins. The longest LCRs are observed and are mostly composed of asparagine repeats in *P. falciparum* proteins. LCRs are shorter and less rich in asparagine residues in other *Plasmodium* species, especially in *P. berghei*. This is why we made the choice to work with this specific strain. (2) The sequences from several model organisms were included in order to identify and delimitate conserved modules in the plasmodial proteins. Sequences from bacterial aaRSs were useful to identify the limits between the aaRS core and the *Plasmodium*-specific extensions and sequences from well-known eukaryotic aaRSs were helpful to establish the nature of some of these extensions. Finally, predictions of secondary and tertiary structures were used to propose a model of the modular architecture of each protein.

2. Evolutionary relationships

The query coverage of BLAST alignments was variable among the 4 proteins analyzed. Higher coverages were obtained for *Pb*tRip (74%) and *Pb*ERS (83%) whereas, *Pb*QRS and *Pb*MRS exhibited lower query coverages (70% and 60%, respectively). Moreover, among the 3 aaRSs, *Pb*MRS showed no more than 38% identity with enzymes from other organisms, and the best matches for *Pb*ERS and *Pb*QRS presented 45 to 50% identity, suggesting that these aaRSs are quite different from their homologues.

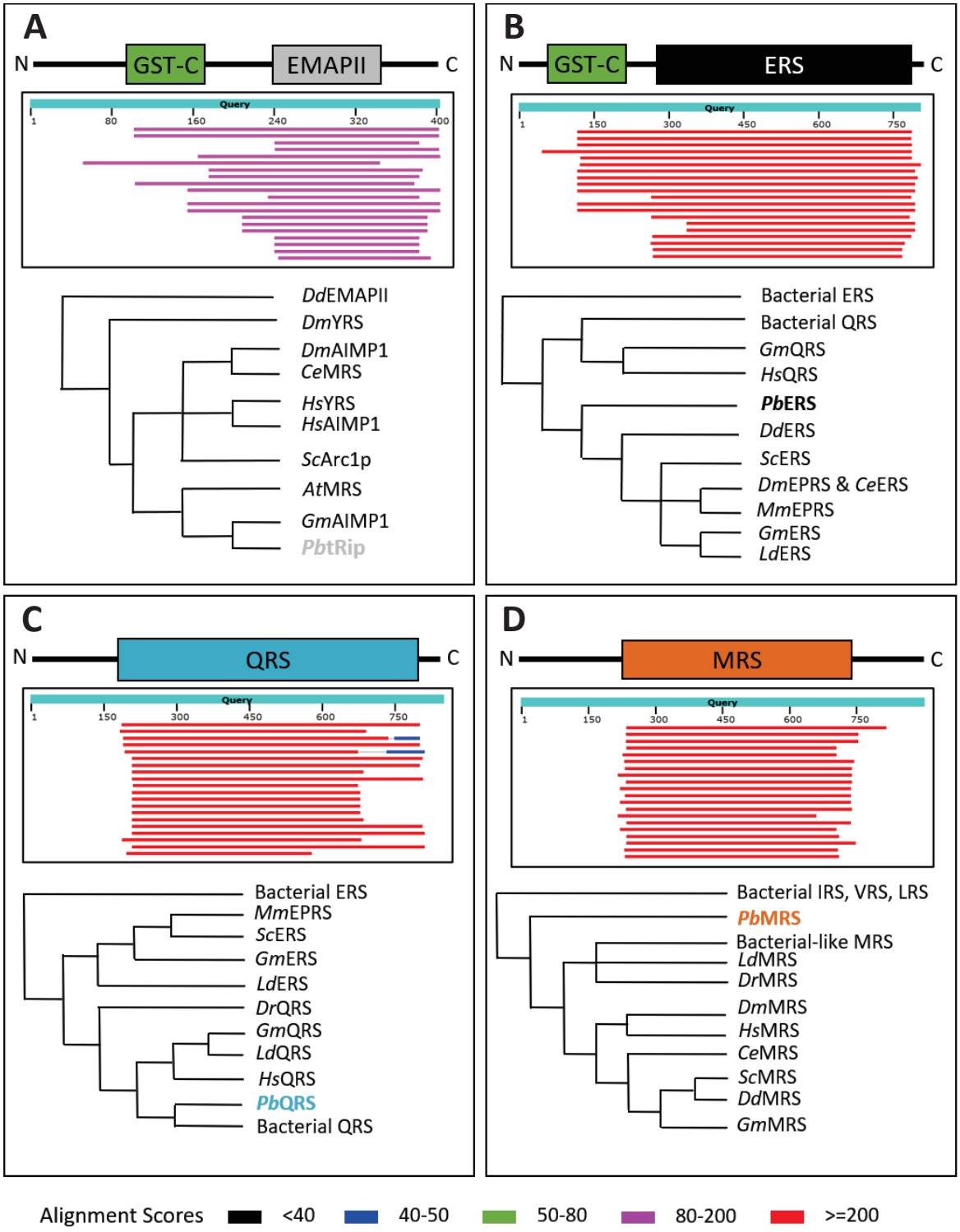


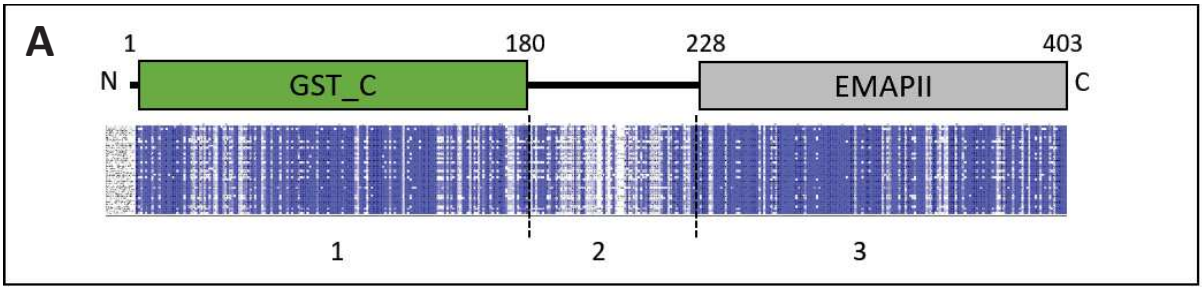
Figure 20. BLAST analysis of proteins of the *Pb*MARS complex. The sequences of *Pb*tRip **(A)**, *Pb*ERS **(B)**, *Pb*QRS **(C)** and *Pb*MRS **(D)** were used to find similar proteins in the landmark database of model organisms. For each protein, two types of results are shown (i) a graphic summary showing the domains identified in the query sequence and the extent of alignments and (ii) a distance tree clustering the hits according to their similarity to the query sequence. *Pb: Plasmodium berghei, Dd*: *Dictyostelium discoideum*, *Dm*: *Drosophila melanogaster*, *Ce*: *Caenorhabditis elegans*, *Hs*: *Homo sapiens*, *Sc*: *Saccharomyces cerevisiae*, *At*: *Arabidopsis thaliana*, *Mm*: Mus musculus, *Dr: Deinococcus radiodurans*, *Ld*: *Leishmania donovani* and *Gm*: *Glycine max.*

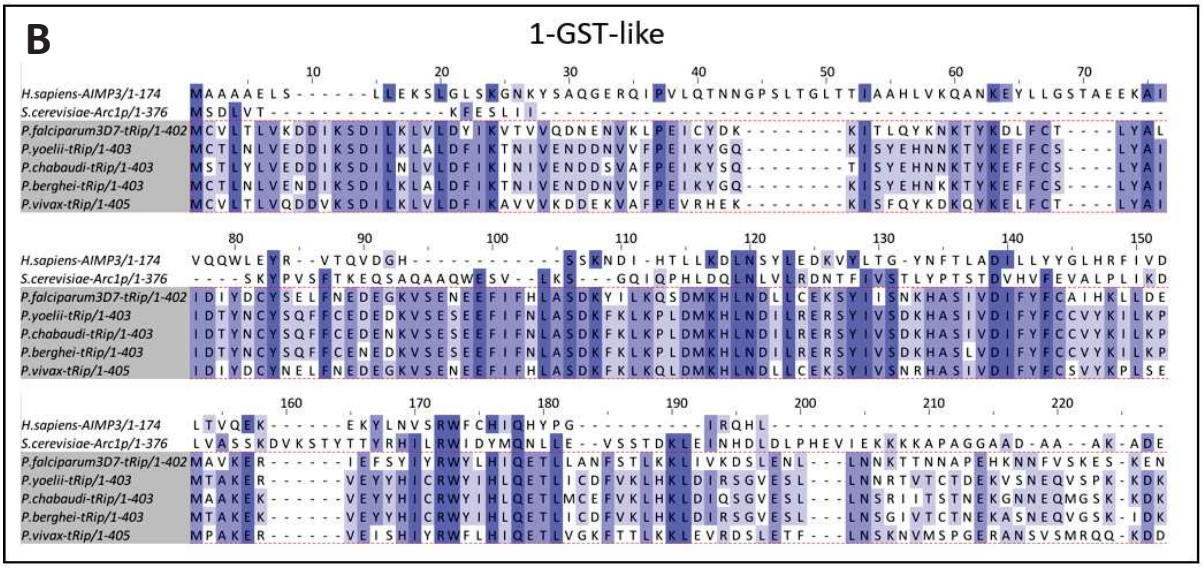
BLAST search using *Pb*tRip produced pairwise alignments with moderate scores (Figure 20A). Consistent with previous analysis (Bour et al., 2016; Kapps et al., 2016), no bacterial proteins were found in this search, reflecting the fact that *Pb*tRip closest homologous proteins are restricted to eukaryotes. All hits aligned primarily with the C-terminal part of *Pb*tRip, which contains the extracellular tRBD, and thus confirmed that *Pb*tRip is an EMAPII-containing protein. Among these homologous proteins, not only AIMPs but also metazoan YRSs (*D. melanogaster* and *H. sapiens*) and MRSs from nematodes and plants (*C. elegans* and *A. thaliana*) were retrieved, because these enzymes contain EMAPII domains appended to their C-terminus (Kaminska et al. 1999; Crépin et al., 2002; Havrylenko et al., 2010). Interestingly, the distance tree indicated that *Pb*tRip is more related to EMAPII-containing proteins from plants (*G. max* AIMP1 *and A. thaliana* MRS) and from *S. cerevisiae*, all of them containing a N-terminal GST domain, indicating that *Pb*tRip is a GST-EMAPII fusion protein like *S. cerevisiae* Arc1p and *T. gondii* Tg-p43.

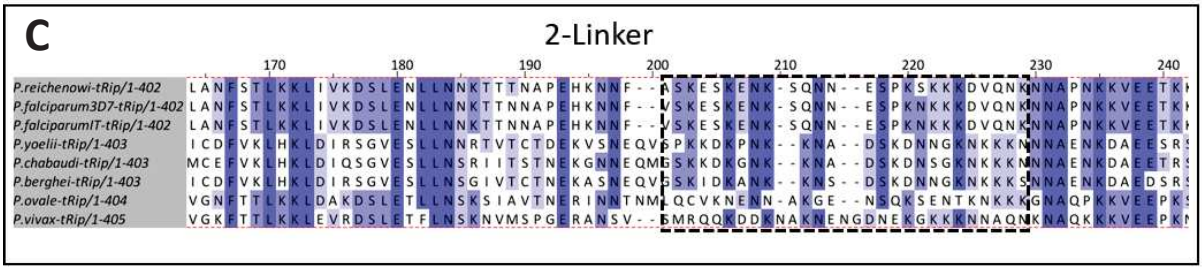
The BLAST search using *Pb*ERS as query sequence produced high alignment scores and retrieved results including both ERS and QRS enzymes (Figure 20B). This is not surprising since ERS and QRS share a common evolutionary history (Lamour et al., 1994; Siatecka et al., 1998; Brown & Doolittle, 1999; Hadd & Perona, 2014). Consistent with this model, *Pb*ERS is separated from bacteria homologues and clusters with both eukaryotic ERS and QRS enzymes in the distance tree. The best matches for *Pb*ERS were ERSs from unicellular eukaryotes (*D. discoideum* and *S. cerevisiae*). This is coherent with the identification of a GST domain fused to the *Pb*ERS N-terminus, since this feature is conserved in all eukaryotic ERSs (except in Kinetoplastida) (Gowri et al., 2012; Cestari et al. 2013).

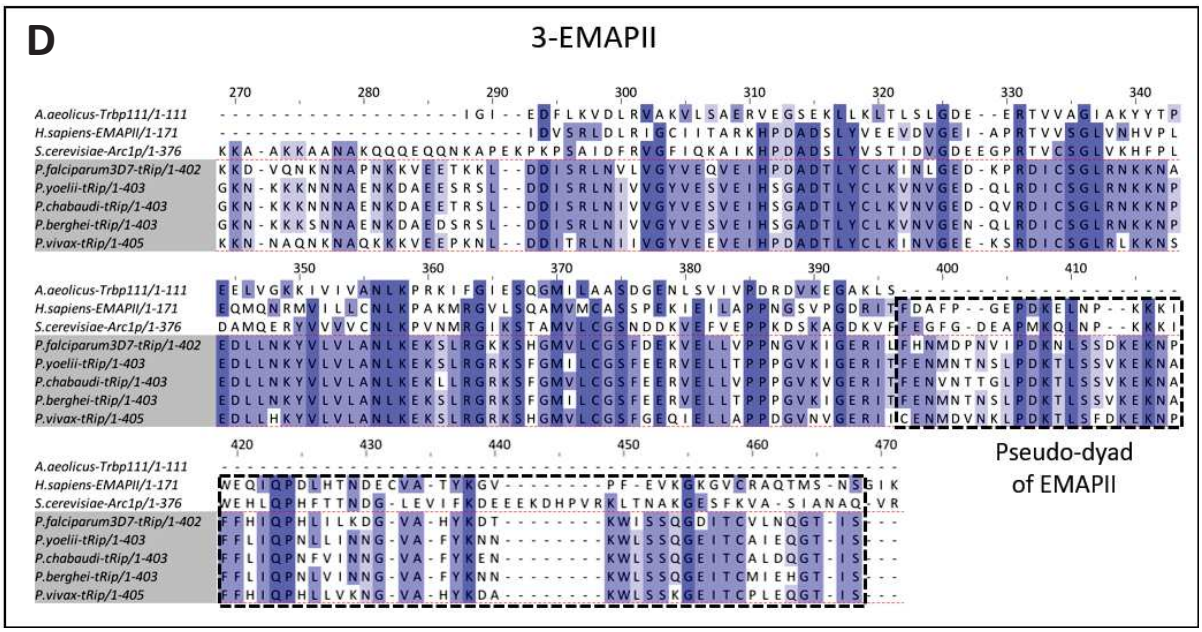
As with *Pb*ERS, the BLAST search with *Pb*QRS as query sequence produced results including both QRS and ERS enzymes (Figure 20C), but in this case *Pb*QRS unambiguously clustered with QRS enzymes, with the best matches being bacterial QRSs, indicating that the additional sequences found at the N- and C-terminals of *Pb*QRS could not be attributed to any additional domain already known in other eukaryotic QRSs. As of today, the only domain specifically appended to QRSs is the YqeY domain (also known as UNE-Q) that is found either at the N-terminus of eukaryotic QRSs (Hadd & Perona, 2014) or at the C-terminus of rare bacterial QRSs like in *Deinococcus radiodurans* (Deniziak et al., 2007).

*Pb*MRS was the most divergent protein in this analysis. It occupied its own branch in the distance tree and was separated from all eukaryotic MRSs (Figure 20D). However, the three other class Ia aaRSs (IRS, VRS and LRS) as well as bacterial-like MRSs, including *T. maritima* MRS and several MRSs from chloroplasts and mitochondria stood out from other hits.









Legend in next page.

Figure 21: Bioinformatic analysis of *Pb*tRip. A. Conservation of tRip in *Plasmodium* species. Schematic representation of tRip highlighting the different domains identified by BLAST and MSA. The MSA graphical summary includes up to 43 sequences from different *Plasmodium* species and *P. falciparum* strains. Blue-colored columns represent conserved residues. **B. Identification of a N-terminal GST domain.** Sequences of two eukaryotic GST domains are aligned with a selection of *Plasmodium* sequences of the tRip N-terminus (highlighted in grey). **C. A linker with RNA binding capacities.** Sequences rich in charged amino acids are framed. **D. Identification of an EMAP-II-like domain.** The C-terminal domain of *Plasmodium* tRip only aligned with proteins containing EMAPII-like domains at their C-terminus.

Once again, significant portion of the N- and C-terminal regions of *Pb*MRS were not covered by any of the hits, suggesting the presence of *Plasmodium*-specific extensions. Indeed, the structural diversity of MRSs is a landmark feature (Deniziak & Barciszewski, 2001) and several additional modules can be identified at their N- and C-terminals, such as WHEP, EMAPII, GST or Trbp111-like domains (Kaminska et al., 1999).

3. Analysis of structural features

3.1. More precise information about tRip

In Bour et al. (2016), using sequences of *S. cerevisiae* Arc1p, *H. sapiens* AIMP1 and *A. aeolicus* Trbp111, authors identified a N-terminal domain with no homology and a C-terminal domain homologous to EMAPII separated by a linker. Although these previous MSA provided some clues, full identification of specific domains was achieved only when more sequences were submitted to the NCBI CD search service in batch mode (Figure 21A). With this approach, I could confirm that the sequence of the linker located between the N-terminal GST domain and the C-terminal EMAPII-like domain is not a LCR (Figure 21C). Indeed, even if the sequence of this insertion varies among the different *Plasmodium* species, its size stays about the same. Likewise, the absence of long asparagine repeats in the *P. falciparum* sequence supports this idea. In addition, the linker of tRip contains conserved stretches of arginine (R) and lysine (K), essential to support the high affinity of tRip for tRNAs. Moreover, I could show that the N-terminal part of tRip aligned with *H. sapiens* AIMP3 and the N-terminus of *S. cerevisiae* Arc1p, which are two well-characterized GST domains (Simader, Hothorn & Suck, 2006; K. J. Kim et al., 2008) (Figure 21B) and that the C-terminal tRNA binding domain contains the characteristic eukaryotic pseudo-dyad of EMPAII (Figure 21D). The EMAPII-like domain of tRip from different *Plasmodium* species was compared with similar domains from other organisms, including *A. aeolicus* Trbp111, *S. cerevisiae* Arc1p and several EMAPII-like with cytokine properties (Figure 21D). All these domains possess non-specific tRNA binding properties.

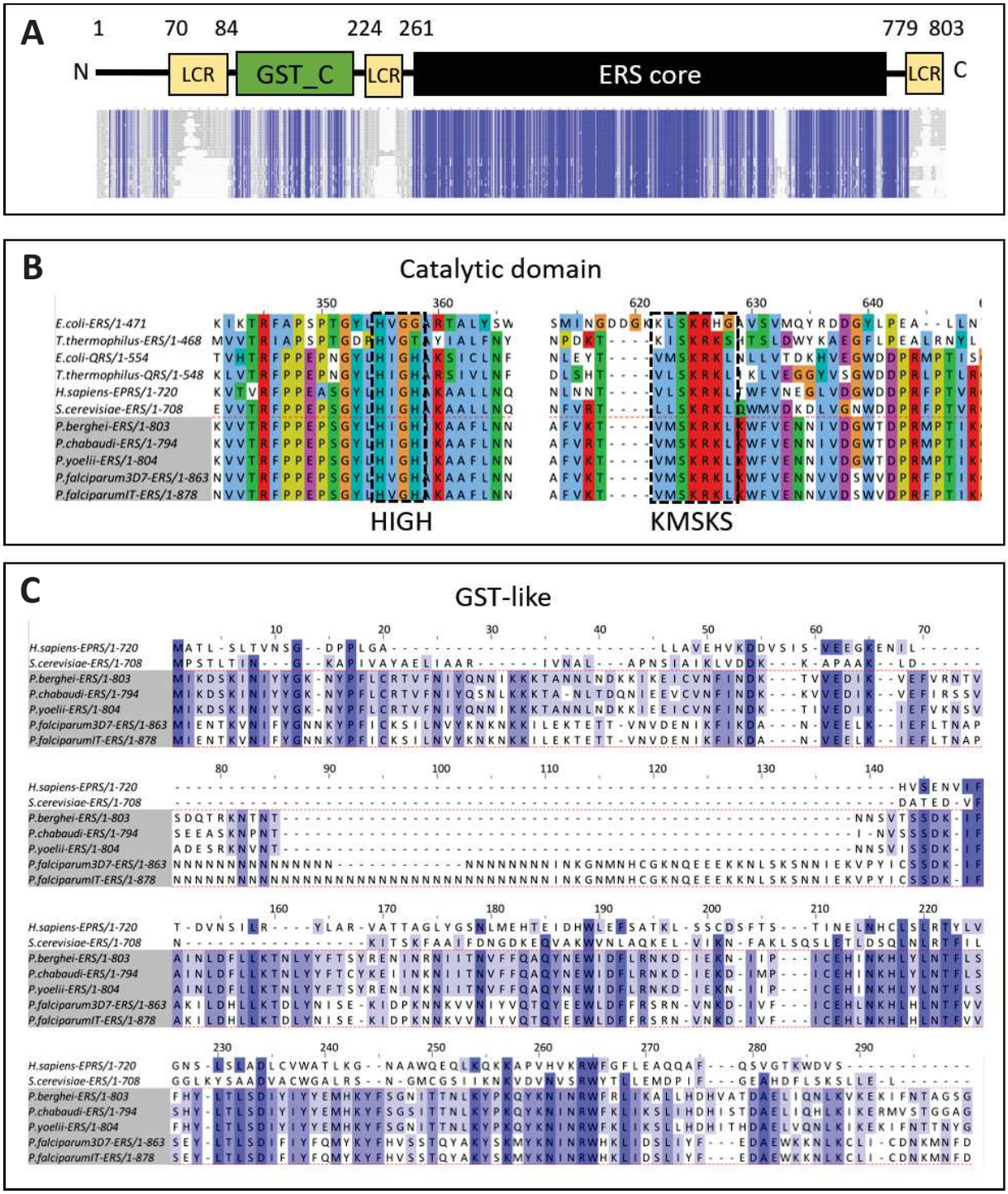


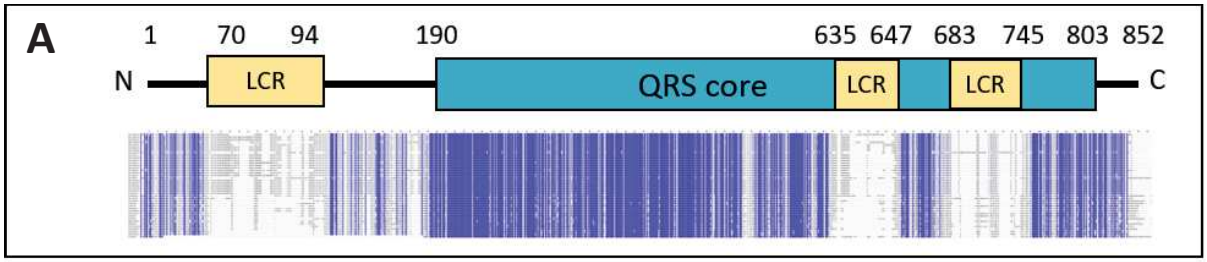
Figure 22: Bioinformatic analysis of *Pb*ERS. A. Conservation of ERS in *Plasmodium* species. Schematic representation of *Plasmodium* ERS highlighting the different domains identified by BLAST and MSA. The MSA graphical summary includes up to 45 sequences from different *Plasmodium* species and *P. falciparum* strains. Blue-colored columns represent conserved residues. **B. Catalytic domain of *Plasmodium* ERSs.** Residues are colored according to Clustal X color code. The two conserved signatures HIGH and KMSKS of class I aaRSs are framed. **C. Identification of a N-terminal GST domain.** Sequences of two GST domains from eukaryotic ERSs are aligned with a selection of N-terminal domains of *Plasmodium* ERSs (highlighted in grey).

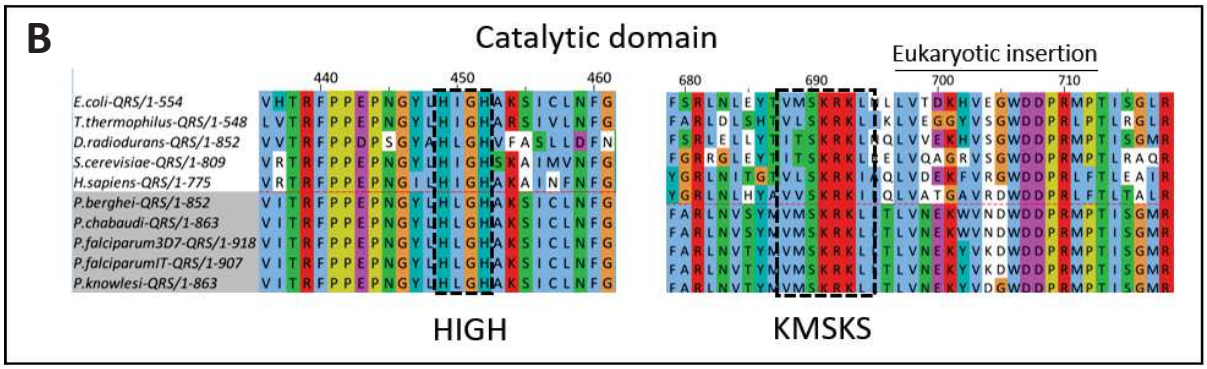
In the case of Trbp111, dimerization provides the interface for tRNA binding. Based on mutational analysis, Swairjo et al. (2000) proposed a model for the interaction. Trbp111 binds the elbow of tRNAs and important residues for the interaction include K33, E45, N71, R75 and especially S82 and M85. Critical residues are strictly conserved in tRip and all the other EMAPII domains, suggesting a similar mode of tRNA binding. Unlike Trbp111, the EMAPII of tRip and other EMAPII are monomers in solution (Renault et al., 2001; unpublished data from the team). They possess a C-terminal extension (pseudo-dyad) that creates the interface for tRNA binding without dimerization.

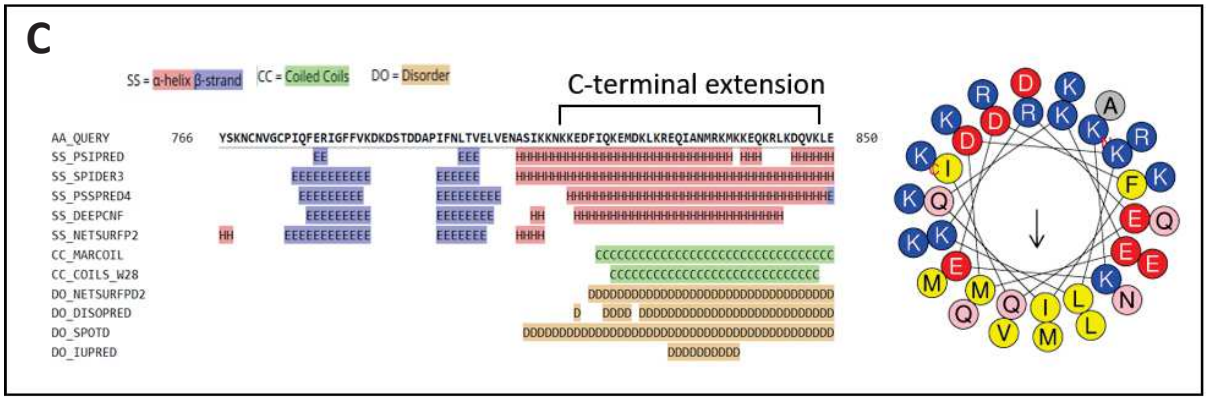
It is known that cytokine activities of EMAPII domains depend upon a heptapeptide sequence called "migration motif". In EMAPII domains derived from human AIMP1 and YRS, the sequence of the migration motif is "RIGCIIT" (Kao et al., 1994) and "RVGKIIT" (Wakasugi & Schimmel, 1999), respectively. This motif is missing in tRip. Although cytokine properties of EMAPII are generally restricted to mammalian systems, the parasite *Entamoeba histolytica* is an exception to the rule. Two aaRSs from this organism, KRS and MRS, possess appended EMAPII domains that can be processed to generate polypeptides with cytokine activity (Castro de Moura et al., 2011).

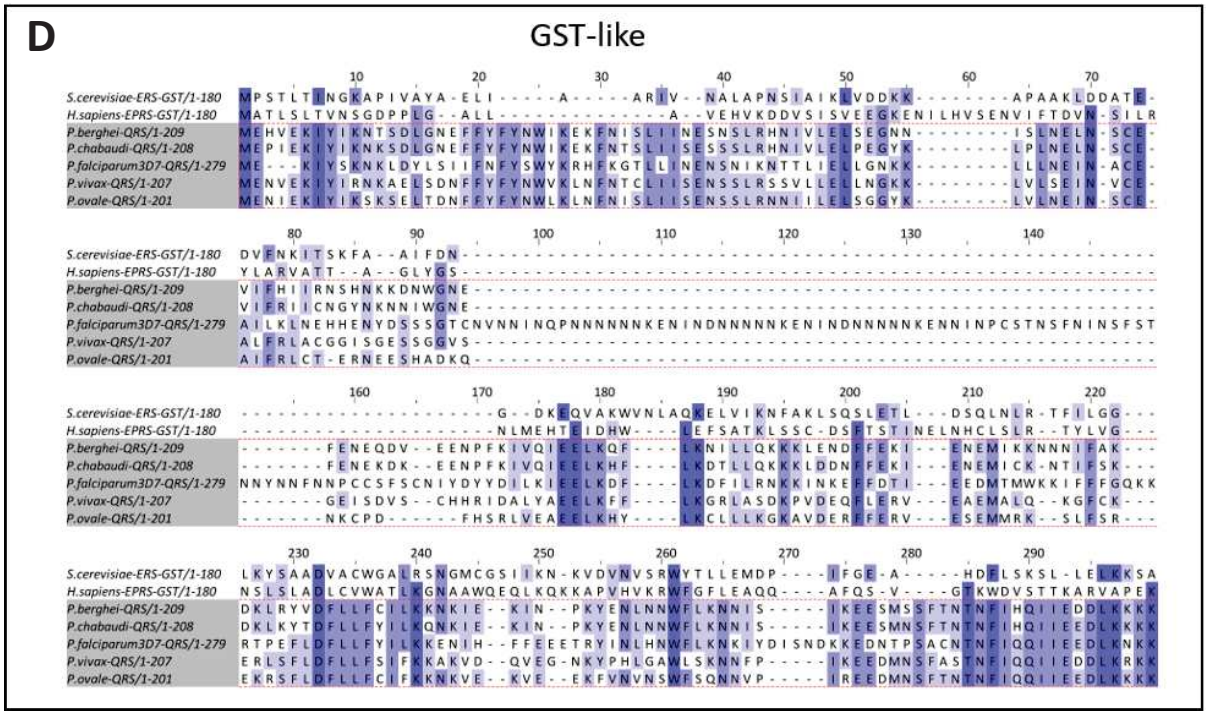
3.2. Nothing really new about *Pb*ERS

*Pb*ERS contains 3 well-delineated LCRs (Figure 22A). Two of them are localized in the N-terminal extension of the protein and the third one is located at the C-terminus of the enzyme. Like in the two other aaRSs, the most conserved part of the protein corresponded to the functional domain aaRS core, which correspond to the catalytic domain (CD) followed by the anticodon binding domain (ABD). The two signature-sequences characteristic of class I aaRSs are well conserved in *Plasmodium* ERSs (Figure 22B). The motif HIGH is almost invariant and the KMSKS loop has a sequence of type "VMSKR", which is characteristic of enzymes from the eukaryotic/archaeal lineage (Sekine et al., 2003). *Plasmodium* ERSs share the same anticodon binding domain topology than other eukaryote organisms. Moreover, all the residues involved in the recognition of the tRNA, either the CCA-end or the anticodon sequence, are also conserved (data not shown). Blast analysis and MSA identified the presence of a GST domain in *Pb*ERS (Figure 22C). BLAST analysis indicated similarities only with the C-terminal half of the GST domain probably because the N-terminal sequences are less conserved. Indeed, the C-terminal part of the GST domain of *Plasmodium* ERS aligned well with GST domains from of *H. sapiens* EPRS and *S. cerevisiae* ERS (Simader, Hothorn & Suck, 2006; Cho et al., 2015), the plasmodial protein being more similar to its human counterparts.









Legend in next page.

Figure 23: Bioinformatic analysis of *Pb*QRS. A. Conservation of QRS in *Plasmodium* species. Schematic representation of QRS highlighting the different domains identified by BLAST and MSA. The MSA graphical summary includes up to 45 sequences from different *Plasmodium* species and *P. falciparum* strains. Blue-colored columns represent conserved residues. **B. Catalytic domain of *Plasmodium* QRS.** Residues are colored according to Clustal X color code. The two conserved signatures HIGH and KMSKS of class I aaRSs are framed. **C. Prediction of secondary structure of the C-terminal sequence.** The secondary structure of the last 50 amino acids of *Pb*QRS was predicted using the tool Quick2D. Although a disordered structure is also possible, this extension has potential to form α-helices. Moreover, the helical wheel projection (https://heliquest.ipmc.cnrs.fr/) corresponding to the last 40 amino acids shows clearly that one side of the predicted helix is positively charged. **D. Identification of a N-terminal GST domain.** Sequences of two GSTs from eukaryotic ERSs are aligned with a selection of *Plasmodium* sequences of QRS N-terminal domains (highlighted in grey), revealing the sequence variations and the presence of insertions that might be responsible for the difficulty to identify a GST domain.

3.3. A *Pb*QRS with unsuspected additional domains

Plasmodium QRS displays 3 well-defined LCRs (Figure 23A). The first one is found in the middle of the N-terminal extension and the 2 others inside the QRS ABD, which explains why many BLAST hits did not align well with this part of *Plasmodium* QRS. As for *Plasmodium* ERS, the two signature sequences HIGH and KMSKS (of type "VMSKR") are present in the catalytic site (Figure 23B) and the residues that specifically recognize the tRNA are also conserved (not shown). Interestingly, the two LCRs divide the ABD in three blocks, each containing essential residues involved in the specific recognition of the tRNA anticodon.

A small extension was identified at the C-terminus of the enzyme. It is ~100 residue-long and did not aligned with any other QRSs in the MSA. This extension is thus *Plasmodium*-specific since it is highly conserved among *Plasmodium* species. Interestingly, it contains several positively charged residues and has the potential to form helices, suggesting RNA binding properties (Figure 23C). However, this C-terminal extension is too short and too different to be a YqeY domain similar to those fused at the C-terminus of bacterial *D. radiodurans* QRSs (Deniziak et al., 2007).

The N-terminal extension of *Plasmodium* QRS is about 200 amino acids-long. No convincing alignments were obtained with any of the YqeY-containing QRSs considered in MSA (from bacterial or eukaryotic origins). In fact, the N-terminal extension of the *Plasmodium* QRS only present a few blocks of conserved residues. However, when submitting QRS sequences from all *Plasmodium* species to the NCBI CD-search in batch mode, a GST-like domain was detected in 8 out of the 45 sequences submitted (Figure 23D), suggesting that a GST domain may be fused at the N-terminus of *Pb*QRS.

It is also interesting to note that residues K402 and R403, involved in discrimination against $tRNA^{E}$ in *E. coli* QRS, are present in *Plasmodium* QRS; R403 is strictly conserved but K402 can be replaced by F or Y as it is the case in *S. cerevisiae* QRS (Grant et al., 2013).

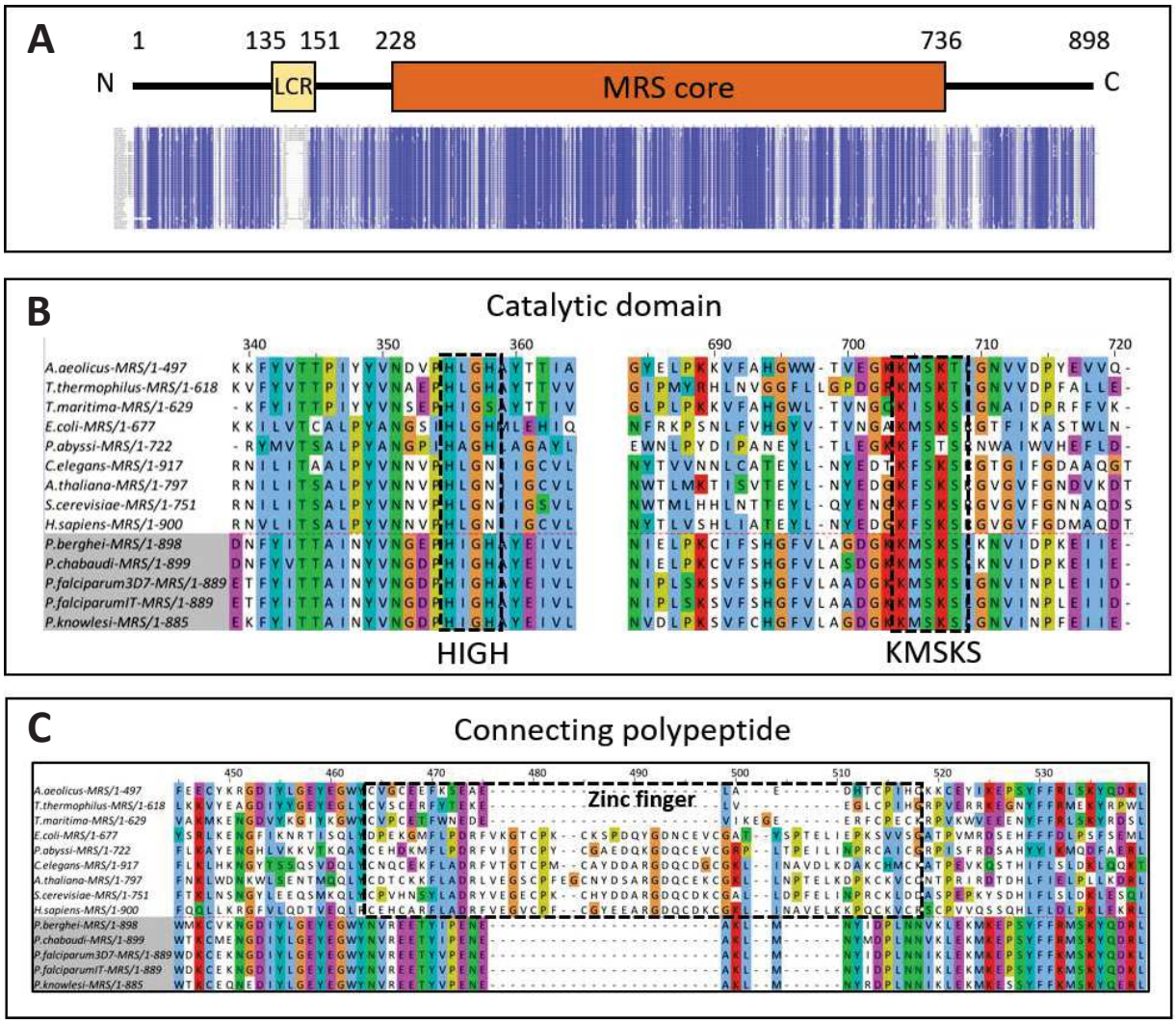


Figure 24: Bioinformatic analysis of *Pb*MRS. A. Conservation of MRS in *Plasmodium* species. Schematic representation of MRS highlighting the different domains identified by BLAST and MSA. The MSA graphical summary includes up to 45 sequences from different *Plasmodium* species and *P. falciparum* strains. Blue-colored columns represent conserved residues. **B. Catalytic domain of *Plasmodium* MRS.** Residues are colored according to Clustal X color code. The two conserved signatures HIGH and KMSKS of class I aaRSs are framed. **C. Absence of a Zn finger domain in the catalytic domain of *Plasmodium* MRS.** *Plasmodium* MRS lacks the zinc finger structure (framed) characteristic of many bacterial and eukaryotic homologues.

(*Figure continues in page 68*).

3.4. *Pb*MRS, the most twisted one

BLAST analysis of *Pb*MRS indicated high similarity between *Plasmodium* MRSs and bacterial MRSs. MSA analysis showed also that MRS contains only one short LCR (Figure 24A) and based on MSA and crystal structures of bacterial MRSs/tRNA complexes, including *E. coli* MRS (Mechulam et al., 1999; Crepin et al., 2003), *T. thermophilus* MRS (Sugiura et al., 2000) and *A. aeolicus* MRS (Nakanishi et al., 2005), all the amino acids important for the catalysis (HIGH and KMSKS) (Figure 24B) and the specific recognition of $tRNA^M$ (not shown) are conserved in *Plasmodium* MRSs. However, one of the most distinctive features of MRSs, the connective polypeptide 1 (CP1), is modified in *Plasmodium* MRSs. In both bacterial and eukaryal MRSs, CP1 contains a zinc finger, this structure is important for the activity of the enzyme and play a key role in methionine activation and correct positioning of the tRNA acceptor stem (Fourmy et al., 1995; Deniziak & Barciszewski, 2001). In *E. coli* MRS, a cluster of cysteines (C145, C148, C158 and C161) is essential for coordinating the Zn^{2+} ion and mutation in any of the residues destabilizes the binding of the ion (Fourmy et al., 1993). This cluster of cysteines is not present in *Plasmodium* MRSs (Figure 24C), indicating that this enzyme does not need zinc to catalyze methionylation. Other MRSs from parasites such as *Leishmania major* and *Trypanosoma brucei* also lack this zinc finger (Larson et al., 2011; Koh et al., 2014).

Interestingly, both extremities of *Plasmodium* MRS show an important level of conservation, which supports the hypothesis of additional structured domains appended to this protein. As it was the case for *Plasmodium* QRSs, I could identify a GST domain only in 6 of the 45 *Plasmodium* MRS sequences submitted to NCBI CD search. But, thanks to MSA, I could show that the *Plasmodium* MRS N-terminal extension aligned well with the GST domain of *H. sapiens* MRS (Cho et al., 2015) and showed also some similarities to the GST of *S. cerevisiae* MRS (Simader, Hothorn & Suck, 2006) (Figure 24D).

I could also observe that the C-terminal extension of *Pb*MRS aligned with the C-terminal EMAPII domain of *A. thaliana* and *C. elegans* MRSs (Kaminska et al., 1999; Havrylenko et al., 2010) and only partially with the C-terminal Trbp111-like domain of MRSs from *E. coli* and *P. abyssi*, which further supports the EMAPII nature of this extension in *Plasmodium* (Figure 24E). In *P. abyssi* and *E. coli*, the extension similar to Trbp111 dimerizes and provides additional tRNA binding properties to the MRS (Crepin et al, 2002). In *Plasmodium*, like in other eukaryotes, the C-terminal EMAPII contains a "pseudo-dyad" extension that mimics the Trbp111 dimer (Renault et al., 2001) and, hence, would not dimerize.

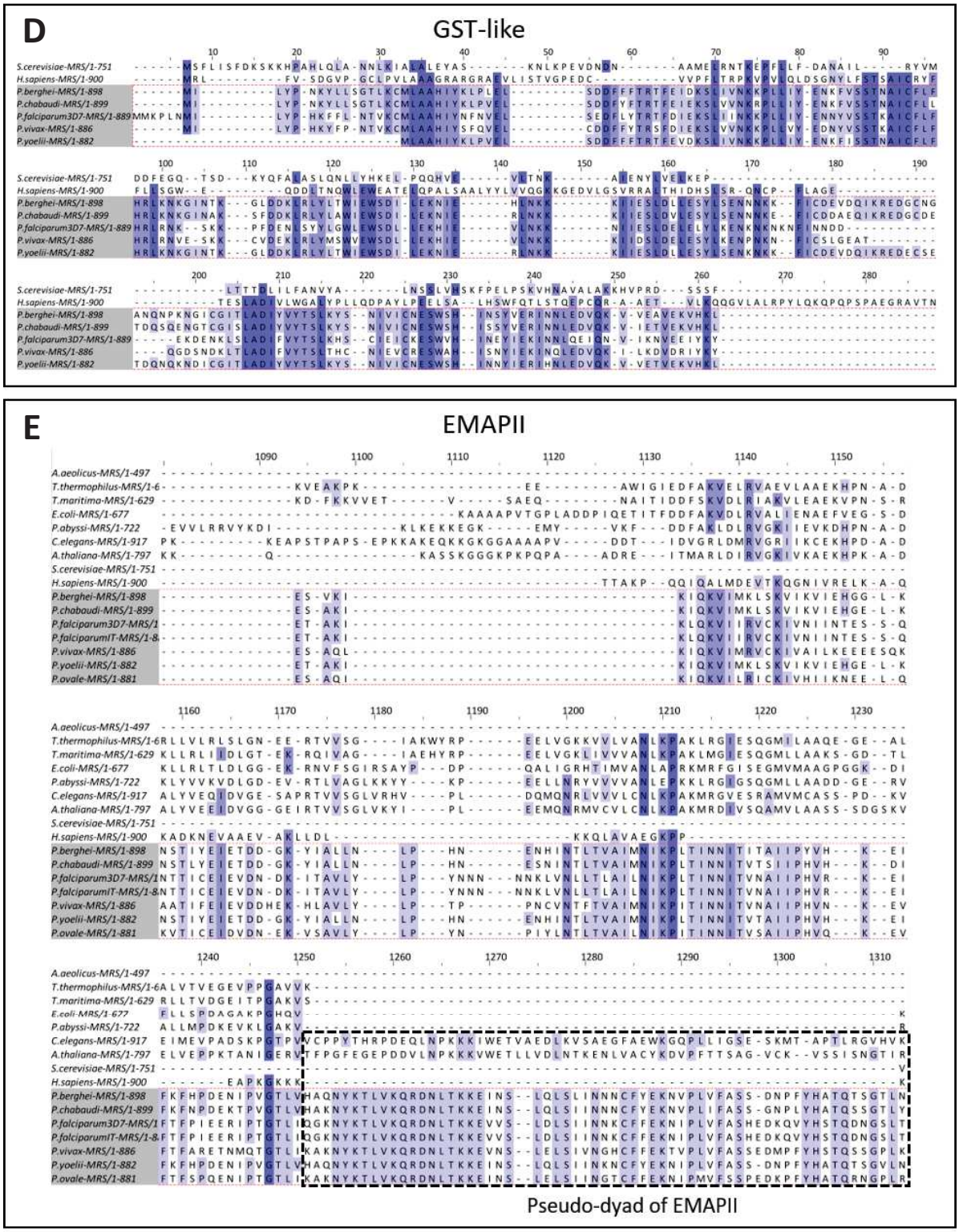


Figure 24 (*Continuation*): Bioinformatic analysis of *Pb*MRS. D. Identification of a N-terminal GST domain. Sequences of two GST domains from eukaryotic MRSs are aligned with a selection of *Plasmodium* sequences of MRS N-terminal domains (highlighted in grey). **E. C-terminal extension of *Plasmodium* MRS.** Several types of MRSs were aligned with MRSs from different *Plasmodium* species (highlighted in grey). Residues are colored by percentage of identity, where blue colors represent conserved residues. The C-terminal extension of *Pb*MRS only aligned with enzymes containing EMAPII-like domains at their C-terminus. Enzymes such as human MRS (C-terminal WHEP) and MRSs lacking a C-terminal additional domain (*S. cerevisiae* and *A. aeolicus*) did not align.

3.5. Modeling of individual proteins

The sequences of *P. berghei* tRip, ERS, QRS and MRS were submitted to the web server Raptor X for structure prediction. This software is particularly appropriate because it searches for different domains in the query sequence and these domains are predicted independently before Raptor X generates a final merged model from individual domains. For each protein, the top templates used for modeling the detected domains are shown in Table 3.

Table 3. Templates selected by Raptor X to model *Pb*tRip, *Pb*ERS, *Pb*QRS and *Pb*MRS. Two (*Pb*tRip, *Pb*ERS, *Pb*QRS) and three (*Pb*MRS) independent domains were detected. Each one was predicted using up to 5 templates. They are presented by decreasing rank and the corresponding PDB ID is indicated.

	N-terminal domain	aaRS core	C-terminal domain
***Pb*tRip**	*Hs*AIMP3 (4BJV-B)	-	EMAPII from *Hs*YRS (1NTG-A)
	GST from *Hs*EPRS (5A1N-A)	-	EMAPII from *Hs*AIMP1(1E7Z-A)
	GST from *Sc*EF-1γ (1NHY-A)	-	EMAPII from *Sc*Arc1p (4R1J-A)
	Sm GST (4NHW-A)	-	Trbp111-like from *Pa*MRS (1MKH-A)
	Af GST (5F8B-A)	-	EMAPII from *Ne*MRS (5H34-A)
***Pb*ERS**	*Bm* δ-class GST (3VK9)	*Ec* QRS (2RD2-A)	-
	Dm GST isozyme E7 (4PNG-A)	*Psa* QRS (5BNZ-A)	-
	Dm GST δ2 (5F0G-A)	-	-
	Md ε-class GST(3VWX-A)	-	-
	Ad δ-class GST (1V2A-A)	-	-
***Pb*QRS**	*Bm* ω-class GST (3RBT-A)	*Psa* QRS (5BNZ-A)	-
	Bm diazinon GST (5ZFG-A)	*Ec* QRS (1O0B-A)	-
	Ce GST (1YQ1-A)	*Dr* QRS (2HZ7-A)	-
	Ov S-crystallin (5B7C)	*Tg* QRS (4P2B-A)	-
	*Hp*GST (1TW9-A)	-	-
***Pb*MRS**	*Bm* diazinon GST (5ZFG-A)	*Ms* MRS (2X1L-A)	EMAPII from *Hs*AIMP1(1E7Z-A)
	At GST U20 (5ECH-B)	*Sa* MRS (4QRD-A)	EMAPII from *Hs*YRS (1NTG-A)
	Ad δ-class GST (1V2A-A)	*Brm* MRS (4DLP-A)	EMAPII from *Sc*Arc1p (4R1J-A)
	Pt GST (5J4U-A)	*Tt* MRS (3VU8-A)	EMAPII from *Ne* MRS (5H34-A)
	At GST U25 (5G5A-A)	*Aa* MRS(2CSX-A)	*Ec* Trbp111 (3ERS-X)

Species abbreviations are Hs: *Homo sapiens*, Sc: *Saccharomyces cerevisiae*, Sm: *Sinorhizobium meliloti*, Af: *Aspergillus fumigatus*, Bm: *Bombyx mori*, Dm: *Drosophila melanogaster*, Md: *Musca domestica*, Ad: *Anopheles dirus*, Ce: *Caenorhabditis elegans*, Ov: *Octopus vulgaris*, Hp: *Heligmosomoides polygyrus*, At: *Arabidopsis thaliana*, Pt: *Populus trichocarpa*, Pa: *Pyrococcus abyssi*, Ne: *Nanoarchaeum equitans*, Ec: *Escherichia coli*, Psa: *Pseudomonas aeruginosa*, Dr: *Deinococcus radiodurans*, Tg: *Toxoplasma gondii*, Ms: *Mycobacterium smegmatis*, Sa: *Staphylococcus aureus*, Brm: *Brucella melitensis*, Tt: *Thermus thermophilus* and Aa: *Aquifex aeolicus*

The Raptor X models were consistent with both the BLAST and the MSA analysis. *Pb*tRip is a GST-EMAPII fusion, the three aaRSs contain N-terminal GST domains and MRS contains an EMAPII domain appended to its C-terminus. Interestingly, the GST domains of *Pb*tRip and the three aaRSs were not modeled with the same set of templates. The N-terminal domain of *Pb*tRip was modeled using mostly GST structures associated to aaRSs or elongation factors, while the N-termini of the aaRSs were modeled based on catalytically active GST enzymes (Table 3). Additionally, the top templates for modeling the EMAPII of *Pb*tRip and *Pb*MRS were also different, since the top template for *Pb*tRip was the EMAPII domain of human YRS while the top template for *Pb*MRS was the EMAPII domain of human AIMP1. No model was produced for the C-terminal extension of QRS, probably because it is too short for Raptor X to consider it as an independent domain. The different structural features identified in this bioinformatic analysis are shown in Figure 25. Especially, the long LCR (62 residues) is visible in the *Pb*QRS model, it is present in the ABD and stands out particularly. The localization of this LCR coincides with the long insertion (27 residues) found in the QRS from *Deinococcus radiodurans*, but the functional relevance of this (if any) is unknown (Deniziak et al. 2007).

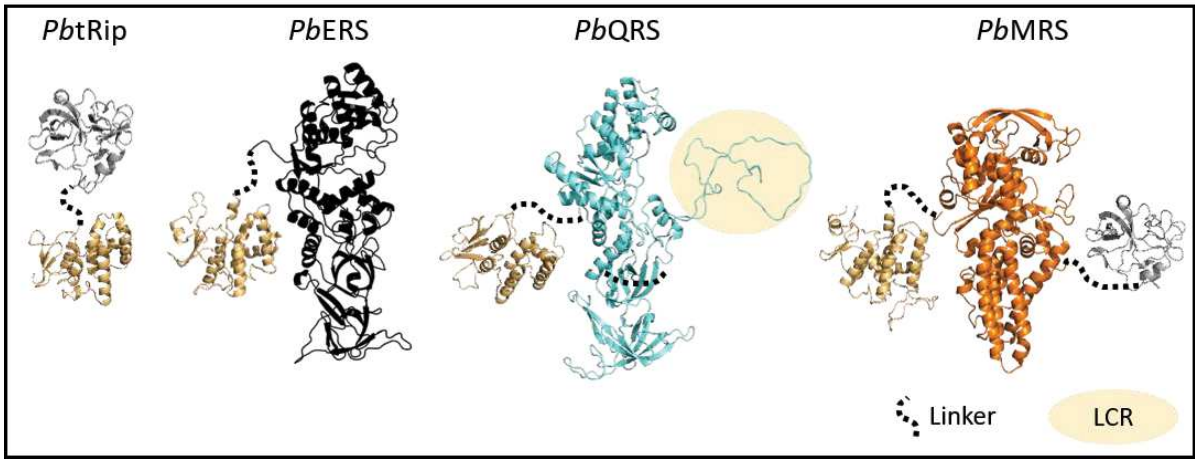


Figure 25: Modeling of individual components of the *Plasmodium* MARS complex (Raptor X). The catalytic cores of *Pb*ERS, *Pb*QRS et *Pb*MRS are colored in black, cyan and orange, respectively. The N-terminal GST domains are shown in gold, EMAPII domains are in silver and the long LCR in the ABD of *Pb*QRS is highlighted in pale yellow.

III. RNA binding properties of the C-terminal domains of *Pb*QRS and *Pb*MRS

The prediction of C-terminal domains with potential tRNA binding abilities in *Pb*MRS and *Pb*QRS led us to test whether these domains can actually bind tRNAs *in vitro*. As a preliminary test, the C-terminal domain of *Pb*MRS (residues 730 – 898) was cloned and expressed in *E. coli*. This domain could be purified to homogeneity but did not bind to RNA (not shown). This observation led us to modify our strategy. For *Pb*MRS and *Pb*QRS, peptides that encompass both the enzyme ABD and the C-terminal domain were cloned and expressed in *E. coli*. As controls, the same constructs without the C-terminal domain were also designed and produced (Figure 26A).

Each polypeptide was affinity purified on Ni-NTA column, using their 6-His N-terminal tag (Figure 26B). Each pair of polypeptides (ABD control and ABD-C-terminal domain) was tested in parallel on a particular native polyacrylamide gel. Increasing amounts of polypeptides were immobilized in the gel together with a competing nucleic acid (here, 3 times more poly-T DNA than the polypeptide) and radiolabeled tRNA (total human tRNA) was run through the gel. In the presence of any RNA-protein interaction, the migration of the radioactive tRNA is slowed down. Examples of this kind of gels can be seen in Figure 26C. This technique allows the detection of interactions that are not stable, but does not allow the determination of any affinity constant.

Slower migration is observed with the ABD alone around 1.5 and 2 µM, suggesting that the ABD binds to total tRNA. However, what we observe looks more like traces than well-defined bands, suggesting that these interactions are only transient. In contrast, when the C-terminal domains are fused to their respective ABD, tRNA migration is significantly reduced and results in much better defined bands, indicating that the presence of these C-terminal domains significantly increases the tRNA binding affinities of *Pb*MRS and *Pb*QRS ABDs. The same experiments were also performed with *in vitro* transcribed $tRNA^M$ and $tRNA^Q$ instead of total tRNA, but this did not change the binding profile (not shown), confirming that the binding of the C-terminal domains is non-sequence specific.

Binding capacity is generally low for isolated EMAPII domains, whether it is the human AIMP1 domain (Quevillon et al., 1997), or the *O. sativa* MRS domain (Kaminska et al., 1999), which have affinities for tRNA of about 20 µM and 15 µM, respectively. Similarly, the tRNA binding capacity of an isolated WHEP domain from the human EPRS is also characterized by a very low affinity, about 10 µM (Cahuzac et al., 2000). However, such domains, when fused to the C-terminus of

aaRSs, are sufficient to provide a much higher apparent affinity to these enzymes for their specific tRNAs. By analogy, we propose that, with their general tRNA binding properties, the C-terminal domains of *Pb*MRS and *Pb*QRS could also act synergistically to confer to their respective native enzymes the ability to bind tRNA with higher affinities.

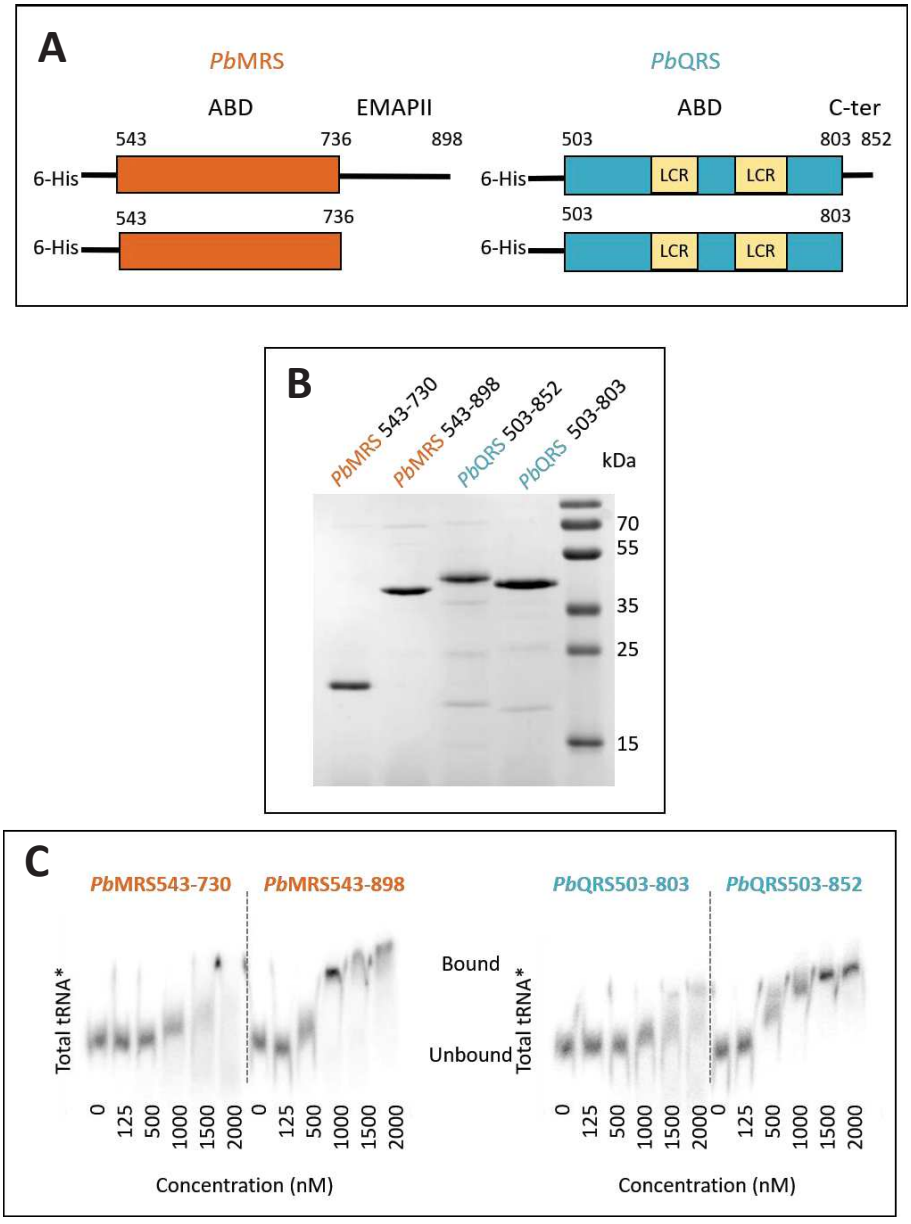


Figure 26: Design, purification and RNA binding properties of the C-terminal domains of *Pb*MRS and *Pb*QRS. A. Design of recombinant domains. Anticodon-binding domains (ABD) of *Pb*MRS and *Pb*QRS were produced in *E. coli* either as free polypeptides or fused to their corresponding C-terminal extension. The four constructs contain a N-terminal 6-His tag important for their purification. **B. Purification of recombinant domains.** The four proteins were analyzed by SDS-PAGE. **C. Binding properties of C-terminal domains of *Pb*MRS and *Pb*QRS.** ^{32}P-labeled total human tRNA was run through a native gel containing increasing concentrations of *Pb*MRS ABD or *Pb*QRS ABD fused or not (control) to their corresponding C-terminal domains. After electrophoresis at 4°C, the mobility shift of tRNA was visualized by autoradiography.

IV. Determination of an interaction network *in vitro*

1. Our strategy to express *Plasmodium* proteins

Since the AT content in the wild-type DNA of most *Plasmodium* proteins is above 70%, sequences were optimized to increase the chances of expression of *Pb*tRip and the three aaRSs in heterologous systems. Although use of *E. coli* was planned, sequences were adapted to the human codon usage instead. Previous work from the team (Cela et al., 2018) indicated that codon optimization for *E. coli* might be detrimental to the solubility of proteins from *Plasmodium*. The correct folding of these proteins might require some decrease in the translation rate at codons corresponding to rare tRNAs species in the parasite (Frugier et al., 2010). This is discordant with common approaches of codon optimization, where rare codons are minimized in order to avoid ribosome stalling. By adapting the *Plasmodium* sequences to the codon usage in human, the content of AT is decreased while keeping some rare codons that will slow down the translation of the recombinant protein in *E. coli*. In addition, induction of protein expression at low temperature (16°C) was crucial to obtain soluble proteins, suggesting that reduction in global translation rate might compensate, in part, the absence of appropriate "tRNA sponges" for the correct folding of *Plasmodium* proteins in *E. coli*.

Despite sequence optimization, production of full-length *Pb*ERS, *Pb*QRS and *Pb*MRS was challenging. Indeed, expression was low, solubility was limited and proteins were prone to proteolysis even in the presence of protease inhibitors. Purification of low amounts of full-length *Pb*MRS was possible, but *Pb*ERS and *Pb*QRS were always shorter than expected. Mass spectrometry analysis revealed that these two proteins lacked their entire N-terminal GST domains. Some minor improvements were obtained when changing the plasmid vector and/or the bacterial strain. However, preliminary experiments showed that N-terminally truncated *Pb*ERS and *Pb*QRS as well as *Pb*MRS were unable to interact with *Pf*tRip *in vitro*, suggesting that the absence of GST domains in both *Pb*ERS and *Pb*QRS could impair the formation of the complex. Indeed, these GST domains are essential in the formation of the yeast and human MARS complexes (Simader, Hothorn & Köhler, 2006; Cho et al., 2015). This is also supported by the observation that the sole GST domain of Tg-p43 was sufficient to precipitate all the components of the MARS complex in *Toxoplasma gondii* (van Rooyen et al., 2014).

We decided to express the GST domains of the 3 aaRSs individually to investigate their interactions with each other and with *Pb*tRip. Hereafter, these domains are referred to as *Pb*ERS-N, *Pb*QRS-N and *Pb*MRS-N. Several constructions with and without 6His-tag were engineered (Table 9 in Methods). *Pb*tRip, the different constructs of its GST domain (referred as *Pb*tRip-N) and *Pb*ERS-N were expressed in *E. coli* as soluble proteins with no difficulties, although the growth rate of bacteria expressing *Pb*ERS-N constructions was significantly reduced. On the other hand, *Pb*MRS-N and *Pb*QRS-N were well expressed but their solubility was limited (Figure 27A).

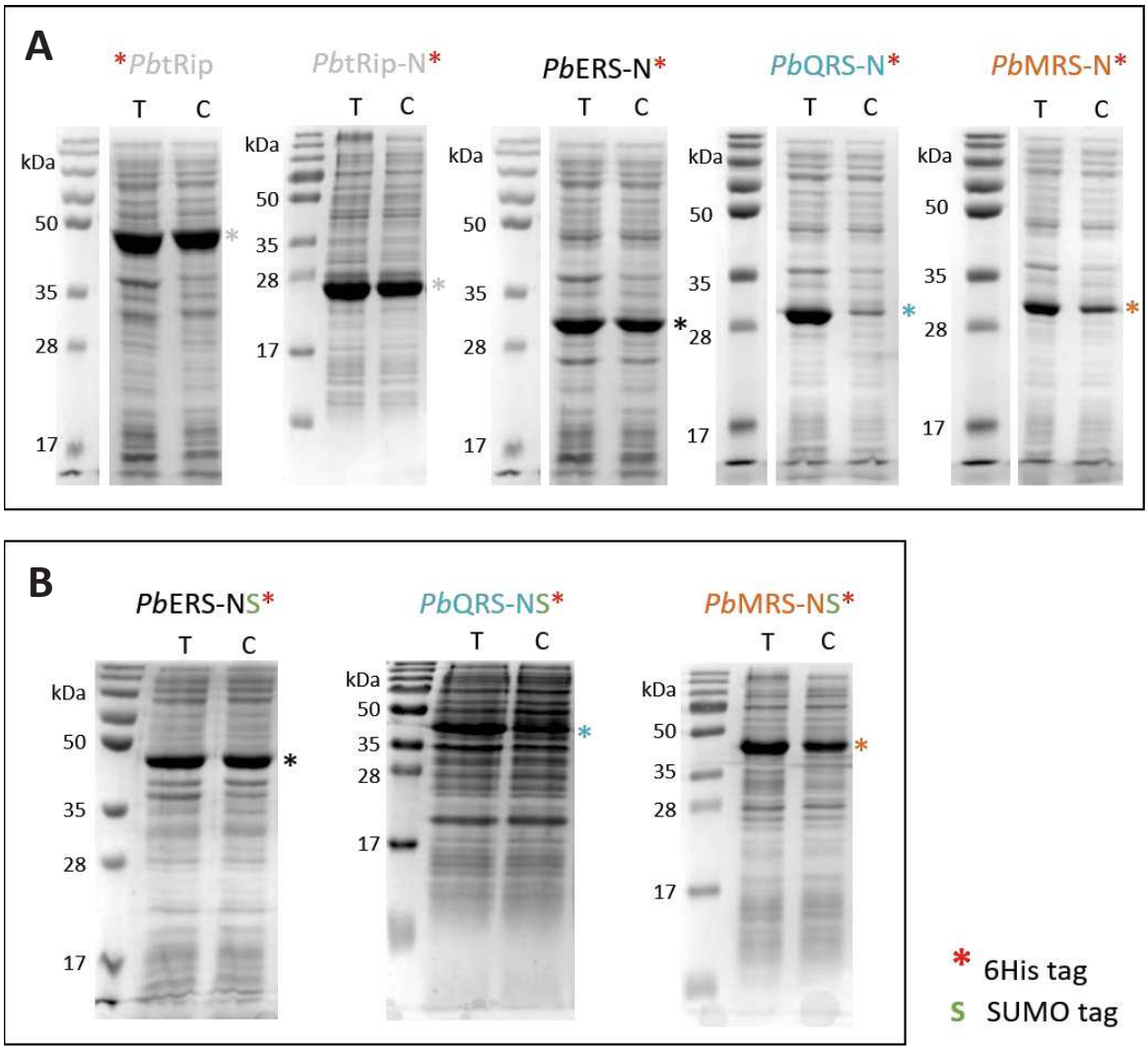


Figure 27. Expression and solubility of the different GST domains. Protein extracts from bacteria after induction were systematically analyzed by SDS-PAGE. The solubility of the recombinant protein was assessed by comparing the intensity of the overexpressed band in the total (T) and centrifuged (C) extracts (see section 3 in Methods). **A. Expression and solubility of His-tagged versions of full-length *Pb*tRip and the GST domains *Pb*tRip-N, *Pb*ERS-N, *Pb*QRS-N and *Pb*MRS-N. B. Expression and solubility of His-tagged versions of GST-SUMO fusions of *Pb*ERS-N, *Pb*QRS-N and *Pb*MRS-N.** The red asterisk indicates the 6His-tag and the green S indicates the fusion of the SUMO tag at the C-terminus of the proteins.

In addition, a set of constructs was produced in which a SUMO protein is fused to the C-terminal end of these GST domains (Figure 27B). The purpose of the SUMO protein was dual. On the one hand, it enhanced the solubility of the proteins, although only a slight improvement was observed for *Pb*QRS-N. On the other hand, it provided proteins of different sizes that can be distinguished one from each other by SDS-PAGE. This property was crucial to interpret pull-down experiments and other types of biochemical analysis. In some cases, the expression of the 6His-tagged and non-tagged version of a same construct was appreciably different, since the addition of 6 histidines may change significantly the isoelectric point (pI) of the protein (Table 13 in Methods).

2. Looking for domain-domain interactions

The capacity of *Pb*tRip and the 3 aaRSs to interact was investigated by pull-down experiments. A mixture of cells expressing (i) the 6His-tagged "bait" protein and (ii) one or more "prey" proteins was lysed and the protein extract incubated with a Ni-NTA resin. This resin binds only the bait protein and would indirectly capture any prey protein interacting with the bait. After incubation, the resin is extensively washed and bound proteins are eluted and analyzed by SDS-PAGE. If an interaction occurred, bait and prey(s) proteins appear together in the gel.

Growth and conditioning of the *E. coli* cultures were critical to obtain reliable results in pull-down experiments. Since the expression and solubility were not the same for the different partners, the volume of bacteria had to be adjusted according to the expression level of each protein. Thus, whenever a new culture was grown, the expression level of the overexpressed protein was verified by SDS-PAGE in order to have similar amounts of soluble recombinant proteins in the pull-down mixture. In the case of *Pb*tRip and *Pb*ERS-N, equivalent volume of bacterial culture was used, as these proteins were expressed at similar levels and had comparable solubility. In contrast, since the solubility of *Pb*MRS-N and *Pb*QRS-N were more limited (Figure 27), 2 and 4 volumes of bacterial culture were used for *Pb*MRS-N and *Pb*QRS-N, respectively. Moreover, we observed that excessive amounts of GSTs lead to unspecific interactions where each domain interacts with any other domain.

2.1. Pairwise interactions

First, we tested the capacity of each partner to interact with itself and with other proteins. A summary of these results is shown in Figure 28A.

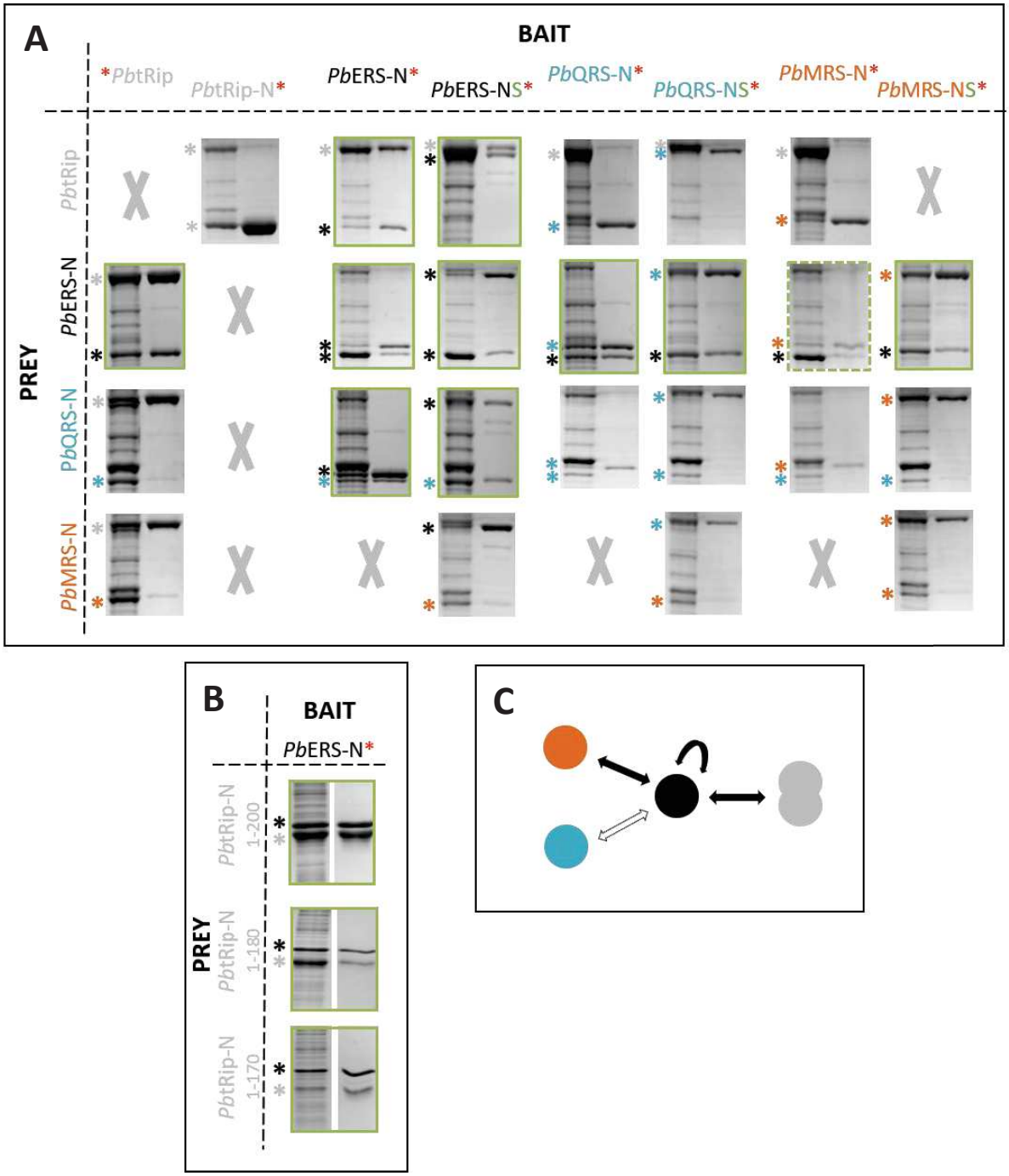


Figure 28. Pairwise interactions. A. 24 different combinations. The results of different pull-down experiments involving 2 proteins are shown (bait as x-axis and prey as y-axis). For each combination, both the initial mixture and the proteins captured in the Ni-NTA resin appear on SDS-PAGE. Some combinations couldn't be tested because both bait and prey proteins have the same length and thus co-migrated in the gel; they are indicated with grey crosses. When an interaction could be detected, the corresponding gel was framed in green. The bands corresponding to bait and prey proteins in the initial mixture are indicated with asterisks of the corresponding color (grey for *Pbt*Rip, black for *Pb*ERS, cyan for *Pb*QRS and orange for *Pb*MRS). Each interaction has been tested at least 3 times. **B. *Pb*ERSN:*Pbt*Rip-N complexes.** Interaction between *Pb*ERS-N-6His (bait) and *Pbt*Rip-N constructs of different lengths (the number of residues is indicated). **C. Preliminary interaction network.** Based on these data, we propose a first interaction network for the 4 proteins.

Each partner was used alternatively as bait or prey. The prey proteins were always non-tagged GST domains, whereas bait proteins were tested in two different versions. They corresponded either to the GST domain with a 6His-tag directly fused at its C-terminus or to the GST-SUMO fusion with a C-terminal 6His-tag. Likewise, the interaction capacities of the GST domain of *Pb*tRip was tested either in the context of the full-length tRip (hereafter simply referred as "*Pb*tRip") or as the GST domain *Pb*tRip-N.

*Pb*ERS-N was the only aaRS GST domain that interacted with *Pb*tRip in a reciprocal mode, whether the 6His-tag was on *Pb*ERS or on *Pb*tRip (Figure 28A). The C-terminal domain of *Pb*tRip was dispensable for the interaction with *Pb*ERS-N (Figure 28B). Although the GST domain of *Pb*tRip was previously delimited within the first 180 amino acids, the more efficient interaction was observed with the construction including the first 200 residues, suggesting that critical interaction residues lie within the first 170 amino acids of the GST domain and that residues 171 to 200 might be important to stabilize this interaction *in vitro*.

Neither *Pb*QRS-N nor *Pb*MRS-N did show any interaction with *Pb*tRip, but both interacted with *Pb*ERS-N. However, this interaction is only unambiguously observed when *Pb*MRS-N-SUMO-6His is used as a bait; otherwise, it appears that the presence of a tag (6His or SUMO-6His) on *Pb*ERS-N does not allow an interaction stable enough to be identified by pull-down. Interestingly, *Pb*ERS interacts with itself, suggesting that this GST domain can oligomerize. Finally, no interaction between *Pb*MRS and *Pb*QRS could be detected with this technique. Based on these observations, a first interaction network was proposed (Figure 28C), where *Pb*ERS plays a central role in this complex and binds *Pb*tRip, *Pb*QRS and *Pb*MRS, although less efficiently.

2.2. Understanding domain organization within the complex

2.2.1. Pull-down experiments with four domains

Four different pull-down experiments were performed, the only difference being the nature of the bait protein used. Surprisingly, the results definitely depended on the bait selected (Figure 29A). The use of *Pb*tRip or *Pb*ERS-N as baits allowed the capture of the three other proteins (*Pb*ERS-N or *Pb*tRip, *Pb*QRS-N and *Pb*MRS-N), but when *Pb*QRS-N or *Pb*MRS-N were used, only *Pb*tRip and *Pb*ERS-N were present in the elution. In other words, the use of *Pb*QRS-N as bait excluded *Pb*MRS-N from the complex and *vice versa*.

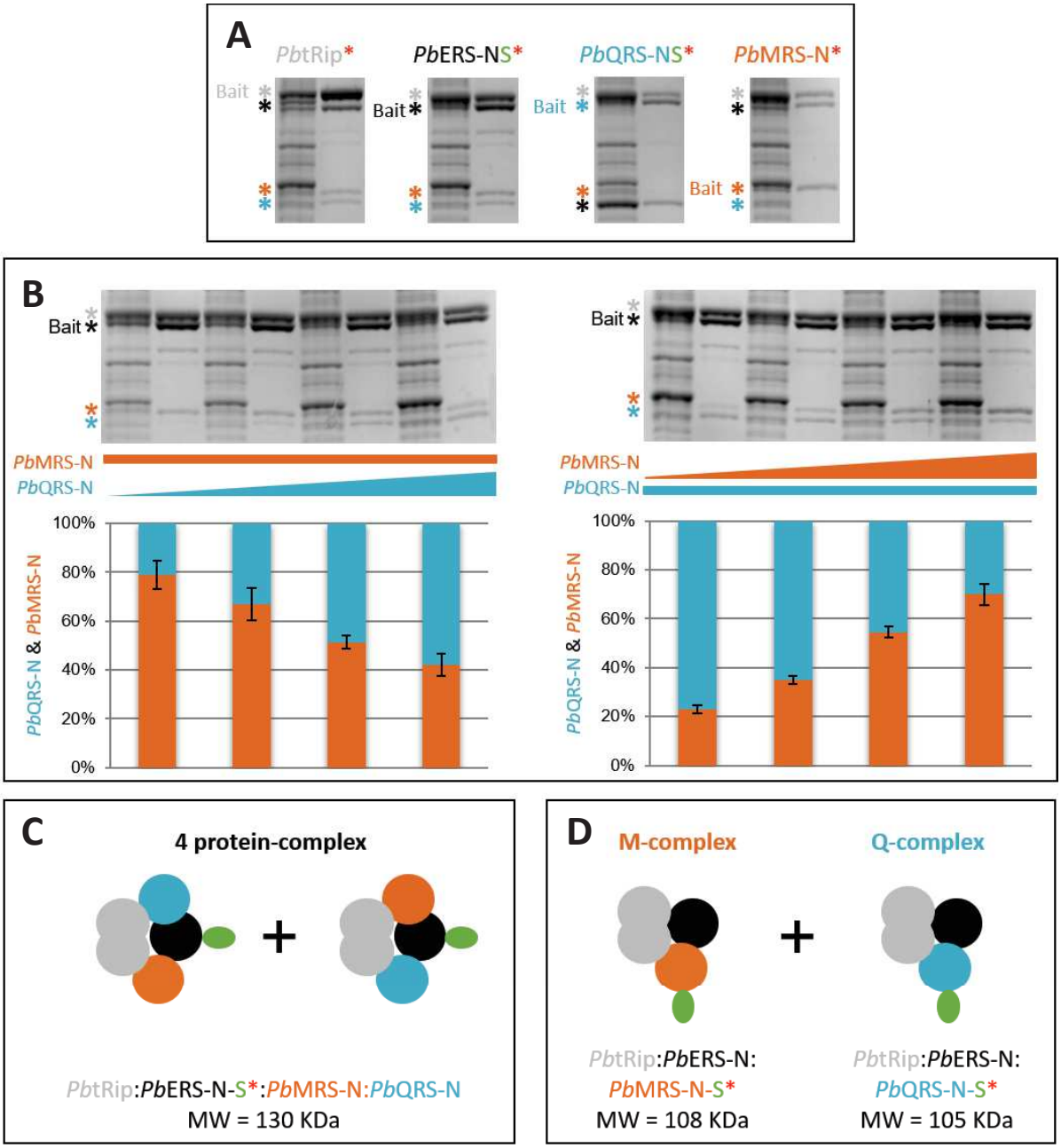


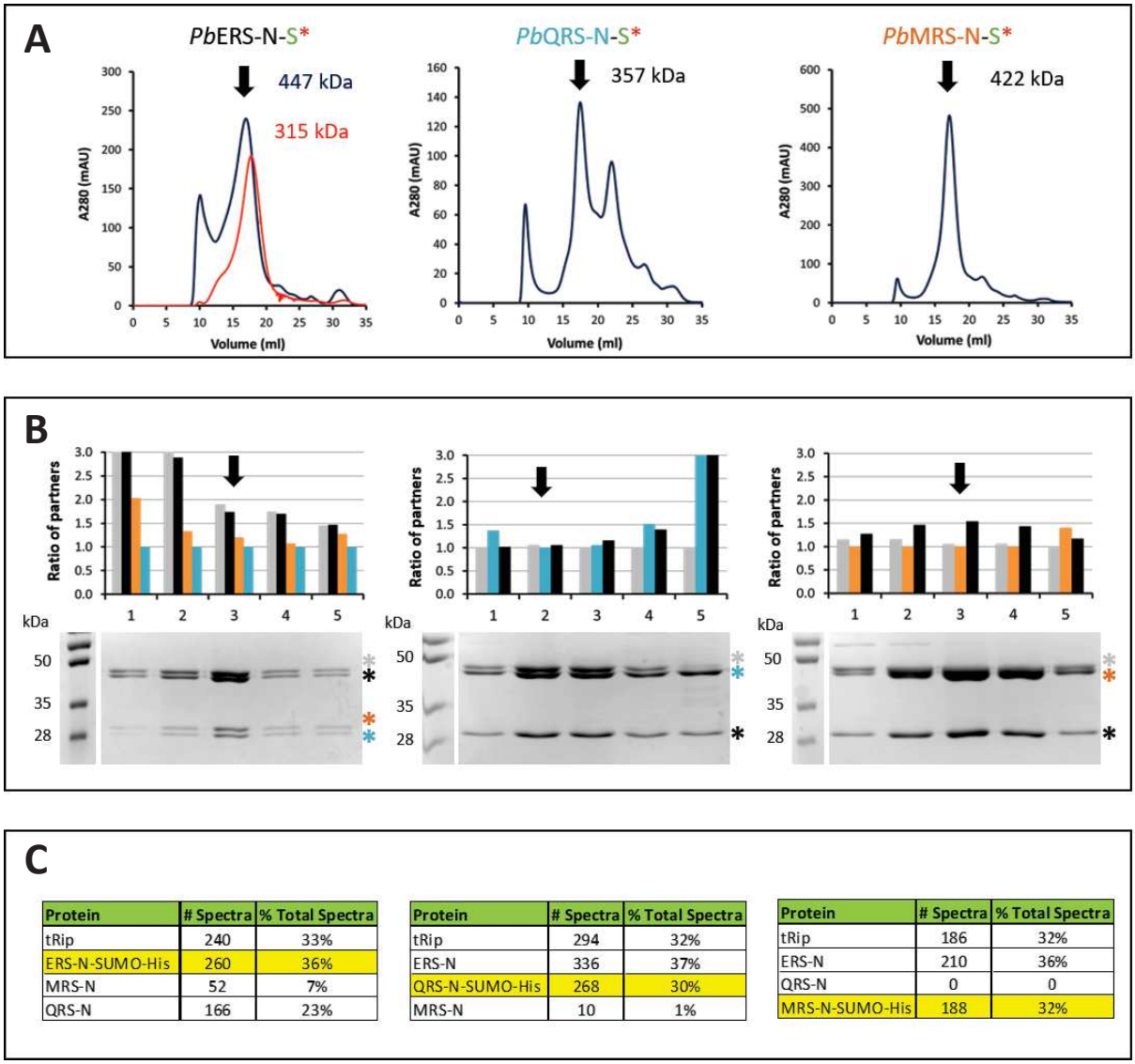
Figure 29. How many domains in the complex? A. Pull-down experiments performed with 4 proteins. Each protein was used as bait to capture the 3 others. The content of the initial mixture and the captured proteins are shown on each gel. Partners are identified with asterisks colored in grey (*Pbt*Rip), black (*Pb*ERS), cyan (*Pb*QRS) and orange (*Pb*MRS) and the bait protein is indicated. **B. Competition experiments between *Pb*QRS-N and *Pb*MRS-N for binding of the *Pbt*Rip:*Pb*ERS-N subcomplex.** Series of pull-down experiments where *Pb*ERS-N-SUMO-6His (bait) was used to capture *Pbt*Rip and either (i) *Pb*QRS in the presence of increasing amounts of *Pb*MRS-N or (ii) *Pb*MRS in the presence of increasing concentrations of *Pb*QRS. The relative amount of *Pb*QRS-N and *Pb*MRS-N captured in each pull-down is represented on the bar plot. Error bars were obtained from three replicates **C. and D. Two alternative models for the organization of *Pb*MARS complex: (C)** The formation of a unique complex containing 4 domains or **(D)** the formation of two independent ternary complexes, M- and Q-complexes**.** Approximate MWs of each complex are indicated.

The existence of a unique complex containing the four domains was challenged by varying the concentrations of *Pb*QRS-N and *Pb*MRS-N. The concentrations of *Pb*QRS-N or *Pb*MRS-N were increased in the initial mixture while keeping constant the other 3 proteins: *Pb*tRip, *Pb*ERS-N-SUMO-6His (bait) and *Pb*MRS-N or *Pb*QRS-N. By increasing *Pb*QRS-N, less *Pb*MRS-N was retained in the complex. Reciprocally, increasing *Pb*MRS-N in the initial mixture lead to the enrichment of this domain in the complex at the expense of *Pb*QRS-N (Figure 29B). In other words, *Pb*QRS-N and *Pb*MRS-N are in competition for binding the binary complex *Pb*tRip:*Pb*ERS-N-SUMO-6His. This behavior is compatible with 2 scenarios, (i) there are 2 binding sites on *Pb*tRip:*Pb*ERS-N, where both *Pb*QRS-N and *Pb*MRS-N can bind alternatively to form one complex with four proteins (Figure 29C) or (ii) *Pb*QRS-N and *Pb*MRS-N bind the same site on *Pb*tRip:*Pb*ERS-N and, hence, two independent complexes exist (Figure 29D). Since no *Pb*MRS-N was observed when *Pb*QRS-N is used as bait and reciprocally (Figure 29A), the second scenario seems more likely.

2.2.2. Purification and analysis of complexes prepared with four domains

A different approach was used to verify the existence of one complex with four proteins or two complexes with three proteins (Figure 29C and 29D). The mixtures of *E. coli* used for pull-down experiments were scaled-up and subjected to two purification steps: Ni-NTA-affinity chromatography followed by size-exclusion chromatography (SEC). In all cases, the formation of soluble complexes was observed on the SEC step. The SEC profiles, shown in Figure 30A, allowed to determine the apparent molecular weight of each complex and the major peaks were analyzed by SDS-PAGE (Figure 30B). The quantification of the bands detected on the gel allowed the determination of the relative amounts of the partners and thus the estimation of the oligomeric state in the different purified complexes. In addition, fractions corresponding to main SEC peaks were analyzed by mass spectrometry to verify the identity of the purified proteins (Figure 30C). Three purifications were performed using different baits. All baits were fused to a SUMO tag to give enough space and avoid any steric hindrance during the capture of on the Ni-NTA resin.

As expected, using *Pb*ERS-N-SUMO-6His as bait allowed the co-purification of the 4 proteins, which eluted as 2 main populations on the SEC chromatogram. The main peak contained the 4 partners. The amount of each partner in this sample was variable among the different replicates, *Pb*tRip and *Pb*ERS-N were always equimolar, whereas *Pb*QRS-N and *Pb*MRS-N showed more variations, but were always significantly less abundant than *Pb*tRip and *Pb*ERS-N. A second peak appeared in the void volume of the column and corresponds to aggregates. Both peaks were quite broad and overlapped, suggesting the presence of intermediary populations. Interestingly, the



C

Protein	# Spectra	% Total Spectra
tRip	240	33%
ERS-N-SUMO-His	260	36%
MRS-N	52	7%
QRS-N	166	23%

Protein	# Spectra	% Total Spectra
tRip	294	32%
ERS-N	336	37%
QRS-N-SUMO-His	268	30%
MRS-N	10	1%

Protein	# Spectra	% Total Spectra
tRip	186	32%
ERS-N	210	36%
QRS-N	0	0
MRS-N-SUMO-His	188	32%

Figure 30. Purification of complexes in the presence of 4 proteins. Three mixtures of bacteria expressing the 4 proteins partners were co-lysed and subjected to Ni-affinity chromatography followed by size-exclusion chromatography (SEC). In each case, a different bait protein (indicated above the figure) was used to capture the 3 other proteins. **A. SEC chromatograms**. They were obtained using a SepFast column (6-5000 kDa). The apparent MW calculated from a calibration curve is shown for each main peak (black arrows). The profiles of two different purifications performed in different conditions (red and blue lines) are shown for the bait *Pb*ERS-N-SUMO-6His. **B. SDS-PAGE analysis.** Contiguous fractions lying within the main SEC peak were analyzed by SDS-PAGE. The relative abundance of each partner in the fraction was calculated and plotted. Colors were conserved grey for *Pb*tRip, black for *Pb*ERS-N, cyan for *Pb*QRS-N and orange for *Pb*MRS-N. The fraction corresponding to the center of the peak is indicated with a black arrow. **C. Mass spectrometry**. Fractions corresponding to the center of each main SEC peak were analyzed by mass spectrometry. The number of spectra and percentage of total spectra are shown for each partner. The bait protein is highlighted in yellow.

intensity of the peak of aggregates correlated with the amount of *Pb*tRip and *Pb*ERS-N-SUMO-6His present in the protein extract. Indeed, using less of these two proteins lead to an important reduction of aggregates (red chromatogram in Figure 30A), suggesting that an excess of *Pb*tRip:*Pb*ERS-N-SUMO-6His causes aggregation.

Using *Pb*QRS-N-SUMO-6His or *Pb*MRS-N-SUMO-6His as baits lead to the purification complexes containing only three proteins. As seen in pull-down experiments (Figure 29A), the use of *Pb*QRS-N as bait excluded *Pb*MRS-N from the complex and reciprocally.

The sample purified using *Pb*QRS-N-SUMO-6His as bait appeared less homogeneous on the SEC analysis than the sample purified with the bait *Pb*MRS-N-SUMO-6His (Figure 30A). Indeed, the sample purified using *Pb*QRS-N-SUMO-6His showed multiple populations: (i) aggregates are found in the void volume and proteins that entered the column separated in 2 other populations. Their analysis by SDS-PAGE showed different ratios between eluted proteins, either a 1:1:1 ratio (*Pb*QRS-NS*:*Pb*tRip:*Pb*ERS-N) in the main peak and a strong enrichment in *Pb*QRS-N-SUMO and *Pb*ERS-N in the late peak. Since *Pb*QRS-N and *Pb*ERS-N interacted efficiently in pull-down experiments (Figure 28A), this population might correspond to a stable binary complex *Pb*QRS-N:*Pb*ERS-N.

The co-purification using *Pb*MRS-N-SUMO-6His as bait produced a more homogenous sample. Proteins eluted from the SEC column as a single peak and only small populations of aggregates were observed. However, *Pb*tRip and *Pb*MRS-N-SUMO co-migrated in the most concentrated fractions and compromise the reliability of SDS-PAGE quantifications.

SDS-PAGE indicated that purity of complexes after 2 chromatographic steps was quite good and this was confirmed by the mass spectrometry analysis. Only a few *E. coli* contaminants were detected and most of them had less than 10 spectra. The chaperone DnaK was more present (~20 spectra) and was recurrently identified in several samples. This might be a reflection of folding issues of LCR-containing proteins. Globally, mass spectrometry results correlated with SDS-PAGE quantifications. In the complex containing the 4 proteins, *Pb*tRip and *Pb*ERS-N-SUMO-6His were detected with similar and high number of spectra which represented 33% and 36%, respectively of the total spectra, whereas *Pb*QRS-N and *Pb*MRS-N were both detected with lower numbers, corresponding to 23% and 7% of the spectra, respectively. In both complexes showing only 3 proteins on SDS-PAGE, mass spectrometry confirmed the absence of the fourth one. *Pb*MRS-N represented only 1% of spectra in the sample purified with *Pb*QRS-N-SUMO-6His. Moreover, the spectra counts confirmed the first impression of a 1:1:1 ratio for both types of ternary complexes.

Finally, the molecular weight (MW) calculated from the elution volume in the SEC column provided a general idea of the oligomeric state of the complexes. All of them were characterized by MWs between 300 and 450 kDa. These values are much higher than the estimations calculated with only one copy of each domain inside each complex (between 100 and 130 kDa), suggesting higher oligomeric forms such as dimers, trimers or even tetramers.

2.2.3. Segregation of complexes upon tRNA binding

The tRNA-binding capacity of the "4-proteins" complex was studied by electrophoretic mobility shift assay (EMSA) in agarose gel. This choice was dictated by the size of the different complexes (between 105 and 130 kDa in absence of any oligomerization, Figures 29C and 29D), which prevented the use of polyacrylamide gels. In these experiments, a fixed amount of complex is incubated with decreasing concentrations of total tRNA from yeast and the mixtures are subjected to electrophoresis in native conditions. The gel was stained with ethidium bromide to visualize tRNA and with Coomassie blue to visualize the proteins (Figure 31).

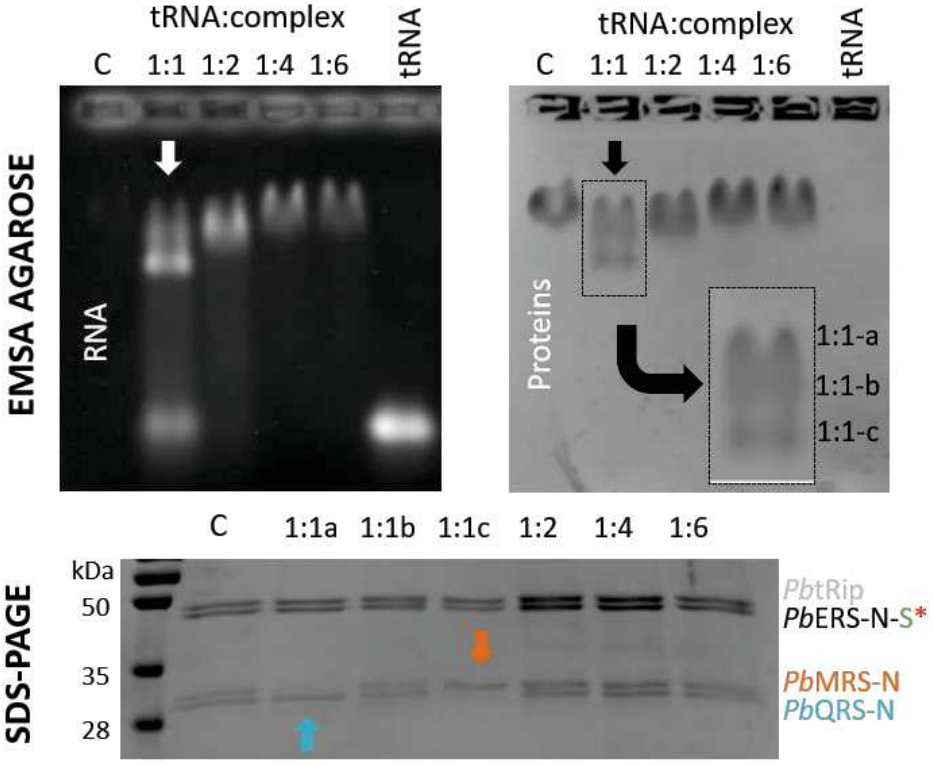


Figure 31. tRNA binding properties of the 4 protein-complex. The 4 protein-complex (C) was purified as in Figure 30, using *Pb*ERS-N-SUMO-6His as bait. Four ratios of total yeast tRNA:complex were used (1:1, 1:2, 1:4 and 1:6) to assess the tRNA binding capacity of the complex. Binding reactions were analyzed on 1% agarose gel. RNA and proteins were visualized with ethidium bromide and Coomassie blue, respectively. Three bands, a, b and c were observed at ratio 1:1 tRNA:complex and analyzed by SDS-PAGE.

tRNA binding was observed since the complex and the tRNA co-migrated in the gel. However, for the 1:1 ratio (tRNA:complex), we could distinguish 3 populations (a, b and c). The 3 bands were sliced from the gel and their protein content was further analyzed by SDS-PAGE. Surprisingly, the upper-band (a) contained *Pb*tRip, *Pb*ERS-N-SUMO-6His and only *Pb*QRS-N and the lowest band (c), contained *Pb*tRip, *Pb*ERS-N-SUMO-6His and only *Pb*MRS-N. Yet, the 4 proteins were observed in the middle band (b), which is most likely a consequence of an overlap of bands a and c (band broadening). This observation provided additional evidence for the existence of two independent ternary complexes. For convenience the *Pb*QRS-N:*Pb*ERS-N:*Pb*tRip will be referred as the Q-complex and *Pb*MRS-N:*Pb*ERS-N:*Pb*tRip as the M-complex.

The two individual complexes were then tested for their tRNA binding capacity and specificity (Figure 32). EMSA experiments involving ternary complexes with and without the C-terminal domain of *Pb*tRip confirmed the role of this domain in tRNA binding. Indeed, tRNA binding occurred only in complexes containing full-length *Pb*tRip and no tRNA shift was observed for complexes lacking the C-terminal domain. This is consistent with EMSA experiments with *P. falciparum* tRip, in which the full-length protein and its C-terminal domain (linker + EMAPII-like domain) were responsible of high affinity interactions with tRNA, but not the N-terminal domain (Bour et al., 2016). Moreover, as predicted from the EMSA performed with the 4 protein-complex, the migration profiles were different for each complex, (compare Figures 32A and 32B). The Q-complex and the M-complex form a unique band at all the tRNA:complex ratio tested, indicating that the population of tRNA binding complexes is homogeneous. However, the tRNA-bound M-complex, especially at the 1:1 ratio, migrated further away in the gel than the M-complex alone (Figure 32B). Although negative charges provided by tRNA certainly contribute to this amplified migration, this could be also the result of a compaction of the complex upon tRNA binding, as it has been observed for the MARS complex in yeast (Koehler et al., 2013).

In addition, team's DLS and SEC data also suggested that *P. falciparum* tRip undergoes a similar compaction when tRNA is present in equimolar concentration (Dr. Anne Théobald-Dietrich unpublished data). This effect disappears when the complex concentration increases to 2 and 4 complexes per tRNA molecule. Since all the tRNAs are bound, it could indicate that the complex relaxes or that the complex tends to oligomerize in the presence of limiting concentrations of tRNA (such an observation was made with the *P. falciparum* tRip, Dr Anne Théobald-Dietrich unpublished data). The "oligomerization" hypothesis is further supported by the fact that an excess of free tRNA was observed in samples containing equimolar amounts of complex and tRNA, suggesting that more than one copy of the Q- or M-complex is required to bind one tRNA molecule.

However, the stoichiometry at which complete tRNA binding occurs is difficult to determine since there is no condition where the tRNA-bound complex can be discriminated from the unbound complex. Neither additional purification steps nor the presence of the SUMO tag influenced the migration profile of M- or Q- complexes in presence of tRNA.

It is important to note that the 3 partners initially present in the Q-complex or in the M-complex remains in tRNA-bound complexes, indicating that tRNA binding does not induce the complex dissociation and the release of any protein *in vitro*.

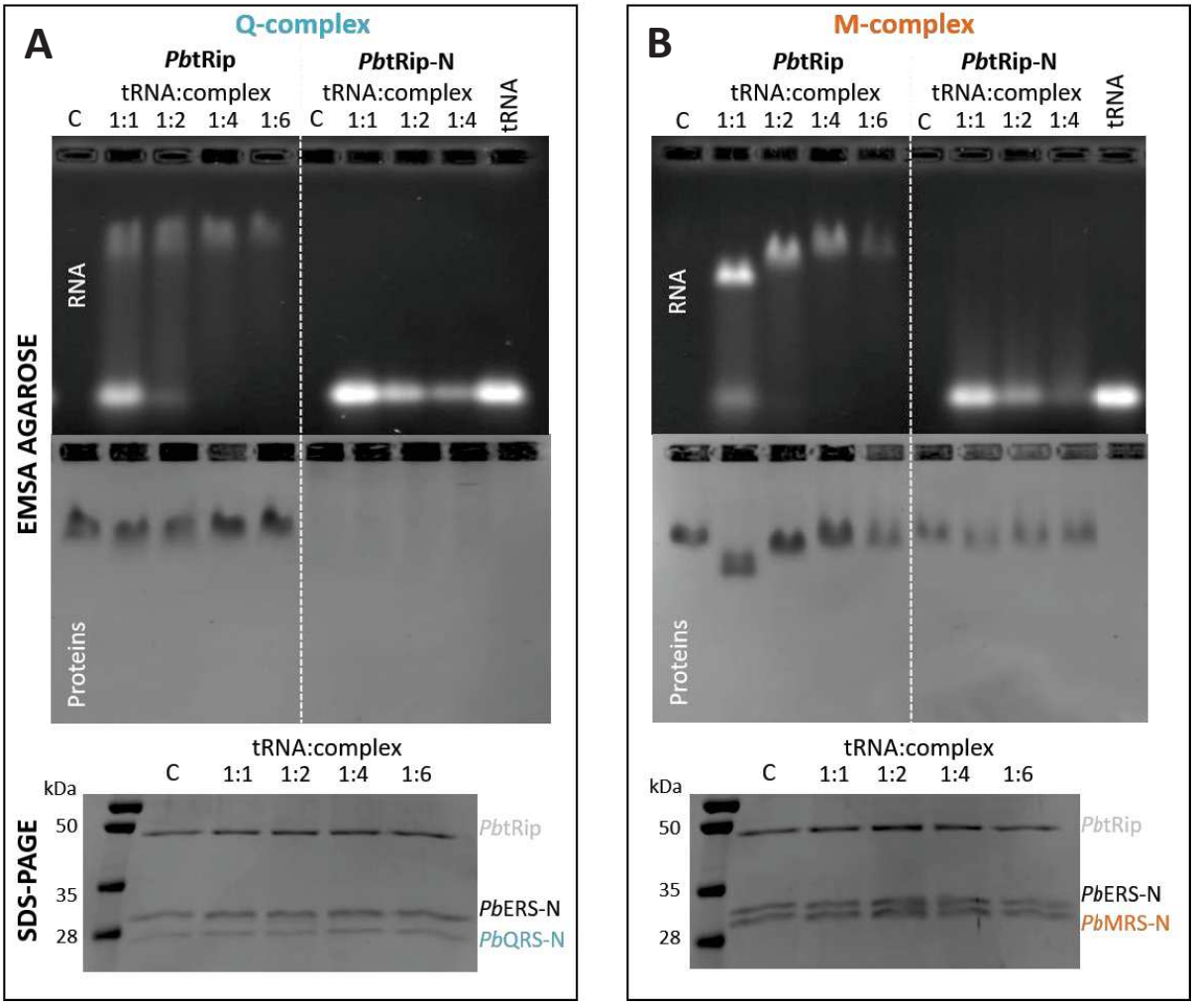


Figure 32. tRNA binding properties of ternary complexes. Different ratios of total yeast tRNA:complex were tested using either the Q-complex **(A)** or the M-complex **(B)**. Two different forms of each complex were used, either with the *Pb*tRip or with *Pb*tRip-N lacking the C-terminal domain. The different populations of molecules were separated on 1% agarose gel and stained with ethidium bromide (RNA) and Coomassie blue (proteins). Each band was sliced from the gel and analyzed by SDS-PAGE. Note that the Q-complex lacking the C-terminal domain of *Pb*tRip did not enter the gel, probably due to positive charges at pH 8.1 in TBE electrophoresis buffer.

2.3. *Pb*tRip:*Pb*ERS-N, the backbone of *Pb*MARS complexes

The two *Pb*MARS ternary complexes share a common feature: the binary complex formed by *Pb*tRip:*Pb*ERS-N. To investigate this complex in more details, both *Pb*tRip and *Pb*ERS-N were individually purified and used to reconstitute a complex under different conditions. In both cases, purification yielded high amounts of pure and homogenous proteins (5 mg from 250 mL culture).

2.3.1. Purification of *Pb*tRip and *Pb*ERS-N

Purification of *P. falciparum* tRip has been previously optimized by the team and these conditions were initially used for the *P. berghei* protein. However, this protein showed tendency to aggregate and some modifications of the protocol were necessary. For instance, *Pf*tRip can be conserved in a buffer containing as little as 75 mM KCl, but *Pb*tRip required more salt to be homogenous. Likewise, *Pb*ERS-N precipitated in the presence of low salt concentrations, but this precipitation was reversible by increasing the NaCl concentration. Thus, the concentration of NaCl in buffers to purify *Pb*tRip, *Pb*tRip-N and *Pb*ERS-N was fixed to 300 mM.

Purifications were performed in 2 steps, a Ni-NTA-affinity chromatography followed by a SEC column (See Material and Methods). During the Ni-NTA column, a washing step with a gradient of NaCl was performed to remove nucleic acid contaminations (Figure 33). This cleaning step was particularly efficient in the purification of *Pb*ERS-N, where UV-absorbing material was released from the column. Since SDS-PAGE analysis revealed only little of protein in the corresponding fractions (Figure 33B, NaCl wash), this suggests that nucleic acids were the main contaminants. This nonspecific binding was attributed to the high isoelectric point of *Pb*ERS-N (pI = 9.38, Table 13 in methods).

Despite the higher NaCl concentration in all buffers, some traces of aggregation were still observed on the *Pb*tRip SEC chromatogram (Figure 33A). *Pb*tRip and *Pb*tRip-N both eluted as a unique peak in the SEC column and their respective apparent MWs were 145 kDa and 63 kDa. These values are consistent with dimers of *Pb*tRip (96 kDa) and *Pb*tRip-N (50 kDa). They are slightly overestimated for *Pb*tRip, suggesting an elongated shape since the 2 domains (GST and EMAPII-like) are separated by a linker. It is also consistent with previous observations by Bour et al. (2016), showing the dimeric nature of the GST domain of *P. falciparum* tRip and the determination of its crystal structure by S. Gupta et al. (2020).

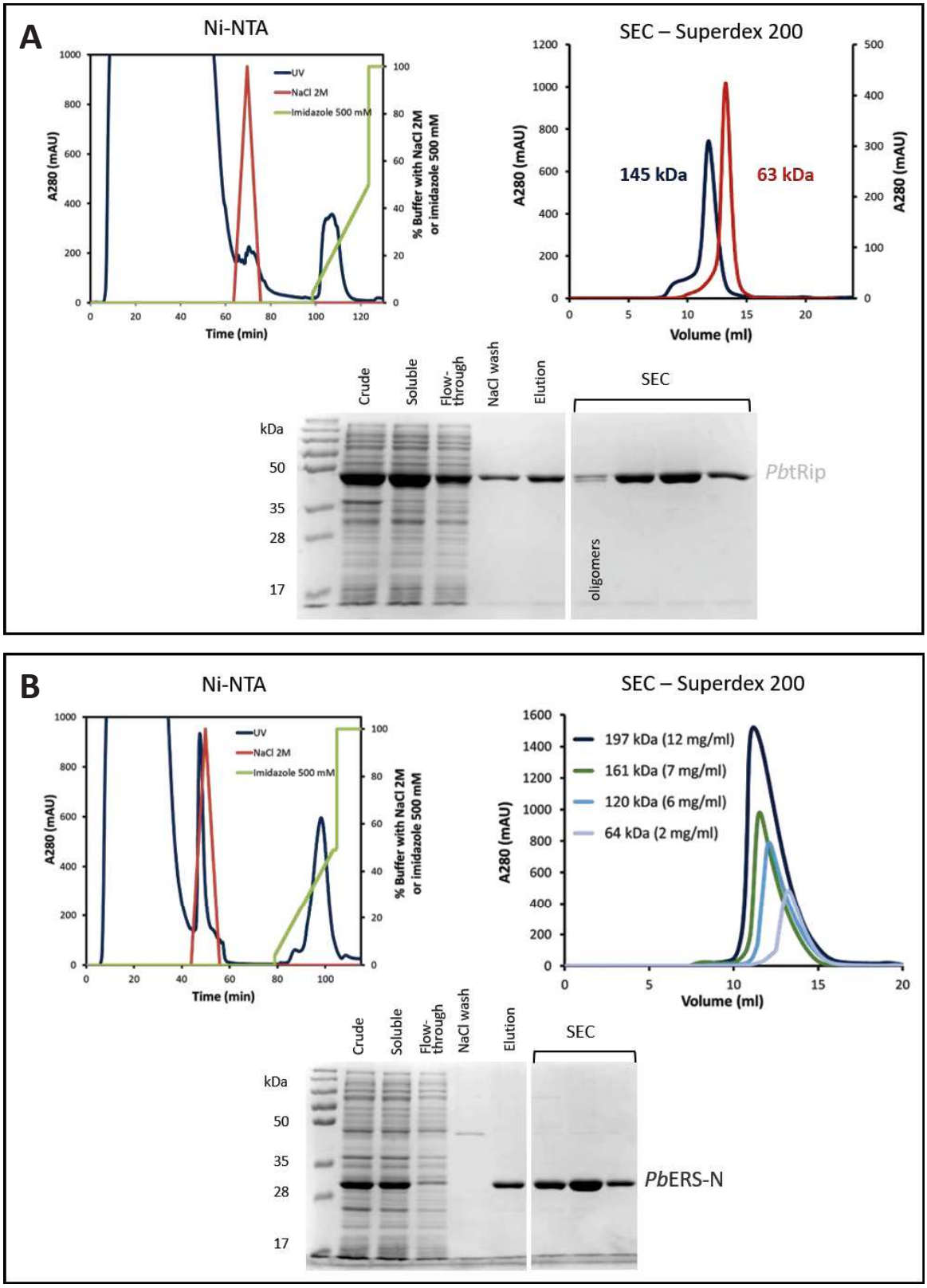


Figure 33. Purification of (A) *Pb*tRip, *Pb*tRip-N and (B) *Pb*ERS-N proteins. Only two chromatographic steps are performed when 6His-tag removal is not required. During the Ni-NTA-affinity chromatography, immobilized proteins are washed with a NaCl gradient (red line) and then eluted with a gradient of imidazole (green line). Elution profiles of *Pb*tRip (dark blue), *Pb*tRip-N (red) and different concentrations of *Pb*ERS-N (different shades of blue) on a Superdex 200 column are shown. The apparent MWs are indicated and SDS-PAGE corresponding to the two purification steps are shown.

*Pb*ERS-N showed a different behavior on the SEC column; the elution volume of the protein was dependent on the concentration at which the sample was applied (Figure 33B). At the lowest concentrations, *Pb*ERS-N (MW = 30 kDa) appeared as a dimer (64 kDa). At higher concentrations, the protein eluted at positions corresponding to higher oligomers (120 kDa, 161 kDa and 197 kDa). Moreover, SEC peaks in concentrated samples exhibited significant tailing, suggesting a mixture of oligomeric species. The same behavior was observed with PbERS-N-SUMO (data not shown). The ability of *Pb*ERS-N to form oligomers was also evidenced in pull-down experiments, in which the protein was able to interact with itself.

2.3.2. Together, *Pb*tRip and *Pb*ERS-N precipitate irreversibly

Both *Pb*tRip and *Pb*ERS-N are stable molecules in solution and can be stored for several days at 4°C. However, when *Pb*tRip is mixed with *Pb*ERS-N and loaded on a SEC column, they both eluted in the column void volume (Figure 34). This happened also with *Pb*tRip-N and *Pb*ERS-N-SUMO, suggesting that the tRNA-binding domain of tRip is not implicated in aggregation. Numerous chromatographic conditions were tested to stabilize/solubilize the complex *Pb*tRip:*Pb*ERS-N (e.g. pH, salt concentration, detergents, addition of tRNA), but none of them allowed to obtain convincing results. For instance, addition of 30% glycerol only delayed the time before the apparition of precipitates.

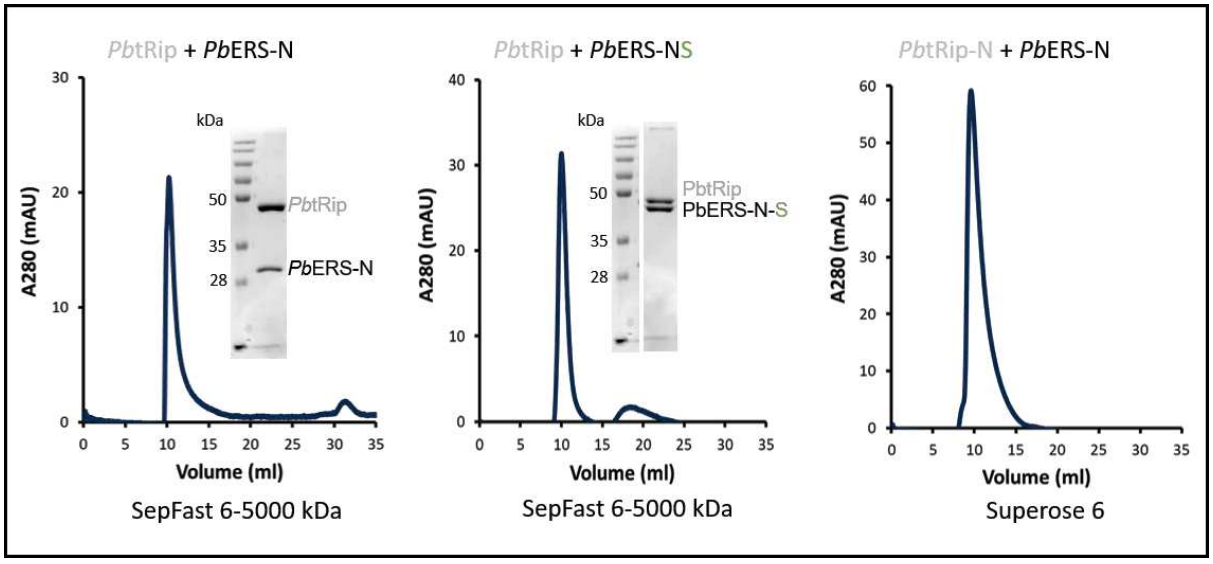


Figure 34. Reconstitution of *Pb*tRip:*Pb*ERS-N from individual partners. SEC profile of equimolar mixtures of *Pb*tRip + *Pb*ERS-N, *Pb*tRip + *Pb*ERS-N-SUMO and *Pb*tRip-N + *Pb*ERS-N are shown, either on SepFast 6-5000 kDa column (void volume ~9.5 mL) or on a Superose 6 column (void volume ~8.2 mL). For the two first conditions, the peak was analyzed by SDS-PAGE.

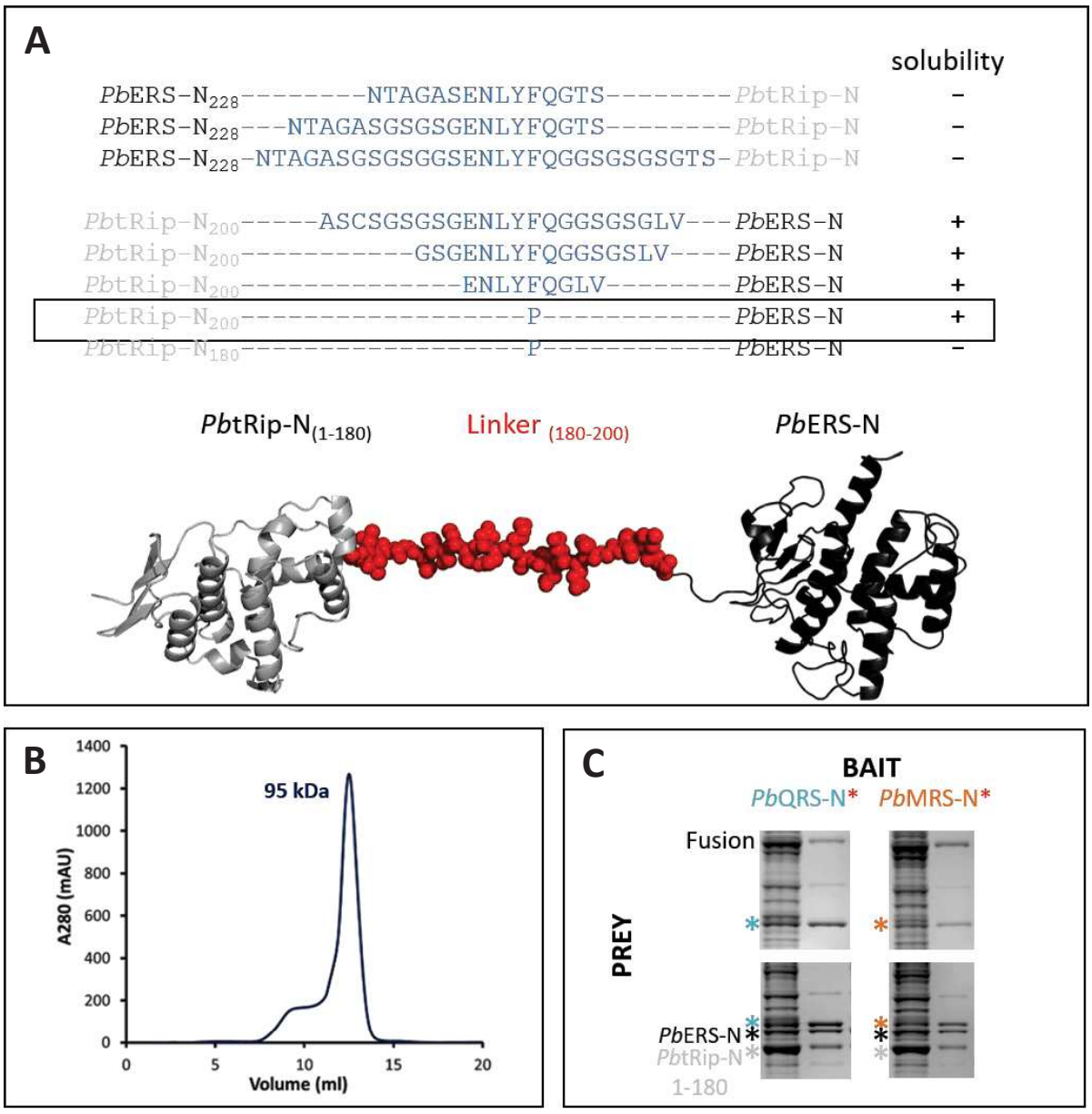


Figure 35. Chimeric protein *Pb*tRip-N-*Pb*ERS-N. A. Different designs and modeling of the fusion proteins. Different constructs with *Pb*ERS-N fused to *Pb*tRip-N or *Pb*tRip-N fused to *Pb*ERS-N were tested for their expression and solubility in *E. coli*. The shortest soluble construct is framed and modelized (Raptor X). *Pb*tRip (PDB 5ZKF) and *Pb*ERS-N (Raptor X) are represented in cartoon and colored in grey and black respectively. The last 20 residues in the *Pb*tRip-N moiety are predicted to be disordered and are represented with red spheres. **B.** Elution profile of *Pb*tRip-N-*Pb*ERS-N fusion protein on a Superdex 200 column. The apparent MW is indicated and would correspond to a dimeric protein. **C.** Pull-down experiments using *Pb*MRS-N or *Pb*QRS-N as bait and fused or unfused *Pb*tRip-N and *Pb*ERS-N as preys.

2.3.3. The *Pb*tRip-N-*Pb*ERS-N fusion

A way to prevent aggregation of *Pb*tRip-N and *Pb*ERS-N was to fuse both polypeptides into a unique chimeric protein. Several constructions were engineered and expressed in *E. coli* (Figure 35A), but only those in which *Pb*tRip-N was fused to the N-terminus of *Pb*ERS-N showed solubility. The size of the linker between the 2 proteins was not important as long as the C-terminal sequence of *Pb*tRip-N covered amino acids 180 to 200. Since these last 20 amino acids are predicted to be disordered, they probably act as a flexible linker and are sufficient to accommodate the *Pb*ERS-N moiety in the fusion without steric hindrance.

The fusion protein with no linker (only a proline residue between the two peptides) was produced, purified by Ni-NTA-affinity chromatography and analyzed on SEC (Figure 35B). It eluted with an apparent MW of about 100 kDa on the SEC column, suggesting that it dimerizes in solution. Moreover, the elution profile recalled those of individual proteins (Figure 33). For instance, some fronting (as *Pb*tRip) and tailing (as *Pb*ERS-N) can be noted in the shape of the chromatogram, suggesting small amounts of different oligomeric species. The protein did not precipitate and remained homogenous for several weeks at 4°C.

Functionally, the fusion protein behaves like individual *Pb*tRip-N and PbERS-N in pull-down experiments and interacts with both *Pb*MRS-N or *Pb*QRS-N (Figure 35C). Although this fusion protein needs to be further characterized, especially in its ability to interact with *Pb*MRS-N and *Pb*QRS-N (apparent size of the complexes, oligomerization, etc.), it is a good candidate to produce and crystallize the binary complex *Pb*tRip-N:*Pb*ERS-N.

V. Reconstructing *Pb*MARS complexes from *Pb*ERS-N structure

Beyond the identity of the domains implicated, knowledge of the structure and organization of *Pb*MARS complexes should provide valuable information on their functional role, especially in the context of tRNA import into the parasite. X-ray crystallography was the first approach chosen for several reasons. Ternary complexes, particularly those lacking the C-terminal domain of *Pb*tRip, are stable and remain homogenous for several days (as assessed by DLS) and the host laboratory is equipped with a nanoliter pipetting robot which allows to multiply the number of experiments. Indeed, several commercial and customized crystallization screens, each one containing 96 different conditions, were tested at least once for several forms of ternary complexes (Table 17 in Methods). In spite of these advantageous conditions and extensive efforts, no crystal of any complex was obtained during my PhD work. However, several encouraging results were observed and I believe that crystallization of these complexes is possible, but still requires more work. Although Cryo-EM is a powerful alternative, especially for complexes with full-length *Pb*tRip and aaRSs, we decided to use instead an integrative approach and combine information from different experiments to propose a model based on the crystal structures of individual partners, multiple scattering measurements (DLS/SLS and SEC-SAXS) and further pull-down experiments with proteins mutated at potential interactions sites.

1. *Pb*ERS-N, a plastic structure

1.1. DLS analysis

The ability of *Pb*ERS-N to oligomerize was first shown by SEC analysis. It was further confirmed by multiple DLS/SLS measurements at different concentrations (Figure 36). The hydrodynamic diameter (d_h) of particle increased linearly with protein concentration, from ~10 nm (2 mg.mL^{-1}) to ~12.5 nm (6 mg.mL^{-1}). Estimates of MW, either from DLS (MW-R) or SLS (MW-S), behaved in the same way. At the lowest concentration, *Pb*ERS-N appeared as a tetramer (MW-S ~125 kDa) and the MW increased linearly with protein concentration, indicating sequential association of *Pb*ERS-N molecules to form larger oligomeric assemblies. Regardless of the concentration, the MW-R/MW-S ratio stays the same ~1.2 and extrapolation at zero concentration gave a d_h of 8.8 nm and MW estimates of 96 (MW-R) and 88 (MW-S) kDa. Therefore, at low concentration, *Pb*ERS-N is an elongated dimer (60 kDa), as it is expected for most GST proteins. Although no aggregates were detected in the regularization analysis of DLS data, the polydispersity of samples was larger than 20%, suggesting a mixture of oligomeric states of similar dimensions.

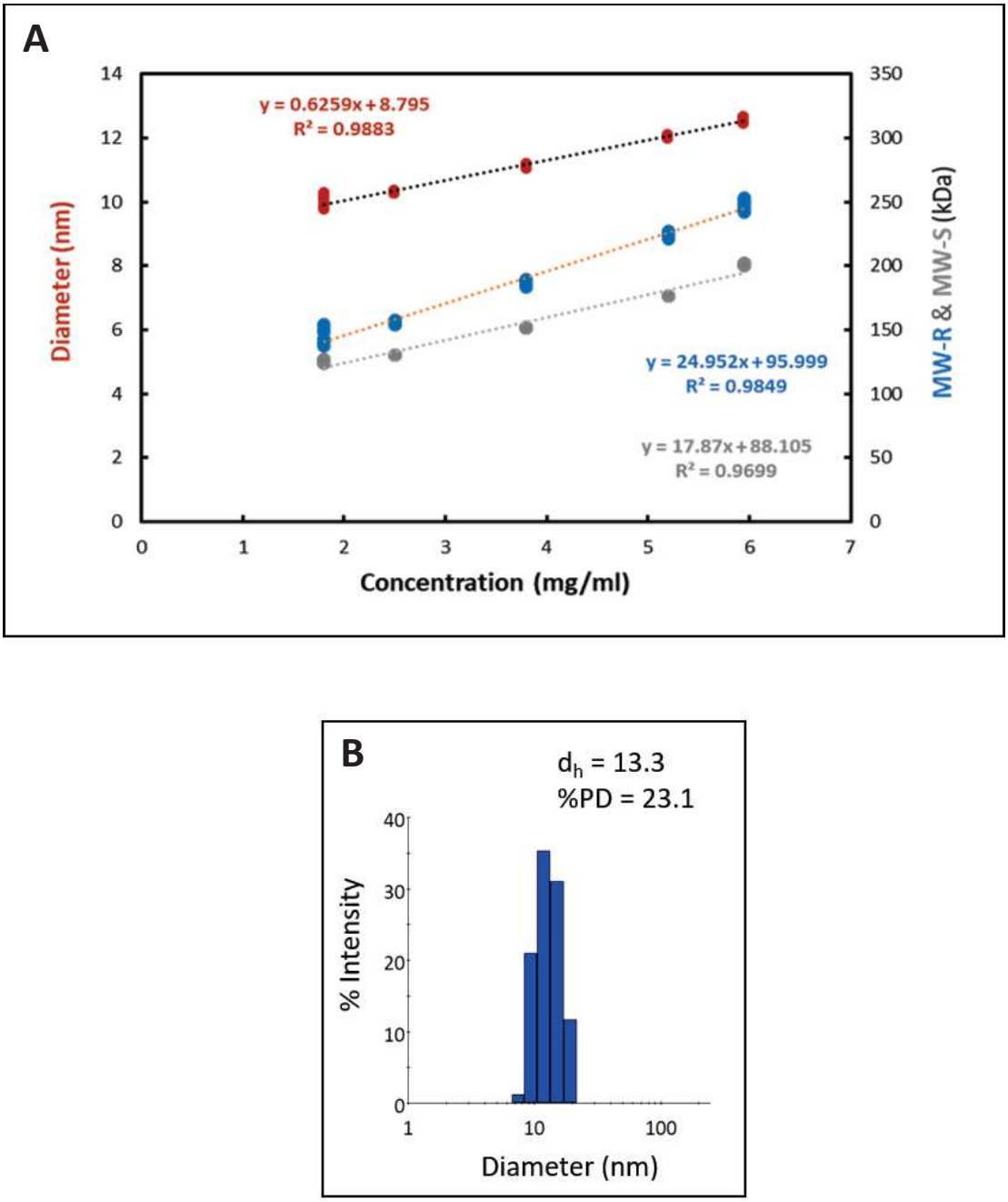


Figure 36. Oligomerization properties of *Pb*ERS-N. **A. Cumulant analysis.** DLS/SLS measurements of *Pb*ERS-N were performed at different concentrations. The diameter (red dots) (left y-axis) and molecular weight estimates from DLS (MW-R, blue dots) and SLS (MW-S, grey dots) (right y-axis) are plotted as a function of the concentration. Ten DLS/SLS measurements are shown for each concentration, most of them overlapping into a single point. A linear regression is performed for each parameter in order to verify linear dependency and obtain values at zero concentration. In each case, the equation of the regression line and its coefficient of determination (R^2) are indicated. **B. Regularization graph.** DLS measurement of a sample of *Pb*ERS-N (6 $mg.mL^{-1}$). The hydrodynamic diameter (d_h) and percentage of polydispersity (%PD) are shown. No aggregates were detected at any measured concentration.

1.2. SEC-SAXS analysis

SEC-SAXS data are consistent with DLS in the sense that *Pb*ERS-N exists as a mixture of different oligomeric states in solution. The protein was analyzed in a buffer containing 1M NaCl in order to reduce particle interactions as much as possible. The SAXS chromatogram exhibited the characteristic tailing of *Pb*ERS-N (Figure 37A, also seen in Figure 33B) and the Rg (not shown) and MW estimates varied a lot across the peak. In the most concentrated part of the peak, MW estimates (obtained from the volume of correlation, Vc) correlated with a tetrameric *Pb*ERS-N (~120 kDa) whereas the estimations went down to 60 kDa at the end of the peak, corresponding to a dimeric *Pb*ERS-N. Despite this heterogeneity, a SAXS curve for the tetramer could be obtained and *ab initio* bead models evidenced an elongated shape (Figure 37B).

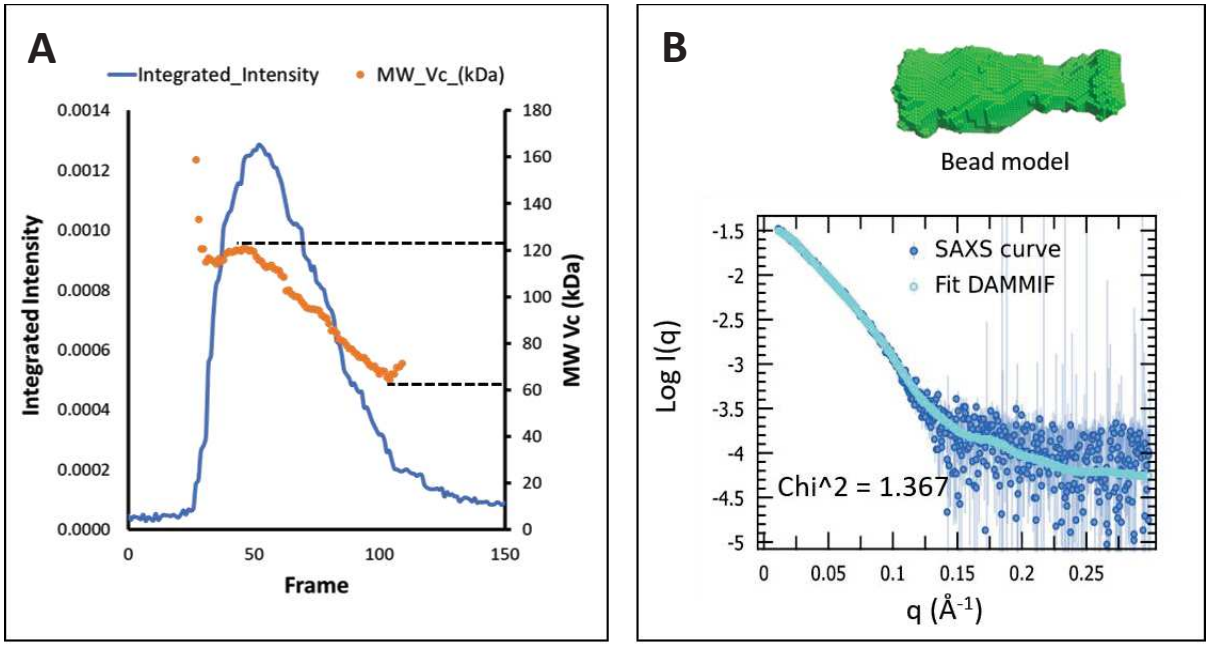


Figure 37. SEC-SAXS analysis of *Pb*ERS-N. A sample (4.3 mg.mL^{-1}) was subjected to SEC in a high salt buffer (1M NaCl) and multiple scattering frames were collected as the protein eluted from the column. **A. SAXS chromatogram.** The integrated intensity (blue line) and the MW (orange dots) estimated from the correlation volume as a function of the frame number (i.e. elution volume) are shown. Frames with a stable Rg values are indicated with a green bar. **B. SAXS curve and 3D reconstruction.** The SAXS curve corresponding to the tetrameric *Pb*ERS-N and *ab initio* 3D reconstruction using DAMMIF are shown; AMBIMETER indicated a potentially unique solution.

1.3. *Pb*ERS-N displaces all other GST domains in pull-down experiments

*Pb*ERS-N was tested for its ability to replace any of the other components in the *Pb*MARS complexes using pull-down assays. Interestingly, *Pb*ERS-N was able to displace any partner in the different complexes tested. For example, addition of increasing amounts of prey *Pb*ERS-N to the M-complex resulted in the capture of this protein on the resin at the expense of *Pb*MRS-N (Figure 38, bottom). This might suggest that *Pb*MRS-N and *Pb*ERS-N were in competition for the same binding site on the complex *Pb*tRip:*Pb*ERS-N. Similar observations were made for the Q-complex and the binary complex *Pb*tRip:*Pb*ERS-N (Figure 38, top and middle), where increasing prey *Pb*ERS-N decreased the captured amount of *Pb*QRS-N and *Pb*tRip, respectively.

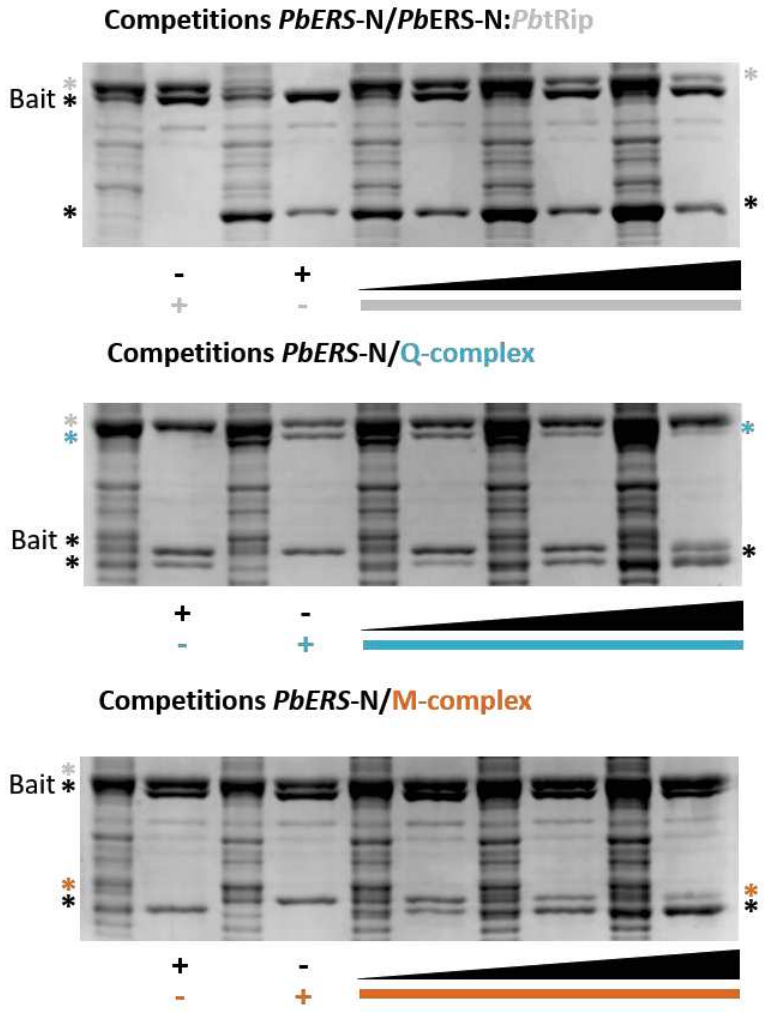


Figure 38. Competition between *Pb*ERS-N and the three other GST domains. Series of pull-down assays where *Pb*ERS-N (with or without SUMO) was used as a bait to capture *Pb*tRip (top, formation of a binary complex), or *Pb*tRip and *Pb*QRS-N (middle, formation of Q-complex) or *Pb*tRip and *Pb*MRS-N (bottom, formation of M-complex). All complexes were challenged with increasing concentrations of prey *Pb*ERS-N. Molecules in competition are indicated by asterisks on the right of the gels.

2. Crystal structure of *Pb*ERS-N

2.1. Crystallization and X-ray analysis

Of all the crystallization tests performed with the different complexes or individual proteins, the only domain that yielded results was the N-terminal domain of *Pb*ERS. Despite its dynamic behavior in solution, *Pb*ERS-N was crystallized in ammonium sulfate conditions. However, crystals were very small and too many of them appeared in a single drop, indicating excessive nucleation (Figure 39A). Addition of 0.5% PEG helped to reduce nucleation, but the size of crystals remained relatively small (< 50 µm) (Figure 39B). While different approaches were explored (e.g. crystallization in gel), only the seeding technique resulted in larger crystals (> 100 µm) with good morphology (Figure 39C). Unfortunately, these conditions were not compatible with cryo-cooling, since X-ray analysis of such crystals directly frozen in liquid nitrogen only showed some diffraction spots at low resolution, and the presence of ice rings was evident. Consequently, different cryoprotectants were explored, including glycerol, sugars, oils, salts and other mixtures (Rubinson et al., 2000; Vera & Stura, 2014; Senda et al., 2016). Cryoprotection with glycerol improved diffraction, but crystals often cracked during soaking. Thus, to skip the soaking step, crystals were directly grown in ammonium sulfate containing glycerol. Moreover, the anti-nucleation properties of glycerol helped to control excessive nucleation of *Pb*ERS-N, the growth rate of crystals was reduced and better crystals were obtained.

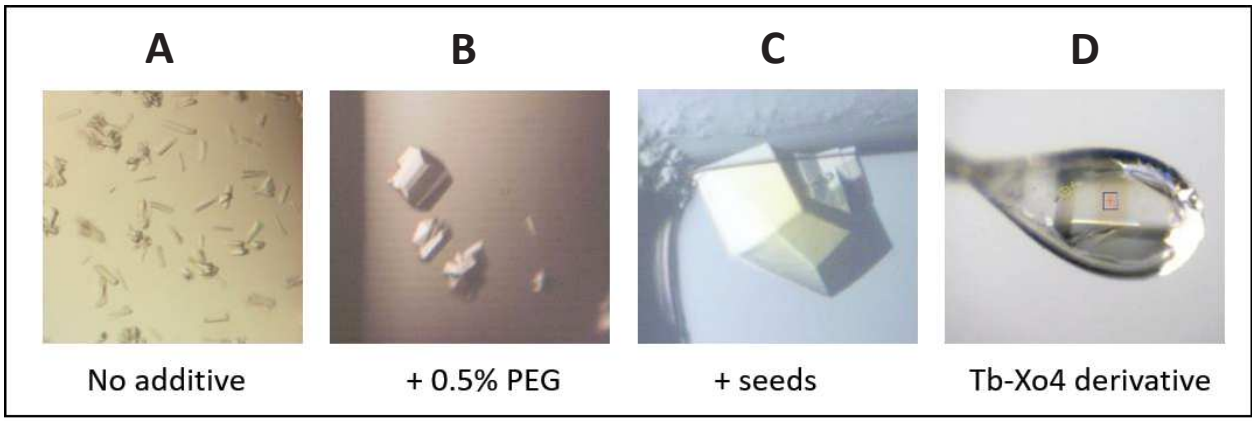


Figure 39. Optimization of crystallization conditions of *Pb*ERS-N. A. Only ammonium sulfate. Many small crystals of *Pb*ERS-N are obtained in 1.4-1.5 M ammonium sulfate and 0.1M HEPES-NaOH pH 7.5. **B. Effect of PEG and glycerol.** Nucleation is reduced by using additives such as 0.5% PEG or 10-20% glycerol. The latter is preferred because it provides also cryoprotection. **C. Effect of seeding.** Addition of seeds in the protein sample is necessary to obtain optimal crystals. **D. Tb-Xo4 derivative.** Crystal derivatized with the nucleating and phasing agent, Tb-Xo4 (after X-ray analysis).

A first complete data set was obtained at 2.73 Å of resolution. Despite the availability of crystal structures of GST domains from other ERSs (Simader, Hothorn & Suck, 2006; Cho et al., 2015), molecular replacement (MR) was unsuccessful. Analysis of solvent content suggested multiple copies of the protein (4 to 7) in the asymmetric unit. The self-rotation function predicted 5 molecules in the asymmetric unit (Figure 64 in Methods). These observations are consistent with SEC and DLS/SLS analysis, in which *Pb*ERS-N appeared as a hexamer at high concentrations. In order to solve the phase problem, crystals derivatized with the crystallophore Tb-Xo4 (Engilberge et al., 2017) were prepared and analyzed at the terbium L_{III} absorption edge (Figure 39D). Although these crystals diffracted only at 3.1 Å of resolution (at best), multiple data sets with high redundancy could be collected and merged to obtain an exploitable SAD (Single-wavelength Anomalous Diffraction) signal. Additionally, several native data sets at resolutions slightly better than 2.73 Å were also obtained. Together, these data provided an interpretable density map, which was further refined to obtain acceptable R_{work} and R_{free} values (Table 4). Most amino acids have been placed in the map, but some of them are still missing and structure refinement is still in progress.

Table 4. Current model (10/21/2020) refinement statistics of *Pb*ERS-N

Data set code	ERS2-1_refine24
Resolution range	44.34 - 2.703
Highest-resolution shell	2.799 - 2.703
Reflections used in refinement	50687 (4900)
Reflections used for R-free	2532 (245)
R-work	0.2085 (0.3455)
R-free	0.2504 (0.3610)
CC (work)	0.956 (0.850)
CC (free)	0.927 (0.802)
Solvent content (%)	61.2
V_M (A^3/Da)	3.26
Number of non-hydrogen atoms	7884
macromolecules	7874
ligands	10
water	0
Protein residues	923
RMS(bonds)	0.003 Å
RMS(angles)	0.58°
Ramachandran favored (%)	91.48
Ramachandran allowed (%)	6.82
Ramachandran outliers (%)	1.7
Rotamer outliers (%)	9.66
Clashscore	6.63
Average B-factor ($Å^2$)	116.85
macromolecules	116.84
ligands	129.43
Number of TLS groups	29

2.2. Structure analysis

2.2.1. *Pb*ERS-N is a classical GST domain

The asymmetric unit is constituted of 5 molecules of *Pb*ERS-N. Each monomer adopts a GST-fold consisting of the two well-known subdomains (Figure 40): the N-terminal subdomain contains the four β-strands of a canonical thioredoxin-fold but lacks one of the α-helices (between β2 and β3) that flanks this structure. The C-terminal subdomain adopts an α-helical structure (α3 to α8) composed of a central helix (α5) surrounded by 5 other helices. Except for helix α7, all of them are parallel one to each other. The central helix α5 is mostly composed of hydrophobic residues and bends at its C-terminal extremity; it exhibits the N-capping box (S/T-X-X-D), which is strictly conserved in all GST proteins and is crucial for the stability of the fold (Aceto et al., 1997). Most of the missing residues in the current structure are localized in the thioredoxin-fold. About 15 of these residues are localized between strands β3 and β4 (Figures 40A and 40B); they correspond to the LCR identified in the bioinformatics analysis (Figure 22A). The fact that these residues are not visible in the crystal confirms the disordered nature of this insertion. As expected, the LCR appeared between 2 structured elements and its presence does not affect the tertiary structure of the GST domain.

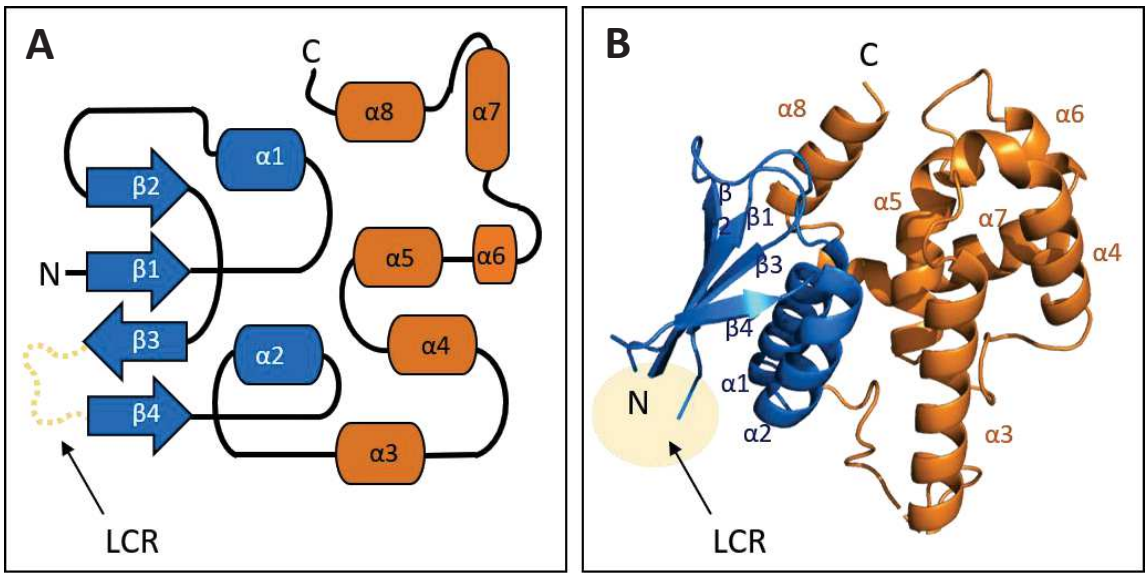


Figure 40: Crystal structure of *Pb*ERS-N. The thioredoxin subdomain is colored in blue and the C-terminal helical subdomain in gold. All secondary structures (α-helices and β-strands) are identified in the structure. The N- and C-terminus are indicated. **A. Topology diagram of *Pb*ERS-N.** Compared to a canonical GST fold, *Pb*ERS-N is missing a helix between β2 and β3 and contains an LCR between β3 and β4. **B. *Pb*ERS-N crystal structure**. The thioredoxin subdomain is constituted of β1-α1-β2-β3-β4-α2. The C-terminal subdomain includes helices α3 to α8 and helix α7 is oriented perpendicularly. The position of the missing LCR is highlighted in pale yellow.

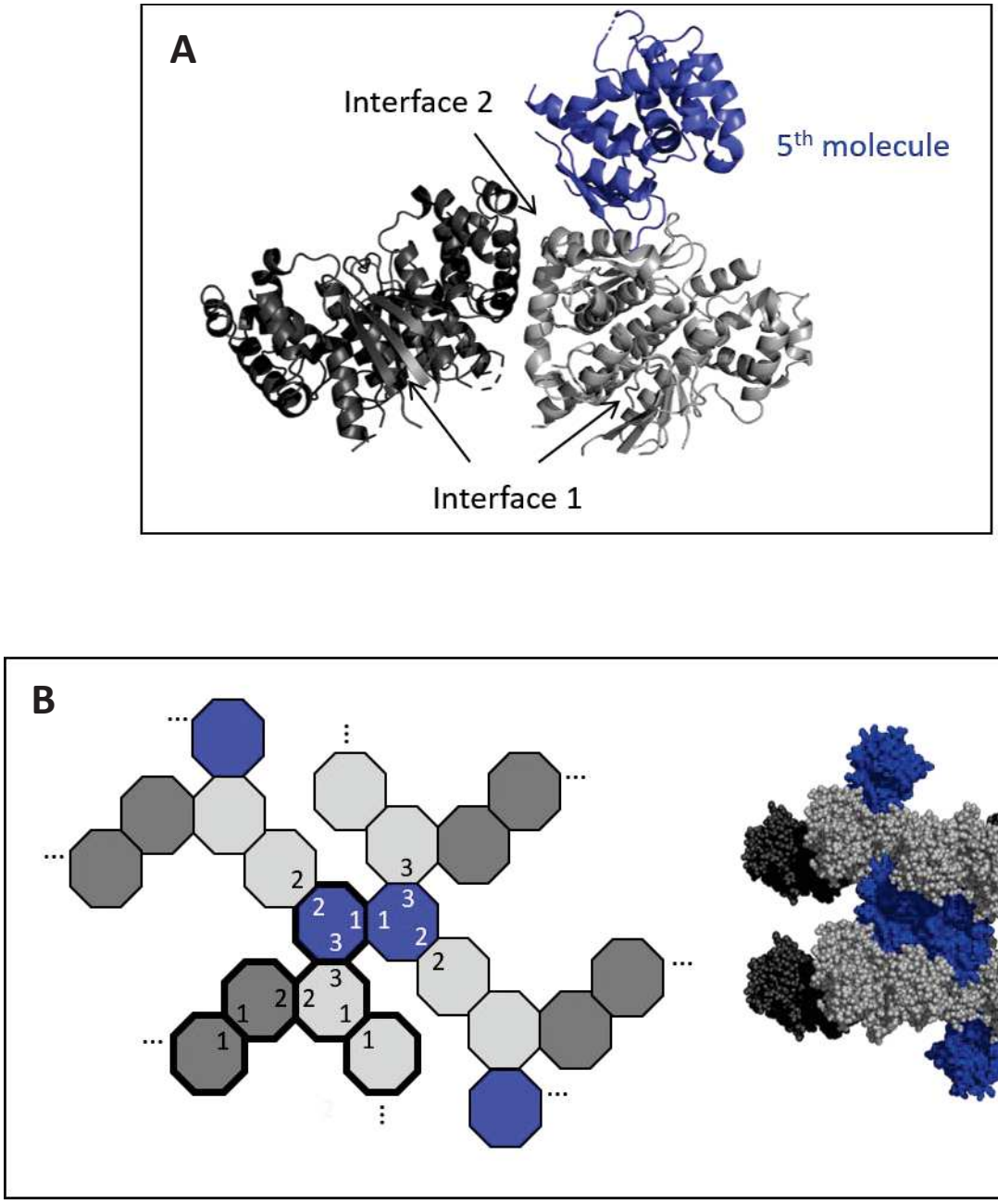


Figure 41. Organization of *Pb*ERS-N in the crystal. A. The asymmetric unit. The five molecules in the asymmetric unit are organized as two canonical GST dimers (dark and light gray) (interface 1), which interact to form a tetramer (interface 2) to which a 5th molecule of *Pb*ERS-N (blue) binds. **B. Arrangement of molecules in the crystal.** Several layers of tetramers (dark and light gray) are linked by dimers (blue). Each dimer is formed by two "5th molecules" from adjacent asymmetric units, which interact through interface 1. Each dimer is able to bind 4 tetramers of *Pb*ERS-N, two of them are bound through interfaces 2 and the two others through a novel interface 3. The molecules in the crystal form continuous chains alternating between interfaces 1 and 2, which are linked through interfaces 3 (dots). One of the asymmetric units is highlighted with ticker lines.

2.2.2. *Pb*ERS-N reveals a new interface for GST domain interaction

In the asymmetric unit, two dimers of *Pb*ERS-N form a tetramer and a 5th monomer contacts one of these dimers (Figure 41A). The crystal is constituted of different layers of tetramers of *Pb*ERS-N, which are linked by dimers (Figure 41B). The presence of different oligomeric assemblies in the crystal structure is consistent with the ability of this protein to oligomerize, as it was evidenced in pull-down, DLS/SLS and SEC-SAXS experiments.

Several interactions are observed between the molecules in the asymmetric unit:

Interface 1: A canonical GST dimerization interface is observed in the dimer of ERS-N. The helices α2 and α3 of one monomer interact respectively with helices α3 and α2 of the second monomer, and all of them are parallel one to each other. The area of this interface is ~1200 Å^2, which is similar to values observed in catalytically active GST enzymes (Dirr et al., 1994). Binding is mainly mediated by van der Waals forces and assisted by several hydrogen bonds (Figure 42A).

Interface 2: Two dimers of *Pb*ERS-N form a tetramer through an interface involving the helix α7 and the loop between helices α4 and α5 of each interacting monomer. The surface of this interface is ~780 Å^2 and two characteristic arginines protruding from helices α7 stack on each other (Figure 42B).

Interface 3: The 5th molecule of *Pb*ERS-N is inserted at the proximity of interface 2. The area of this "novel" interface is ~380 Å^2 and involves helix α8 and the loop between strands β2 and β3 of the 5th molecule and the helices α3 and α4 of one of the subunits in the tetramer (Figure 42C).

The interaction interfaces observed in dimers and tetramers of *Pb*ERS-N are observed in other GST domains. The elongation factor 1Bγ from yeast and the AIMP3 from the human MARS complex also crystallized as canonical GST dimers (interface 1) which further tetramerize through an interface 2. However, unlike PbERS-N, these proteins behave as monomers in solution and do not oligomerize (Jeppesen et al., 2003; K.J. Kim et al., 2008). On the other hand, to my knowledge, the binding site of the 5th molecule of ERS-N in the asymmetric unit is not observed in other GST domains. Although this contact might be an artifact of crystal packing, we cannot exclude a biologically relevant interface. Indeed, in the human MARS complex, a fragment of the catalytic core of DRS binds a similar region on the GST domain of AIMP2 (Cho et al., 2019) (compare blue molecules in Figures 44B and 44C).

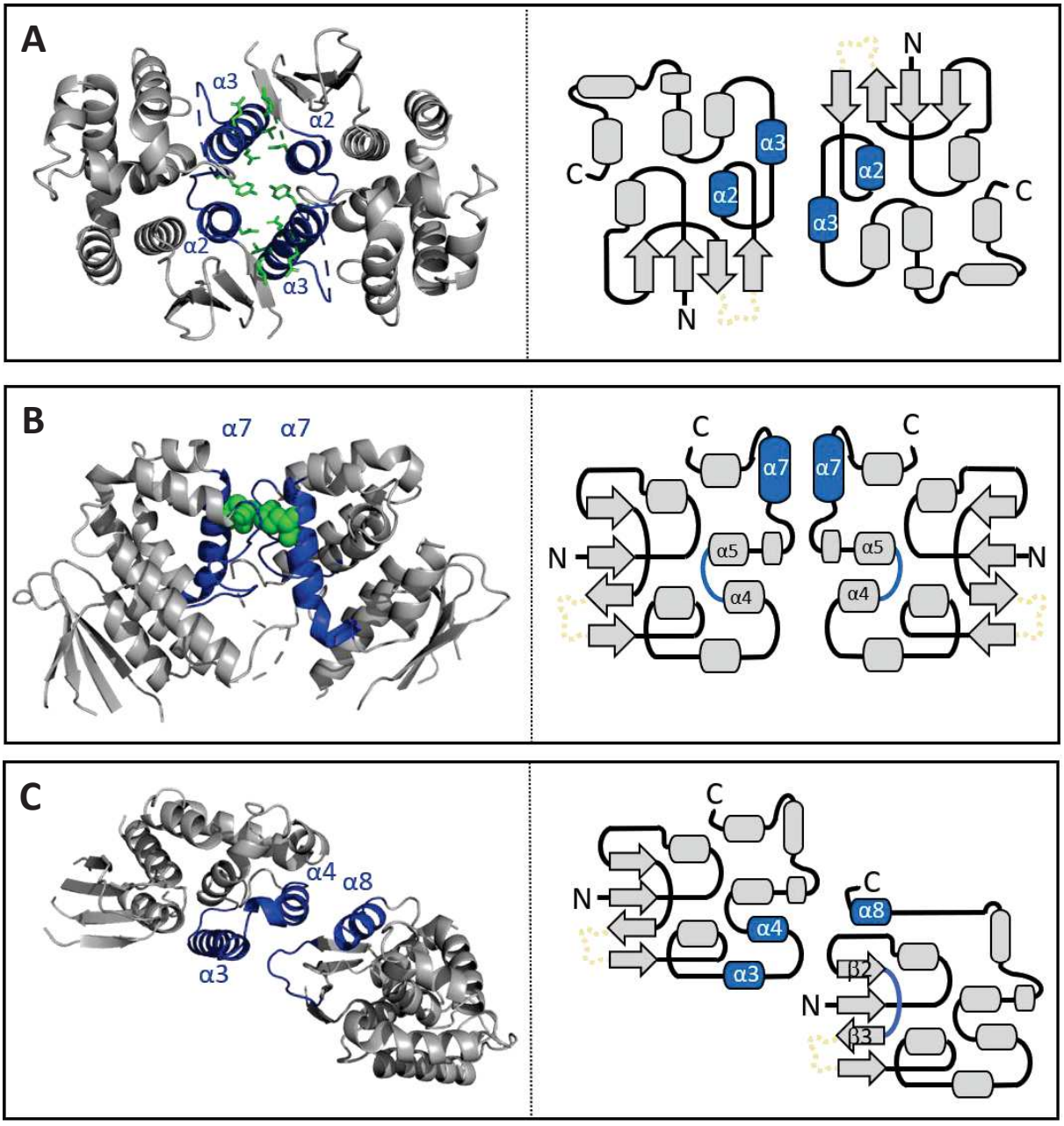


Figure 42. Interfaces between the *Pb*ERS-N molecules in the asymmetric unit. Three types of interactions are observed. **A. Interface 1.** Two monomers of *Pb*ERS-N form a canonical GST dimer by interacting through their helices α2 and α3 (blue). Residues involved in the formation of hydrogen bonds are highlighted as green sticks. **B. Interface 2.** Tetramerization occurs through helices α7 (blue) and the loop α4-α5 (blue) of one monomer from each dimer. A stacking interaction between 2 arginine residues is observed at this interface (green spheres). **C. Interface 3.** Helix α8 (blue) and the loop β2-β3 (blue) of the 5th molecule contact helices α3 and α4 (blue) of one of the monomers implicated in tetramerization through interface 2.

The crystal structure of tRip-N of *P. vivax* was recently solved by Gupta S. et al. (2020). As expected, the protein adopted a GST-fold and form 2 types of dimers in the crystals (Figure 43). The first one resembles interface 1, **but** helices α2 and α3 are arranged differently than in a canonical GST dimer. In *Pv*tRip-N, helices α2 and α3 of one monomer pack against α2 and α3 of the second monomer, and they are oriented perpendicularly to each other (compare Figures 43A and 42A). The consequence is that the two C-terminal ends are oriented in the same direction, a unique situation among GST dimers. This observation will be essential in the design of our model (see section 4.4). The size of the dimerization interface is ~867 Å^2, which is slightly lower when compared to canonical GST dimers. The second type of dimer observed involves interface 2: two monomers of *Pv*tRip-N interact through helices α7 and loops α4- α5 and the same stacking of arginines is observed (Figure 43B).

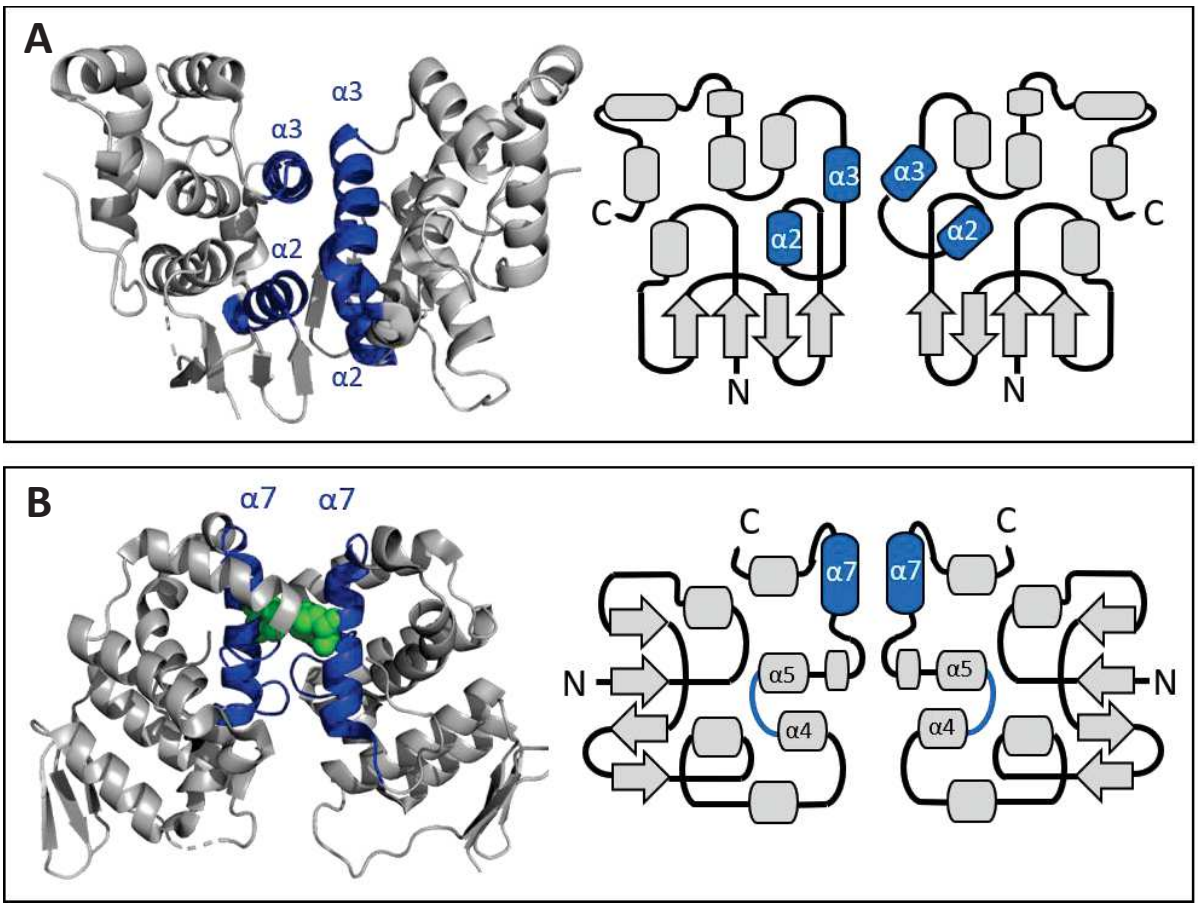
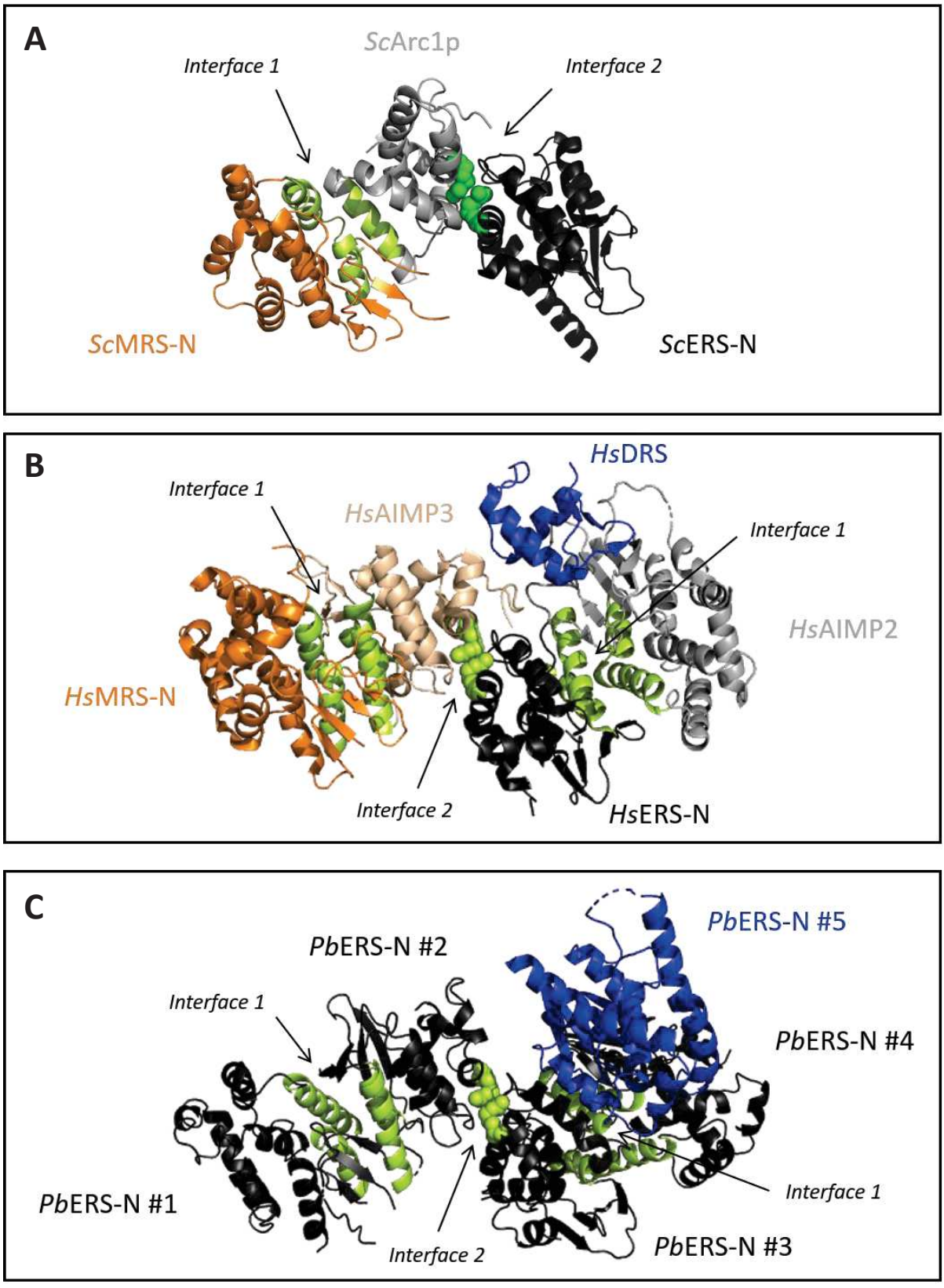


Figure 43: Crystal structure of *P. vivax* tRip-N. Two dimeric forms of *Pv*tRip-N are observed in the crystal structure (PDB ID 5ZKF). **A. Dimerization of *Pv*tRip-N through interface 1**. Contrary to a classical GST dimer, helices α2 and α3 (blue) of each monomer are oriented perpendicularly and contacts α2-α2' and α3-α3' are observed. As a consequence, the β strands of the thioredoxin subdomain are oriented on the same side of the dimer. **B. Dimerization of *Pv*tRip-N through interface 2.** The helix α7 and the loop between α4 and α5 (blue) are implicated in the interaction. 2 arginine residues from helices α7 stack one to another (green spheres).



Legend in next page

Figure 44: Comparison of MARS sub-complexes with the asymmetric unit of *Pb*ERS-N. The sub-complexes shown here involve mostly GST domains. Helices α2 and α3 (interface 1) are colored in green and the arginines from helices α7 (interface 2) are represented as green spheres. **A. The GST domains in the MARS complex from *S. cerevisiae*.** The AIMP Arc1p (gray) interacts simultaneously with MRS (orange) and ERS (black) using its interfaces 1 et 2, respectively. The structure shown here has been reconstituted from the crystal structures of the two sub-complexes *Sc*MRS:*Sc*Arc1p (PDB 2HSN) and *Sc*ERS:*Sc*Arc1p (PDB 2HRK). **B. The GST domains in the MARS complex from *H. sapiens*.** Cho et al. (2019) solved the crystal structure of a heterotetramer of GST domains from *Hs*MRS (orange), *Hs*AIMP3 (light brown), *Hs*EPRS (black) and *Hs*AIMP2 (gray) together with a fragment of the catalytic core of *Hs*DRS (blue) (PDB 5Y6L). **C. The asymmetric unit of the crystal of *Pb*ERS-N.** The tetramer of *Pb*ERS-N is colored in black and the 5th molecule of *Pb*ERS-N in blue.

2.2.3. Comparison with crystal structures of MARS sub-complexes

Several crystal structures of MARS sub-complexes with GST domains are available in the PDB. In all of them, the GST domains interact using the same two types of interfaces. For instance, in the yeast MARS complex, Arc1p interacts with the GST domain of MRS through interface 1 and with the GST domain of ERS through interface 2 (Simader, Hothorn, Köhler et al., 2006) (Figure 44A). Similarly, in the human MARS complex, AIMP3 interacts with the GST of MRS using interface 1 and with the GST of ERS using interface 2, whereas ERS uses its interface 1 to interact with the GST of AIMP2 (Figure 44B). This network of interactions allows the formation of a heterotetramer of GSTs, which constitutes the core of the human MARS complex (Cho et al., 2015; Hahn et al., 2019; Cho et al., 2019).

Interestingly, the homotetramer of *Pb*ERS-N is organized in similar way that the heterotetramer MRS:AIMP3:EPRS:AIMP2 in the human MARS complex (Figures 44B and 44C). Therefore, it is possible that *Pb*ERS-N in the crystal is mimicking the interactions that it would have with its partners *Pb*tRip and *Pb*QRS in the Q-complex and with *Pb*tRip and *Pb*MRS in the M-complex. Moreover, the fact that GST domains involved in MARS complexes are apparently not able to bind simultaneously more than 2 partners is consistent with the inability of *Pb*ERS-N to interact with both *Pb*QRS and *Pb*MRS while interacting with *Pb*tRip.

The yeast MARS complex and the *Pb*M-complex are constituted of the same homologous proteins (i.e. AIMP, ERS and MRS), but they are organized differently. Contrary to the AIMP *Sc*Arc1p, which binds independently and simultaneously *Sc*ERS and *Sc*MRS (Deinert et al., 2001) (Figure 44A), the AIMP *Pb*tRip in the *Pb*MARS complex only binds directly *Pb*ERS-N but not *Pb*MRS-N.

Table 5. Residues found at the interaction interfaces of MARS GST sub-complexes from *Homo sapiens* and *Saccharomyces cerevisiae*. Three GST heterodimers involving interfaces 1 [*Hs*MRS:*Hs*AIMP3 (PDB 4BVX), *Hs*EPRS:*Hs*AIMP2 (PDB 5A34) and *Sc*MRS:*Sc*Arc1p (PDB 2HSN)] and two involving interfaces 2 [*Hs*AIMP3:*Hs*EPRS (PDB 5BMU) and *Sc*Arc1p:*Sc*ERS (PDB 2HRK)] are shown. In each case, five couples of residues coming in close contact at the interaction interface of each dimer are listed. The secondary structure of the GST-fold to which they belong is indicated in parenthesis. Critical residues for interaction according to mutagenesis experiments from Simader, Hothorn, Köhler et al. (2006) and Cho et al. (2015) are colored in red.

H. sapiens		***S. cerevisiae***	
Interface 1			
MRS	**AIMP3**	**MRS**	**Arc1p**
(α2) S61, S63	(α3) E86	(α2) D62	(α3) S35
(α2) R67	(α3) Q72	(α2) A65	(α3) A28
(α2) A64	(α3) A69	(α2) Y69	(α3) E25
(α3) D79	(α2) K53	(α3) Q77	(α3) E34
(α3) E86	(α2) T45, T46	(α3) Q83	(α3) K38
EPRS	**AIMP2**		
(α2) R56	(α2) R215		
(α2) R60	(α3) D234		
(α2) R56	(α3) D238		
(α3) D79	(α2) R215		
(α3) E83	(α2) N212		
Interface 2			
AIMP3	**EPRS**	**Arc1p**	**ERS**
(Loop α4-α5) V106	(Loop α4-α5) T110	(Loop α4-α5) T57	(Loop α4-α5) T127
(Loop α4-α5) Y107	(Loop α4-α5) Y111	(Loop α4-α5) L63	(α7) L170
(α7) L140	(Loop α4-α5) L108	(α7) N109, L110	(Loop α4-α5) L133
(α7) R144	(α7) R149	(α7) H99	(α7) N163
(α7) H148	(α7) F153	(α7) R102	(α7) R166

3. Probing domain-domain interfaces

Table 5 shows some residues participating in contacts between GST domains in the MARS complexes from human and yeast. In order to identify similar residues in the proteins from *Plasmodium*, we performed pairwise structural alignments using the crystal structures of *Pb*ERS-N and *Pv*tRip-N and the prediction models (Raptor X) of *Pb*QRS-N and *Pb*MRS-N. In particular, we looked for exposed polar and small residues in helices α2 and α3 (interface 1) and for arginines or similar residues protruding from the helix α7 (interface 2). We mutated the corresponding residues that were conserved among *Plasmodium* species (summarized in Table 6) and the mutants were tested in pull-down assays for their capacity to bind the other GST domains (Figures 45 and 46).

In the crystal structure of *Pv*tRip-N, homodimerization is mainly mediated by hydrophobic contacts between 2 strictly conserved phenylalanine residues, F58 in α2 and F90 in α3 (Gupta et al., 2020). Several residues potentially involved in interactions through interface 1 have been identified in *Pb*ERS-N. They correspond to N93, D95 and L97 in helix α2 and V120, A124 and F132 in helix α3. Except *Pb*ERS-N D95A, that was not stable enough to perform pull-down experiments, all other mutants were assayed as baits to test their capacity to reconstitute the binary complex *Pb*tRip:*Pb*ERS. None of them had an effect on the heterodimerization (Figure 45A), indicating that the interaction between *Pb*tRip and *Pb*ERS-N does not involve interface 1. Furthermore, mutants *Pb*tRip F58A and F90A were less soluble than the wild-type and the analysis of the mutants by SEC showed that they were aggregated (data not shown), confirming that interface 1 is involved in the homodimerization of *Pb*tRip, as it is the case in the *Pv*tRip-N crystal structure (Figure 43A).

The conserved R found in helix α7 of both *Pb*tRip (R154) and *Pb*ERS-N (R198) were mutated in order to test the role of interface 2 in *Pb*tRip:*Pb*ERS-N interaction. The replacement of R residues led to the dissociation of the heterodimer (Figure 45A), showing that the binary complex forms *via* helix α7.

Interestingly, all these mutations had different consequences on the formation of the two ternary complexes. The disruption of interface 2 between *Pb*tRip and *Pb*ERS-N (mutant *Pb*ERS-N R198A) did not prevent the binding of *Pb*QRS-N, but destabilized the binding of *Pb*MRS-N. Similarly, mutations in *Pb*tRip interface 1 only disrupted the binding of *Pb*MRS-N. Although previous results indicated that *Pb*QRS-N and *Pb*MRS-N compete and therefore bind the same area on *Pb*tRip:*Pb*ERS, the present observations show that binding of *Pb*MRS-N and *Pb*QRS-N do not require exactly the same residues.

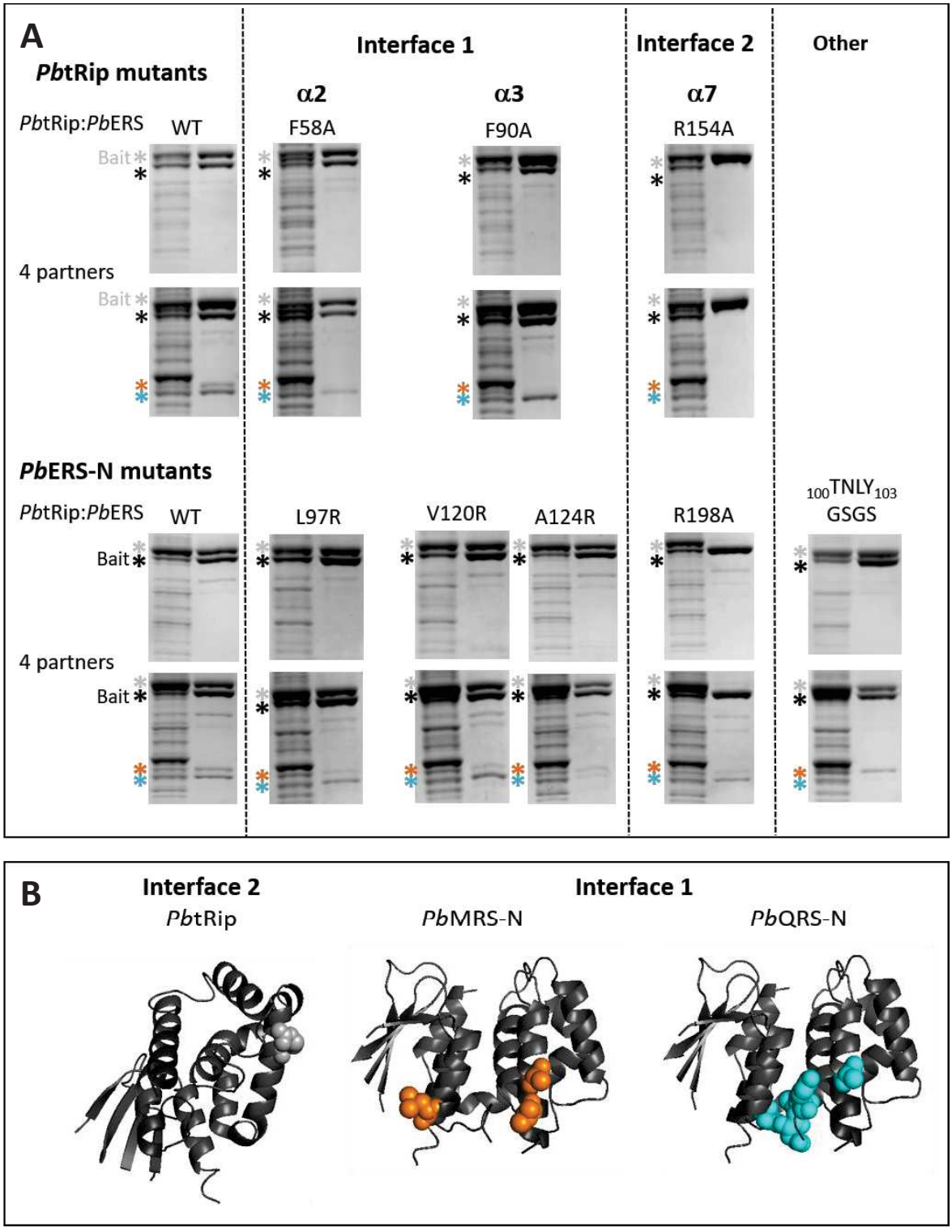


Figure 45. Identification of interaction interfaces in *Pb*ERS-N and *Pb*tRip. A. Pull down experiments. The results of different pull-down experiments with either *Pb*tRip and *Pb*ERS-N or the 4 partners involved in *Plasmodium* MARS complexes are shown. Each mutant protein was used as bait. The bands corresponding to bait and prey proteins in the initial mixture are indicated with asterisks of the corresponding color (gray for *Pb*tRip, black for *Pb*ERS, cyan for *Pb*QRS and orange for *Pb*MRS). Each interaction has been tested at least 3 times. **B. Localization of *Pb*ERS-N residues directly involved in interactions with the three other domains.** Amino acids identified in this study are shown with colored spheres according to the protein partner to which they bind (*Pb*QRS-N or *Pb*MRS-N).

The association of *Pb*MRS-N to the binary complex depends on the presence of a dimeric *Pb*tRip bound to *Pb*ERS-N, whereas *Pb*QRS-N binds directly *Pb*ERS-N. This was confirmed with mutants L97R, V120R and A124R on interface 1 of *Pb*ERS-N that reduced the binding of *Pb*MRS-N (Figure 45B, orange) while only one of them (A124R) had an effect on the binding of *Pb*QRS-N (Figure 45B, cyan). The only mutation that dissociated *Pb*QRS-N from the complex, without dissociating *Pb*MRS-N, was located in the loop between α2 and α3 ($_{100}$TNLY$_{103}$ replaced by the sequence GSGS, Figure 45B, cyan). The choice of this mutation was based on the crystal structure of the catalytically active GST enzyme of *P. falciparum.* This enzyme contains a sequence "TNLF" in a loop located between helices α3 and α4, which allows the tetramerization of the molecule (Liebau et al., 2009; Perbandt et al., 2015). Thus, the similar sequence "TNLY" between helices α2 and α3 of PbERS-N was successfully investigated.

In the raptor X model of *Pb*QRS-N, helix α7 did not contain any conserved R, therefore, only α2 and α3 were tested. Residues V66 in α2 and K97, V99 and E102 in α3 were mutated. Only V66 and V99 had an effect on the binding of *Pb*QRS-N on *Pb*ERS-N, and confirmed the existence of a canonical interaction *via* interface1 between the two GST domains (Figure 46).

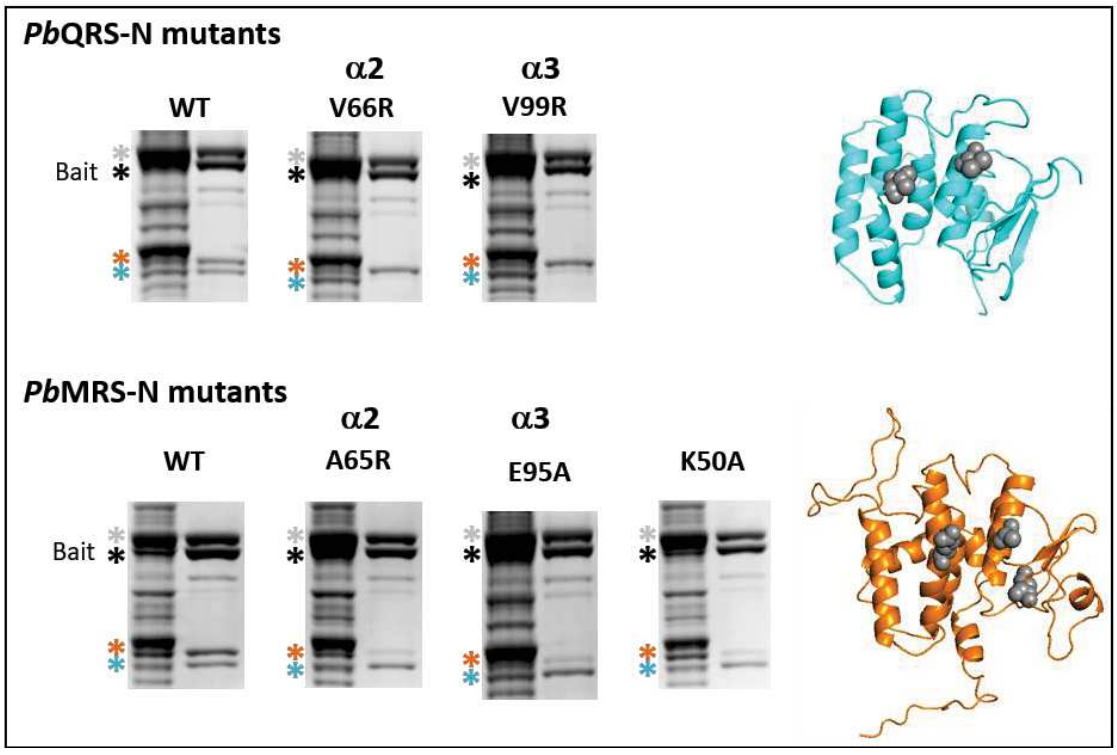


Figure 46. Identification of interaction interfaces in *Pb*QRS-N and *Pb*MRS-N. The results of different pull-down experiments with the 4 partners involved in the *Plasmodium* MARS complexes are shown. *Pb*ERS-N-SUMO was used as bait in all experiments. The bands corresponding to bait and prey proteins in the initial mixture are indicated with asterisks of the corresponding color (gray for *Pb*tRip, black for *Pb*ERS, cyan for *Pb*QRS and orange for *Pb*MRS). Each interaction was tested at least 3 times. Amino acids identified in this study are shown with gray spheres on Raptor X models of *Pb*QRS-N (cyan) and *Pb*MRS-N (orange).

In the case of *Pb*MRS-N, five residues were tested: A65 in α2, E95 in α3, R191 in α7 and N48 and N50 in the loop located between β2 and β3. These two last residues were selected based on the putative interface 3 observed in the crystal of *Pb*ERS-N. As expected, helix α7 in interface 2 did not affect the formation of the M-complex. However, A65 and E95 on interface 1 disrupted the association of *Pb*MRS-N with the binary complex *Pb*tRip:*Pb*ERS-N (Figure 46). More surprisingly, the residue K50 (but not N48) located in the thioredoxin fold of *Pb*MRS-N was also involved in this interaction, suggesting that an "interface 3" might be important for the non-canonical recognition between *Pb*tRip:*Pb*ERS-N and *Pb*MRS-N.

Table 6. Mutants of *Pb*tRip, *Pb*ERS-N, *Pb*QRS-N and *Pb*MRS-N for interfaces probing. For each mutation, the interface and the position in secondary structures (α-helix, β-strand or loop) is indicated. The effect of each mutation in the interaction with the different partners is represented with a color code.

Protein	Interface	Position	Mutation	Effect in interaction with:			
				*Pb*tRip	***Pb*ERS-N**	*Pb*QRS-N	*Pb*MRS-N
*Pb*tRip	1	α2	F58A				
		α3	F90A				
	2	α7	R154A				
***Pb*ERS-N**	1	α2	D95A				
			N93A				
			L97R				
		α3	A124R				
			V120R				
			F132A				
	2	α7	R198A				
	3	α4	H148A				
	other	loop α2-α3	$_{100}$TNLY$_{103}$GSGS				
*Pb*QRS-N	1	α2	V66R				
		α3	K97A				
			E102A				
			V99R				
*Pb*MRS-N	1	α2	A65R				
		α3	E95A				
	2	α7	R191A				
	3	loop β2-β3	N48A				
			K50A				

Disrupted interaction
No effect on interaction
Destabilization of the mutated protein

4. Structural analysis of *Pb*MARS ternary complexes

4.1. Purification of *Pb*MARS ternary complexes

The use of *Pb*QRS-N-SUMO-His and *Pb*MRS-N-SUMO-His as baits allowed the purification of Q- and M-complexes respectively. If needed, the "SUMO-His" particle was cleaved with TEV protease. This was the best strategy optimized to date that has been used to prepare complexes for light scattering and SAXS experiments.

Purification of the Q-complex was more challenging than for the M-complex for two reasons: (i) the SUMO-His in this construct spontaneously cleaved during expression and that limited the yield of the purification. This problem was solved by using a shorter version: *Pb*QRS-N_{1-179}-SUMO-His (Table 9 in Methods), which showed less auto-cleavage. (ii) Unlike the M-complex, where *Pb*MRS-N binds preferentially to the binary complex *Pb*tRip:*Pb*ERS, *Pb*QRS-N interacts directly with *Pb*ERS-N (Figure 28A) and the formation of this accessory binary complex decreased the yield and purity of the sample (Figures 30A and 30B). The ratio of the different partners in the initial mixture was important to control the formation of the Q-complex. Using an excess of *Pb*tRip with equivalent amounts of *Pb*ERS-N and *Pb*QRS-N reduced the formation of this detrimental binary complex. In general, a mixture 2:1:1 of bacteria expressing the new bait *Pb*QRS-N_{1-179}-SUMO-His, *Pb*ERS-N and *Pb*tRip, respectively, increased the purification yield of the Q-complex.

Examples of optimized purifications of Q- and M-complexes are shown in Figures 47 and 48, respectively. The SUMO bait is used to capture the complex on the Ni-NTA resin, which is then extensively washed to remove contaminants, especially nucleic acids that bind *Pb*tRip and *Pb*ERS-N (Figures 47A and 48A). Eluted complexes are treated with TEV protease, which is in turn removed from the sample by running a second Ni-NTA column (Figure 47B and 48B). In the final step, the sample is injected into a SEC column that separates the complex from aggregates and other populations of lower MWs (e.g. free baits, *Pb*QRS-N:*Pb*ERS-N complex) (Figure 47C and 48C). The SEC peak is analyzed by SDS-PAGE, bands are quantified and the more concentrated fractions containing equimolar amounts of each partner are pooled and used for further experiments. This approach allowed the purification of about 2 mg of ternary complexes from 1 L of culture (500 mL of bait, 250 mL of *Pb*ERS-N and 250 mL of *Pb*tRip).

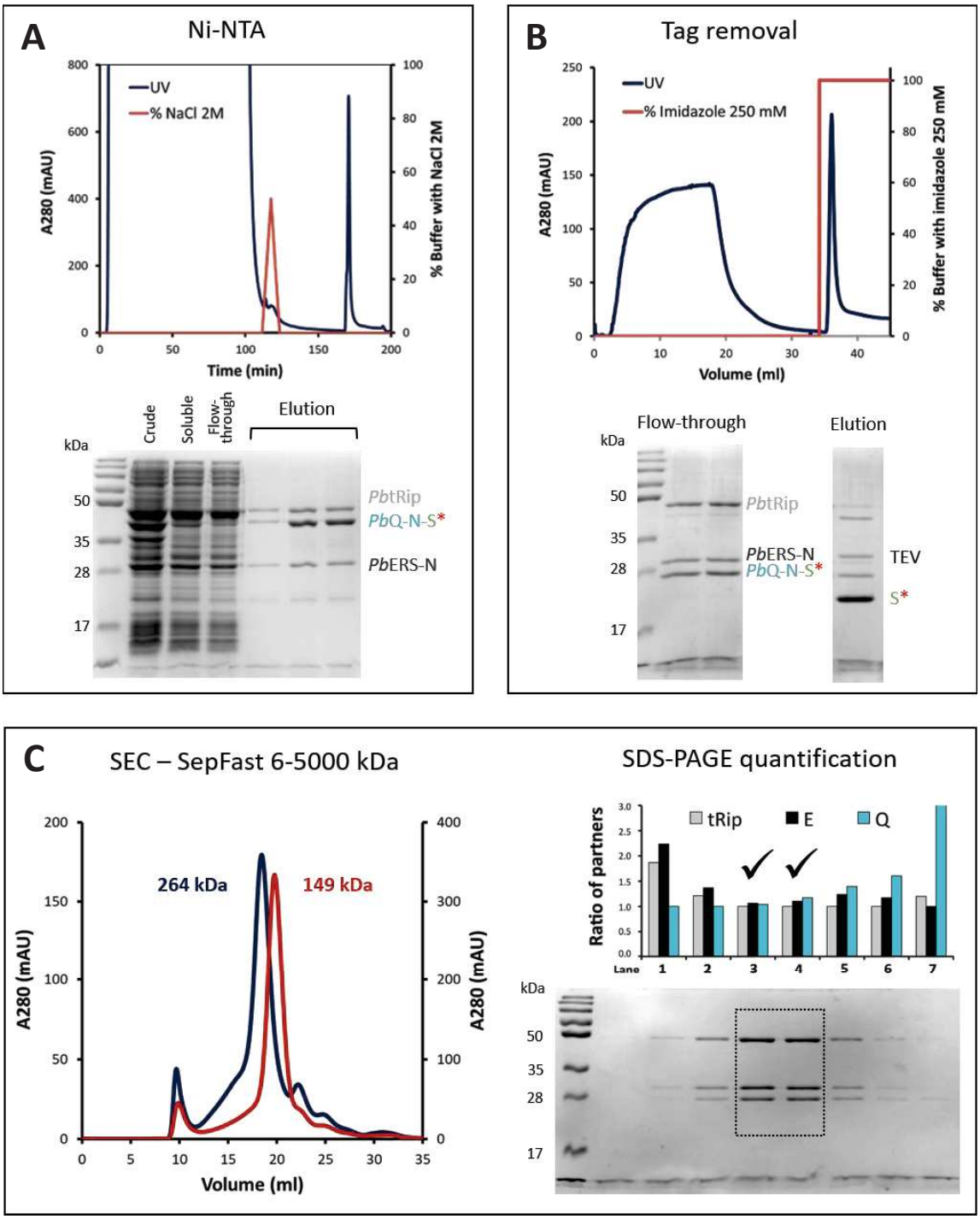


Figure 47. Purification of Q-complex. A mixture of bacteria expressing *Pb*QRS-N$_{1\text{-}179}$-SUMO-His, *Pb*ERS-N and *Pb*tRip was co-lysed and subjected to several purification steps. **A. Ni-affinity chromatography**. During washing, a gradient of NaCl (red line) allowed removal of nucleic acid contaminations. A typical chromatogram (top) and the corresponding SDS-PAGE (bottom) are shown. Fractions analyzed on gel are named and the band corresponding to each protein partner is indicated **B. Tag removal.** Digestion with TEV protease cleaves the SUMO-His from the complex and these fragments can be removed by a second Ni-NTA column. A typical chromatogram (top) and SDS-PAGE (bottom) are shown. Relevant protein bands are identified in the gel. **C. Size-exclusion chromatography (SEC).** The flow-through in **B** is concentrated and applied to a calibrated SepFast 6-5000 kDa column. The elution profiles of Q-complexes with *Pb*tRip (blue) and *Pb*tRip-N (red) are shown, with the corresponding apparent MW. Fractions were analyzed by SDS-PAGE and the relative amount of each partner quantified. Fractions containing equimolar amounts of partners are indicated with checkmarks.

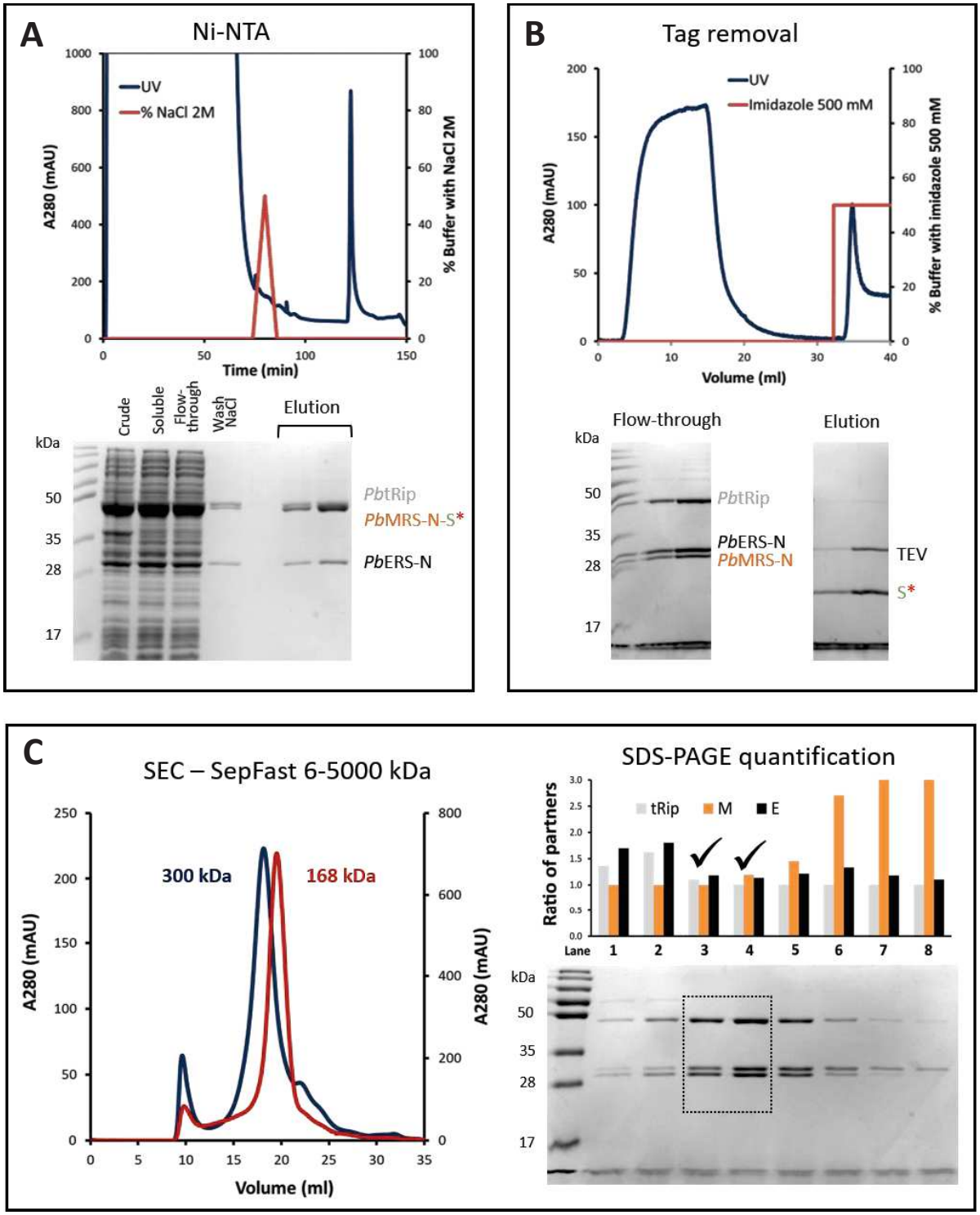


Figure 48. Purification of the M-complex. See the legend of Figure 47

Q- and M-complexes lacking the C-terminal domain of *Pb*tRip, were purified using the same protocol. Since these complexes are more compact, they were mainly prepared for crystallization experiments. Conveniently, the yield of these purifications was at least twice than those involving full-length *Pb*tRip. Judging by their SEC profile, they were also more homogenous (red chromatograms in Figures 47C and 48C). However, SDS-PAGE quantification was not always possible due to the co-migration of some partners on the gel.

Assuming a 1:1:1 stoichiometry, the calculated MW of complexes with full-length *Pb*tRip is about 100 kDa and about 80 kDa for those lacking the C-terminal domain. For both types of complexes, the apparent MW in the SEC column is compatible with dimers. In the case of complexes containing *Pb*tRip-N, estimations matched nicely the expected value for dimers (about 160 kDa). In contrast, these values were overestimated for complexes with full-length *Pb*tRip, especially for the M-complex (300 KDa *versus* 200 kDa). This is probably due to an elongated shape of the complexes, since the EMAPII-like domain of *Pb*tRip is separated from the GST domain by a flexible linker. It is well known that elongated proteins in SEC can easily elute at positions twice its real MW when the column is calibrated with globular proteins (Erickson et al., 2009). However, based only on SEC profiles, we cannot exclude that Q- and M-complexes form trimers in solution.

4.2. Size estimation by dynamic and static light scattering (DLS/SLS)

Multiple samples of M- and Q- complexes, containing either *Pb*tRip or *Pb*tRip-N, were analyzed by dynamic and static light scattering (DLS/SLS) in order to assess their homogeneity and oligomeric state in solution. Ultracentrifugation of freshly purified samples was critical to obtain reliable data, in particular for MW determinations. Indeed, the scattering intensity is proportional to the 6th power of the particle diameter and the presence of even small amounts of aggregates leads to overestimated MW values (Lorber, 2020).

Cumulant analysis of DLS data produced SOS values lower than 1, indicating that samples are composed of a single type of particles (Table 7). As a consequence, hydrodynamic diameters (d_h) showed low standard deviations, expressed as percentages of polydispersity (%PD). The diameter of M-complexes was slightly larger than that of Q-complexes and this was even more evident for complexes containing full-length *Pb*tRip, in which differences of up to 2 nm were observed at several concentrations (data not shown).

In Table 7, two different MW estimates are listed. The first one, MW-R, is derived from d_h and assumes a spherical shape for the particles and the second one, MW-S, is obtained from the SLS measurement and is thus independent of the particle shape. However, care must be taken when interpreting MW-S, since the estimation is based on the total scattering intensity of the sample, including any aggregate that might still be present despite ultracentrifugation of the sample. Moreover, the reliability of MW-S estimates depends also on the precision of the concentration measurements (Nanodrop).

Table 7. Cumulant analysis of DLS and MW estimates from different samples of Q- and M- complexes. Complexes with and without the C-terminal domain of tRip were measured at 5 different concentrations, ranging from 1 to 6 $mg.mL^{-1}$. Although some concentration effects were observed, no dramatic changes were obtained on the particle's diameter and MW estimates were generally consistent with dimers of ternary complexes. Here, the results of samples at similar concentrations are shown.

Complex	mg/mL	MW (kDa)	d_h (nm)	PD (%)	SOS	MW-R (kDa)	MW-S (kDa)	Shape factor
Q-complex (*Pb*tRip)	3.5	96.8	14.3	6.2	0.275	336.7	194	1.74
Q-complex (*Pb*tRip-N)	3.3	74.5	11.4	6.4	0.193	197	163	1.21
M-complex (*Pb*tRip)	3.4	101.5	15.7	12.4	0.200	417.3	274	1.52
M-complex (*Pb*tRip-N)	3.6	79.2	12.3	5.4	0.567	238.1	171	1.39

d_h: hydrodynamic diameter; PD: Polydispersity

The ratio MW-R/MW-S (shape factor) gives an estimation of how much the particle deviates from a sphere and values ranging between 2 and 3 are common for elongates proteins (Lorber, 2020). For all complexes, MW-R were larger than MW-S and therefore shape factors were higher than 1. Logically, the shape factors were higher for M- and Q- complexes containing the full length *Pb*tRip compared to the complexes containing *Pb*tRip-N, confirming that the presence of the C-terminal domain of *Pb*tRip is responsible of their elongated shapes. Although the MW-R estimates for the M-and Q- complexes containing *Pb*tRip were overestimated, all MW-S estimates were rather consistent with dimers of ternary complexes.

On the other hand, regularization analysis of DLS data yielded higher %PD. Especially, complexes containing full length *Pb*tRip showed larger size distributions (d_h= 15.4 and 17.7 nm for Q- and M-complexes, respectively) with a polydispersity up to 30% (Figure 49A and 49B, left side). The heterogeneity of these samples could be the consequence of the presence of the flexible C-terminal domain of *Pb*tRip but it could also be the consequence of a mixture of oligomeric states, whose dimensions are not sufficiently different to be resolved by our apparatus. In contrast, regularization analysis of complexes containing *Pb*tRip-N showed narrower size distributions, around 12 nm, and their %PD were low enough (~15%) to consider these complexes as monodispersed, and thus suitable for crystallization experiments (Figure 49A and 49B, right side)

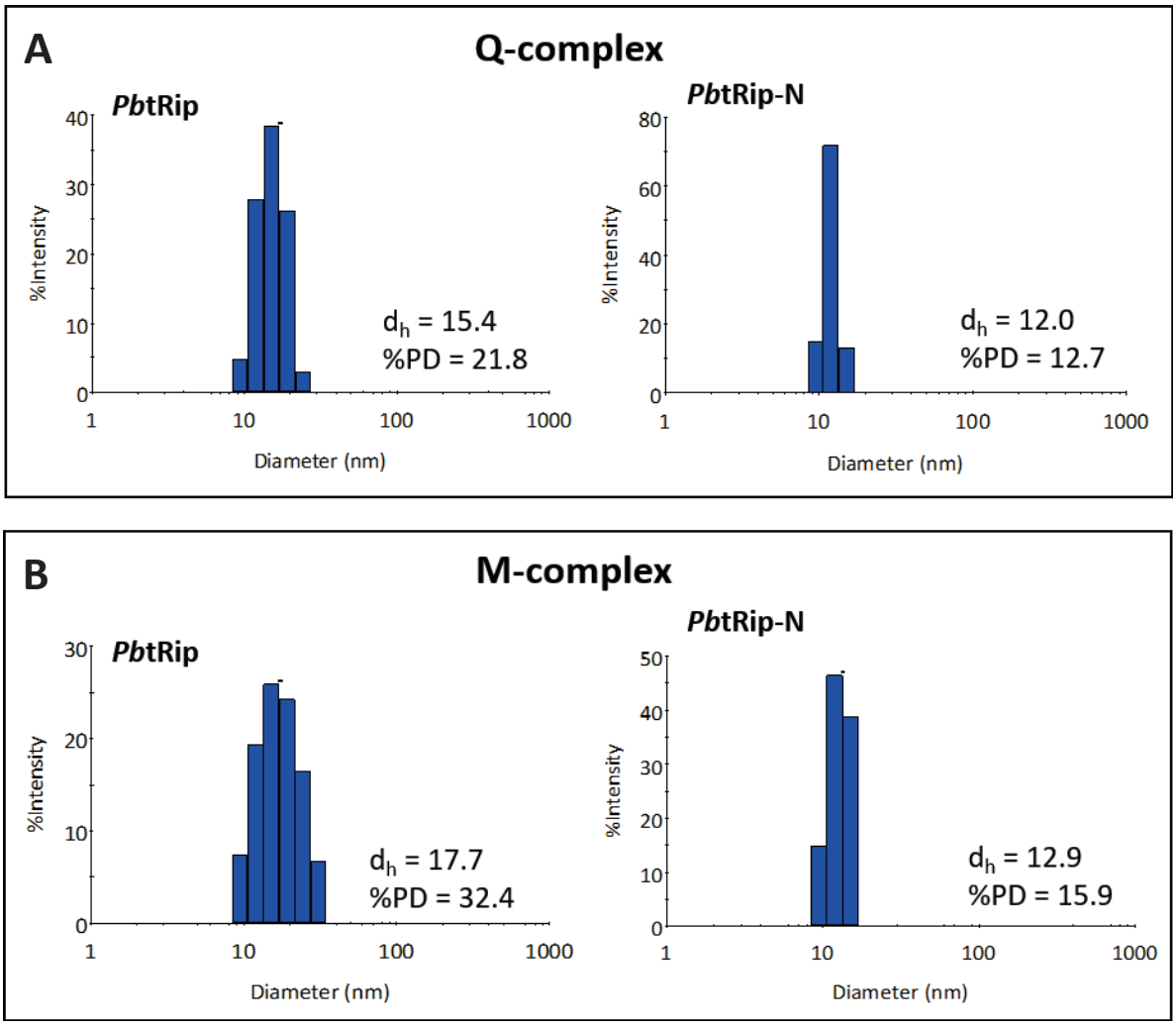


Figure 49. Particle size distribution of ternary complexes. The regularization graphs show the percentage of total scattering intensity as a function of the diameter of particles. **A. Analysis of Q-complexes.** Complexes containing *Pb*tRip (left) or *Pb*tRip-N (right). **B. Analysis of M-complexes.** Complexes with *Pb*tRip (left) or *Pb*tRip-N (right). In each case, the hydrodynamic diameter (d_h) and the percentage of polydispersity (%PD) of the distribution are indicated. Here, the results of samples at similar concentrations are shown (see Table 7).

4.3. Size estimation by Size-Exclusion Chromatography coupled to Small-Angle X-ray scattering (SEC-SAXS)

Different types of ternary complexes were analyzed by SEC-SAXS in order to assess their homogeneity in solution and confirm MW estimates obtained from SEC and DLS/SLS. Details about data collection for the different samples are summarized in Table 16 in Methods. In these experiments, complexes were injected into an analytical SEC column and multiple SAXS frames were collected as the sample was eluted. Each frame contains information about the shape and size of particles and the evolution of these parameters across the SEC peak and is correlated with the sample homogeneity. For instance, in a monodisperse sample, size-related parameters such as the radius of gyration (Rg) and the molecular weight (MW) have a constant value across the peak.

Several types of MW estimates can be obtained from SAXS data. A MW can be derived from the volume of hydrated particle in solution, also known as Porod volume (Vp), but it can be inaccurate if the molecule is not globular (Piiadov et al. 2019). Another type of estimate is obtained from the volume of correlation (Vc) and is independent of the particle shape (Rambo & Tainer, 2013). Other methods are available, but we mainly looked at these two to analyze our data.

Neither Rg (not shown) nor MW estimates were constant across the peak of any sample (Figure 50), but ternary complexes containing *Pb*tRip-N exhibited a better behavior than those containing full-length *Pb*tRip. Indeed, the MW Vc of complexes with *Pb*tRip-N was quite stable in the most concentrated portion of the peak and exhibited values around 150 kDa, which are consistent with dimers of ternary complexes. On the other hand, the MW Vc of complexes with *Pb*tRip was much more variable, even at the center of the peak. Despite this, estimates remained close to the value expected for dimers of ternary complexes. Within the limits of the peak, the MW Vc of Q- and M-complexes containing *Pb*tRip were less than ~210 kDa and ~240 kDa, respectively.

Despite variations in the sample across the peak, data processing managed to find portions of stable Rg and meaningful SAXS curves could be obtained, even for complexes with full-length *Pb*tRip. Both Guinier and P(r) analysis produced similar Rg values and parameters such as MW and Dmax were consistent among the different complexes analyzed (Table 8). Normalized Kratky plots indicated that proteins were folded (data not shown) and the ambiguity of 3D reconstructions estimated by AMBIMETER correlated with the presence of the flexible C-terminal domain of *Pb*tRip.

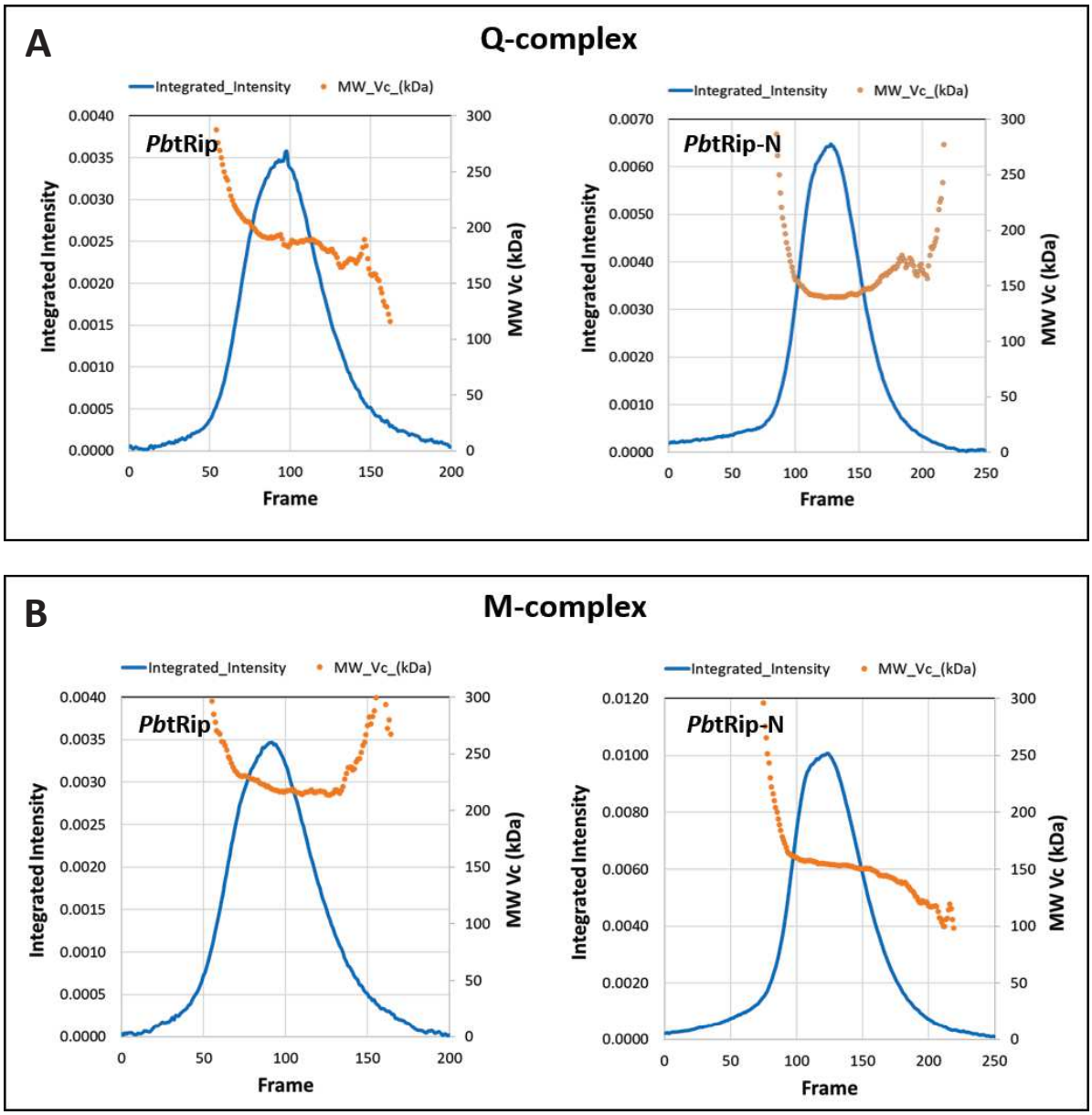


Figure 50. Estimation of molecular weights of ternary complexes by SEC-SAXS. In each graph the integrated scattering intensity (left y-axis, blue line) and the molecular weight estimate from the volume of correlation (MW Vc, right y-axis, orange dots) are plotted as function of frame number. The evolution of scattering intensity with the frame number (SAXS chromatogram) is analogous to the evolution of UV absorbance with the elution time (SEC chromatogram). **A. Q-complexes** and **B. M-complexes.** In each case, complexes with full-length *Pbt*Rip are shown on the left side and complexes with *Pbt*Rip-N on the right side.

Table 8. Structural parameters and MW estimates from SAXS analysis of Q- and M- complexes. SAXS curves were obtained by averaging contiguous frames exhibiting statistically similar Rg values at the center of the SEC peak. Curves for Q- and M- complexes with *Pb*tRip-N are plotted in Figure 51.

Analysis	Sample			
	Q-complex (*Pb*tRip-N)	**Q-complex (*Pb*tRip)**	**M-complex (*Pb*tRip-N)**	**M-complex (*Pb*tRip)**
Guinier analysis				
I0 (cm-1)	0.1753±1.78E-04	0.1051±1.61E-04	0.0839±1.05E-04	0.0894±2.49E-04
Rg (Å)	46.8111 ± 0.0799	54.9788±0.1339	46.9846±0.1122	55.8426±0.237
q range (Å $^{-1}$)	0.01095 – 0.02554	0.01095-0.02189	0.0073-0.02554	0.01231-0.02144
qRg max	1.1956	1.2036	1.2000	1.1971
R^2	0.9996	0.999	0.998	0.9992
Molecular weight				
MW ProtParam (kDa)	74.5	96.8	79.3	101.4
MW Vc (kDa)	143.2	188.5	148.7	215.7
MW Vp (kDa)	173.7	228.3	184.1	244.8
MW Bayesian (kDa)	157.1	208.0	169.6	208.0
MW shape & size	148.7	198.5	156.9	212.8
Corrected Vp (Å ^{3})	2.09E+05	2.75E+05	2.22E+05	2.95E+05
q range (Å $^{-1}$)	0.01095-0.2998	0.01095-0.2998	0.0073-0.2998	0.01231-0.2998
P(r) analysis				
I0 (cm^{-1})	0.1755±1.18E-04	0.1059±1.38E-04	0.0840±1.02E-04	0.0899±1.41E-04
Rg (Å $^{-1}$)	47.7800±0.0339	56.5700±0.1062	47.9100±0.0891	57.4400±0.1002
Dmax (Å)	147	193	166	185
q range (Å $^{-1}$)	0.0109-0.2998	0.01-0.2998	0.0073-0.2998	0.0123-0.2998
Chi^2	1.3406	1.9099	1.1181	1.2373
Total estimate from GNOM	0.7683	0.7603	0.7905	0.8107
GNOM says	a GOOD solution	a GOOD solution	a GOOD solution	a GOOD solution
AMBIMETER				
Compatible shapes	249	849	245	704
Ambiguity score	2.396	2.929	2.389	2.848
AMBIMETER says	3D reconstruction might be ambiguous	3D reconstruction is highly ambiguous	3D reconstruction might be ambiguous	3D reconstruction is highly ambiguous

I0: Scattering intensity at the origin
Rg: Radius of gyration
Vc: Volume of correlation
Vp: Porod volume
Dmax: maximum dimension of particle

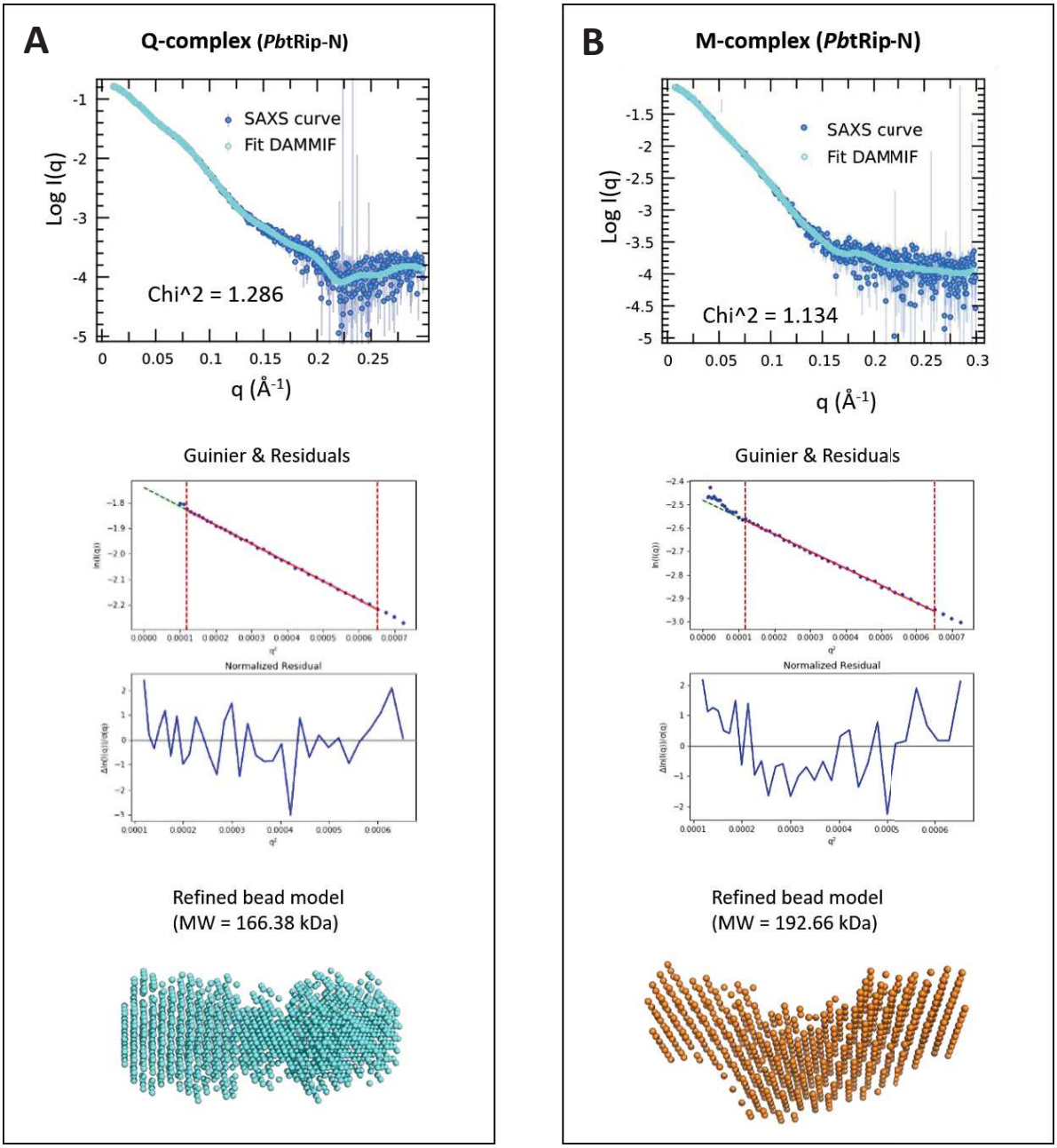


Figure 51. SAXS curves and *ab initio* bead models for ternary (A) Q-complex and (B) M-complex with *Pbt*Rip-N. SAXS curves (blue) were obtained by averaging a zone of frames with constant Rg at the center of the elution peak. The Guinier fit and residuals are shown under each curve. The *P(r)* calculated by GNOM was used to build multiples *ab initio* bead models (DAMMIF), that were then averaged (DAMAVER) and refined (DAMMIN). They are shown at the bottom of each panel. The calculated scattering profile of these 3D reconstructions (cyan) agreed well with the experimental profiles. The molecular weight estimate from the refined bead model is also consistent with dimers of ternary complexes.

Figure 51 shows the SAXS curves and *ab initio* bead models obtained for Q- and M- complexes containing *Pb*tRip-N. Although there is always ambiguity in 3D modelling from SAXS data, the multiple models build from these curves resulted in similar elongated and rather symmetrical particles for each complex. This is consistent with the shape factors observed in DLS/SLS and supports the dimeric nature of these complexes. Interestingly, the SAXS curves and bead models were different between the two complexes, even if proteins bind the same partners and have similar sizes and predicted structures. This suggests different binding modes of *Pb*QRS-N and *Pb*MRS-N when interacting with the binary complex *Pb*tRip-N:*Pb*ERS-N.

4.4. Modeling of one *Pb*MARS complex

Our results, taken together, provided enough information to propose a model at least for Q-complex. We have shown that (i) assembly is exclusively mediated by GST domains, (ii) *Pb*tRip and *Pb*ERS-N interact by interfaces 2, (iii) *Pb*ERS-N binds either *Pb*QRS-N or *Pb*MRS-N using mutually exclusive sites around interface 1, and (iv) homodimerization of *Pb*tRip-N allows the formation of a bisymmetric complex.

Since interaction interfaces in Q-complex were unambiguously determined, I used the crystal structures of *Pb*ERS-N and of *Pv*tRip and the predicted structure (Raptor X) of *Pb*QRS-N to build a rational 3D model (Figure 52). The interaction between *Pv*tRip and *Pb*ERS-N was reconstructed by the superposition of 2 homodimers interacting by their interfaces 2. The interaction between *Pb*QRS-N and *Pb*ERS-N was reconstituted by superposing *Pb*QRS-N to one of the *Pb*ERS-N monomers in the canonical dimeric GST conformation (interface 1). Contacts observed between helices α2 and α3 are consistent with a polar interface 1 and some of them were supported by mutagenesis experiments. For instance, V66 in helix α2 of *Pb*QRS-N aligned well with A124 in helix α3 of *Pb*ERS-N and the polar residues observed at the interface included Y103 from the loop "TNLY" of *Pb*ERS-N and 2 strictly conserved glutamates (E102 and E103) in helix α3 of *Pb*QRS-N (Figure 52A). Validation and optimization of this model using SAXS data is currently underway.

In the case of the M-complex, further experiments are necessary to determine the precise binding site of *Pb*MRS-N on *Pb*ERS-N and to clarify the role of *Pb*tRip in this interaction. In the absence of crystal structures, these questions can be addressed by combining the information from mutagenesis experiments and SAXS data. For instance, the optimal position of *Pb*MRS-N on the binary complex *Pb*tRip-N*:Pb*ERS-N can be determined by performing rigid body modeling around the possible binding sites. We anticipate that *Pb*MRS-N binds a site partially overlapping the binding interface with *Pb*QRS-N but closer to the binding interface with *Pb*tRip.

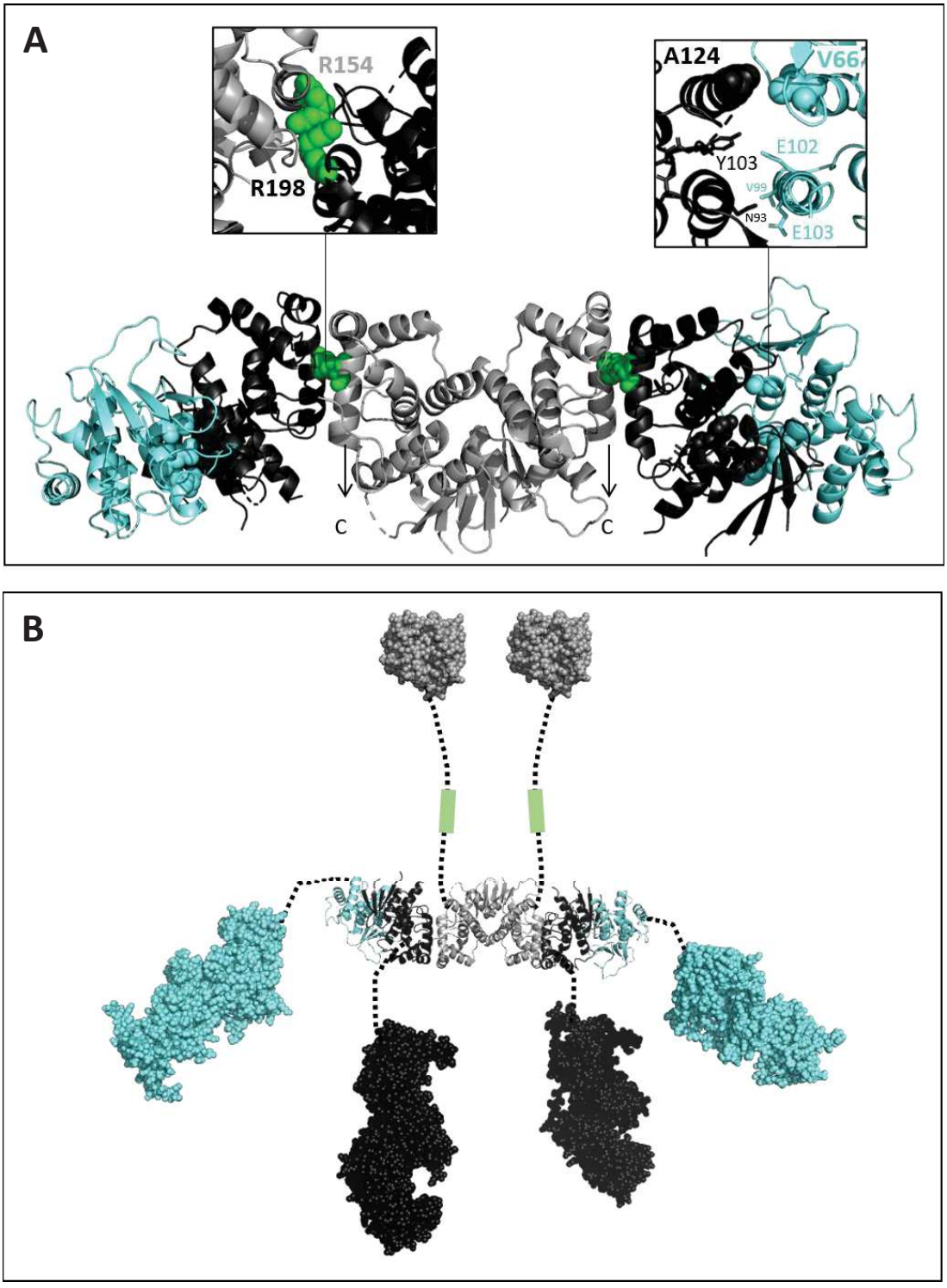


Figure 52. A model of the *P. berghei* Q-complex. A. Organization of the GST domains. The central *Pv*tRip dimer (gray) binds two monomers of *Pb*ERS-N (black), which in turn interact with two molecules of *Pb*QRS-N (cyan). Top views of the interaction interfaces are shown in panels. In each case, the potentially interacting residues are labeled. The arginines in helices α7 and the small amino acids in helices α2 and α3 are highlighted with spheres. **B. Model of the complete Q-complex**. This model includes the Raptor X models of *Pb*ERS aaRS core (black) and *Pb*QRS (cyan) and the EMAPII domain of *Pv*Rip (grey, PDB 6IPA) (Gupta et al. 2020). Interacting GST domains are shown in cartoon and the aaRSs cores are in spheres. Linkers are represented with dashed lines and the position of a putative transmembrane helix is framed in light green.

The original organization in the interaction interfaces of GST domains allow the positioning of the rest of the complex (aaRSs cores and EMAPII-like domain of tRip) without any obvious steric hindrance (Figure 52B). The presence of linkers between GST domains and the cores of aaRSs gives a lot of flexibility for their positioning in the complex. Interestingly, the atypical symmetry observed in the crystal structure of the *Pv*tRip-N dimer orients the C-terminal extremities of each monomer in the same direction, suggesting that fused EMAPII-like domains would also be oriented in the same direction. This would be compatible with the extracellular localization of EMAPII-like domains and intracellular localization of the rest of the MARS complex.

DISCUSSION

1. Specific features of the aaRSs belonging to the *Pb*MARS complexes

Our investigation to identify the tRip interactome in *Plasmodium berghei* is a valuable approach to understand the function of this unique membrane protein. In a previous work we performed co-immunoprecipitation of the endogenous *Pb*tRip with a specific antibody followed by a mass spectrometry analysis. As expected, the identified interactome contains aminoacyl-tRNA synthetases. They are involved in the specific tRNA aminoacylation with glutamate, methionine and glutamine. The composition of this complex is similar to that of multi-synthetase complexes identified in other protozoa, such as *S. cerevisiae* (Arc1p:ERS:MRS) (Simos et al. 1996) or more recently *Toxoplasma gondii* (Tg-p43:ERS:MRS:QRS:YRS) (van Rooyen et al., 2014). The 3 aaRSs identified in the *Pb*tRip complex are characterized by additional domains and insertions specific to *Plasmodium*. Sequence analysis confirmed our choice of working with *P. berghei* aaRSs since they contain shorter insertions, an indication that favors the expression and the purification of corresponding recombinant proteins. However, cloning and expression of these enzymes did not result in sufficient soluble and full-length aaRSs. We therefore chose to work with individual domains, N- and C- terminal domains in order to characterize their role in the complex association and binding of tRNA.

Interestingly the four proteins involved in the complex formation contain GST domains. These domains are restricted to eukaryotes. When searching GST-like domains in the *P. berghei* genome (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), only 6 proteins were retrieved, including *Pb*ERS, *Pb*MRS and *Pb*tRip. The other three proteins are two catalytically active glutathione-S-transferases (one of them is fused to the elongation factor γ) and the elongation factor 1 β that contains a very short (54 amino acid long) GST-C-terminal-like domain. The N-terminal GST domain of *Pb*QRS was not identified as a GST C-terminal-like domain by this website. This is certainly the consequence of poor sequence conservation in this N-terminal domain and the presence of a LCR in the middle of the structural domain. However, further investigations, especially its modeling with Raptor X, allowed its identification. The three aaRSs: *Pb*ERS, *Pb*MRS and *Pb*QRS are the only aaRS*s* containing a GST domains, and all were found associated with *Pb*tRip.

Compared to the other homologous aaRSs, *Plasmodium* ERS, MRS and QRS are characterized by some unique features (Figures 53 and 54). In almost all eukaryotes, ERS acquired a GST domain appended to their N-terminus, the only exception being kinetoplastida (i.e. *Trypanosoma brucei*) (Gowri et al., 2012). In Metazoa, the ERS GST domain is fused to PRS via several repeats of WHEP domains in order to form a large bifunctional enzyme. In *Plasmodium*, ERS does contains a GST domain appended to its N-terminal extremity, but the enzyme is not fused to PRS (Figure 53).

On the other hand, most eukaryotic QRSs acquired a YqeY domain appended to their N terminus (Grant et al., 2012). This additional domain is also present in bacterial tRNA-dependent amidotransferases (AdTs) (Oshikane et al., 2006). The YqeY domain is not present in the *Plasmodium* QRS. Rather, the enzyme displays a GST domain. Additionally, *Plasmodium* QRS contains a small tRNA binding motif appended to its C-terminal extremity.

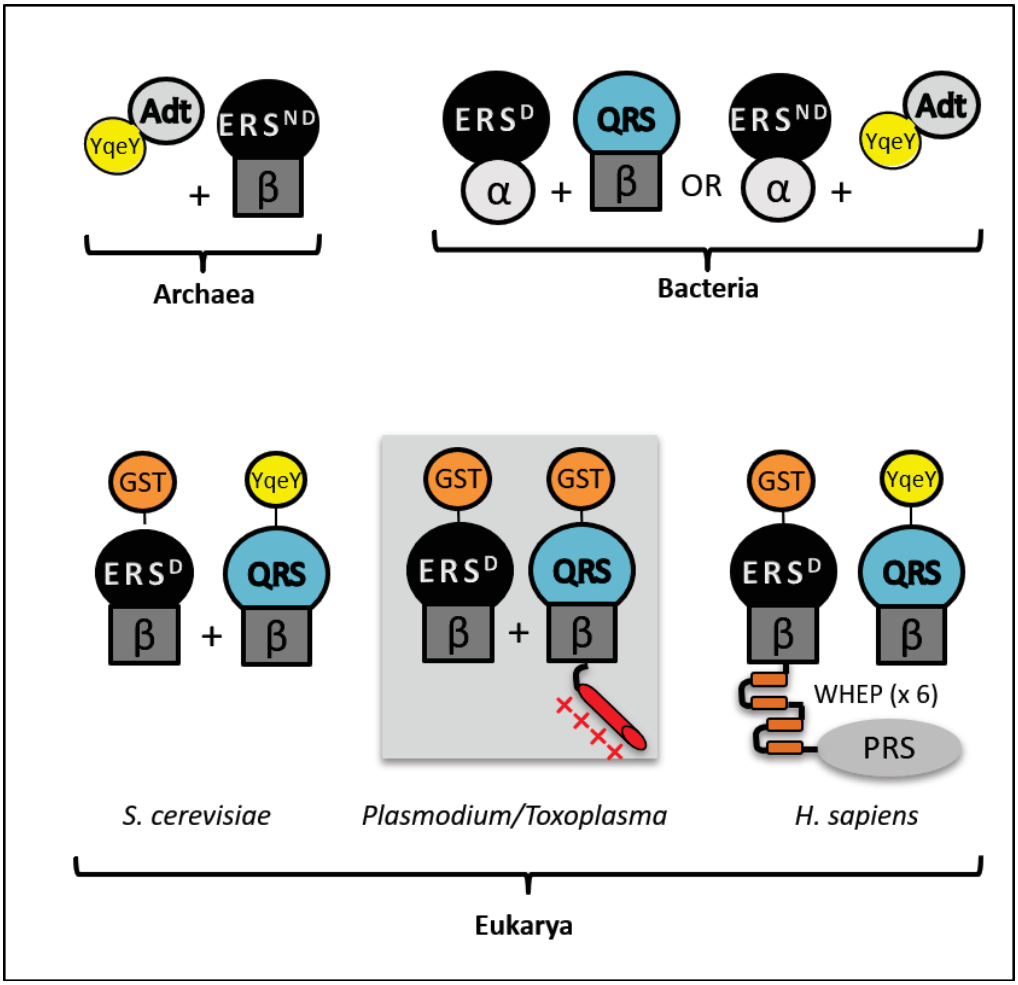


Figure 53. Evolutionary history of ERS and QRS pathways. Since QRS was not present in the last universal common ancestor, glutamine was incorporated into proteins using an indirect tRNA aminoacylation pathway where a non-discriminant ERS (ERS^{ND}) aminoacylated $tRNA^{Q}$ with E and then a tRNA-dependent amidotransferase (AdT) converted the misacylated E-$tRNA^{Q}$ into Q-$tRNA^{Q}$. These ERS^{ND} enzymes have separated in two types that are distinguished by the structure of their ABD, which contains either two β-barrels or α-helices. The α-ERSs are present in contemporary bacteria and β-ERSs passed to archaea and eukaryotes. While β-ERSs in archaea remained non-discriminating enzymes, in eukaryotes a gene duplication event allowed differentiation of β-ERS^{ND} into a discriminating ERS (ERS^{D}) and a QRS. These two enzymes then evolved separately and acquired additional modules such as GST domains and RNA binding domains. The identity of each aaRS is indicated by the one letter symbol of its amino acid substrate, GST domains are green, the YqeY domain is yellow, the positively charged helix at the C-terminus of the *Pb*QRS is shown in red and the repeat of six WHEP domains in human ERS are labeled. Figure adapted from Hadd & Perona (2014).

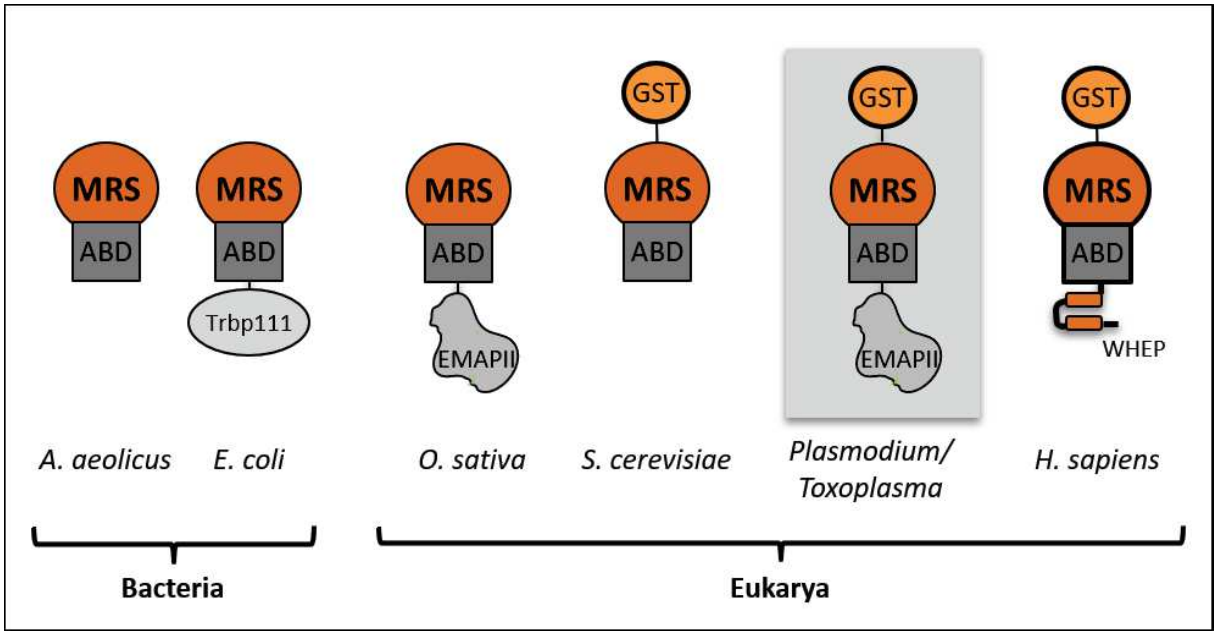


Figure 54. Structural diversity of MRSs. Kaminska et al. (1999) defined 5 groups of MRS. A minimal core enzyme, consisting of the catalytic domain (orange) and the anticodon-binding domain (ABD), is found in organisms such as *A. aeolicus* and organellar MRSs. A large group of bacteria (e.g. *E. coli*) and some archaea (not shown) possess MRSs with a Trbp111-like domain appended to the C-terminus, which allows dimerization and improved tRNA affinity (Crepin et al., 2002). Similarly, in plants (e.g. *O. sativa*) and some eukaryotes (e.g. *C. elegans*), a monomeric EMAPII-like domain is appended to the C-terminus of MRS and provides additional non-specific tRNA binding properties (Kaminska et al., 1999; Havrylenko et al., 2010). Cytosolic yeast MRS possesses a GST domain appended to its N-terminus, which mediates the incorporation of the enzyme into the MARS complex (Simader, Hothorn, Köhler et al.., 2006). Human MRS possess appended domains to both extremities: an N-terminal GST domain that mediates the incorporation of the enzyme into the MARS complex and a C-terminal WHEP domain that provides tRNA-binding properties, which modulates the activity of the enzyme (Kaminska et al. 2001). The GSTdomains are green and the EMAPII-like tRNA binding domains in light grey.

Similarly, the *Plasmodium* MRS is also peculiar (Figure 54). Indeed, in Bacteria and Eukarya, MRSs display either a GST N-terminal domain, or a C-terminal tRNA binding domain, the only exception so far being the human MRS that contains both a GST domain and a WHEP domain. *Plasmodium* is the second example of such a MRS with two additional domains: an N-terminal GST domain and a C-terminal EMAPII-like domain.

In the present study, I was able to show that the C-terminal domains identified in *Pb*QRS and in *Pb*MRS are both non-specific RNA binding domains. The specific modular organization of both *Pb*MRS and *Pb*QRS is conserved in MRS and QRS of the other Apicomplexan parasite *T. gondii* (van Rooyen et al., 2014). These observations lead to the following question: why so many tRNA binding domains in the *Pb*MARS complexes?

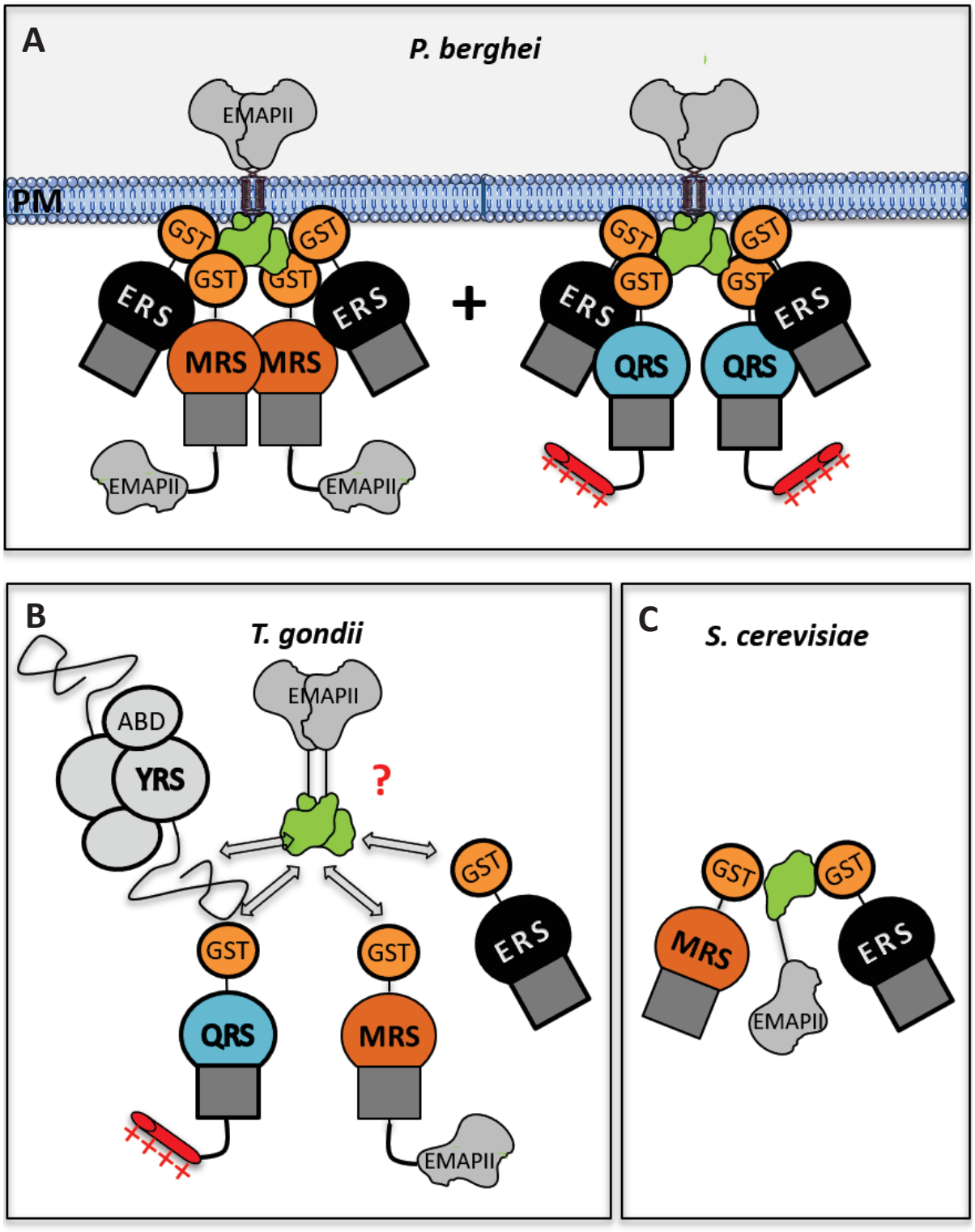


Figure 55: Composition and architecture of protozoan MARS complexes. Schematic views of the MARS complexes described in **(A)** *P. berghei* (this study), **(B)** *Toxoplasma gondii* (van Rooyen et al., 2014) and **(C)** *S. cerevisiae* (Simader, Hothorn, Köhler et al., 2006). The color codes are as described in the legend of Figures 53 and 54. In the complex of *T. gondii*, the YRS is represented as a dimer with an unfolded N-terminal domain. The red question mark indicates that, despite high sequence homology, the Tg-p43 protein was shown to be cytosolic, whereas *Pb*tRip is localized at the surface of *Plasmodium*.

2. Why so many tRNA binding domains in the *Pb*MARS complexes?

Our results show that *P. berghei* express two individual ternary complexes containing either *Pb*tRip, *Pb*ERS and *Pb*MRS (like in *S. cerevisiae*) or *Pb*tRip, *Pb*ERS and *Pb*QRS. These proteins associate through their GST domains and form a dimer *in vitro* (Figure 55A). The two closest examples of protozoa MARS complexes correspond to *S. cerevisiae* and *Toxoplasma gondii*. As in *P. berghei*, the *T. gondii* MARS complex contains Tg-p43 (*Pb*tRip homologue), *Tg*ERS, *Tg*MRS, *Tg*QRS but also a 4th aaRS, the dimeric *Tg*YRS (Figure 55B). This is the only report of a YRS enzyme being in a MARS complex and its inclusion was unexpected because its N-terminal extension is predicted to be highly disordered. The tagged *Tg*MARS complex was localized in the cytosol and its purification led to significantly heterogeneous samples in size and composition (van Rooyen et al., 2014). The poor homogeneity the *Tg*MARS complex was confirmed by electron microscopy, which failed to identify a unique species of particle. Unexpectedly, the different proteins in the complex were clearly separated from each other and not arranged as dense globular structures, as expected for large macromolecular complexes. On the contrary, the *S. cerevisiae* MARS complex is well defined. It contains only two aaRSs, *Sc*MRS and *Sc*ERS, which are organized around Arc1p (*Pb*tRip homologue) (Figure 55C). Interactions between the three GST domains that are responsible for the complex association have been identified par crystallography (Simader, Hothorn, Köhler et al., 2006).

When compared to the *Sc*MARS complex, both the *Tg*MARS and the *Pb*MARS complexes display more tRNA binding domains. Among the three *Sc*MARS proteins, there is only one EMAPII-like domain (in Arc1p) that was shown to strongly increase the affinity of *Sc*MRS and *Sc*ERS for their cognate tRNAs (Simos et al., 1996), whereas each comparable monomeric *Pb*MARS complex contains two tRNA binding domains for three proteins, either two equivalent EMAPII-like domains in the M-complex or one EMAPII-like and one positively charged helix in the Q-complex. Interestingly, this ratio is also lower in the human MARS complex where it reaches nine non-specific tRNA binding domains (two EMAPII-like domains from the two AIMP1 molecules, four positively charges helices in dimeric DRS and KRS, two WHEP repeats in the dimeric EPRS and one WHEP in MRS) among the 18 protein partners.

The enrichment of the *Plasmodium* MARS complex in tRNA binding domains might be the consequence of the particular sub-cellular localization of M- and Q- complexes in the parasite: the *Pb*tRip EMAPII-like domain is exposed outside and the second tRNA binding domain (EMAPII-like or positively charged helix) would be present inside the parasite and thus available to increase the affinity of the aaRSs for their cognate tRNAs.

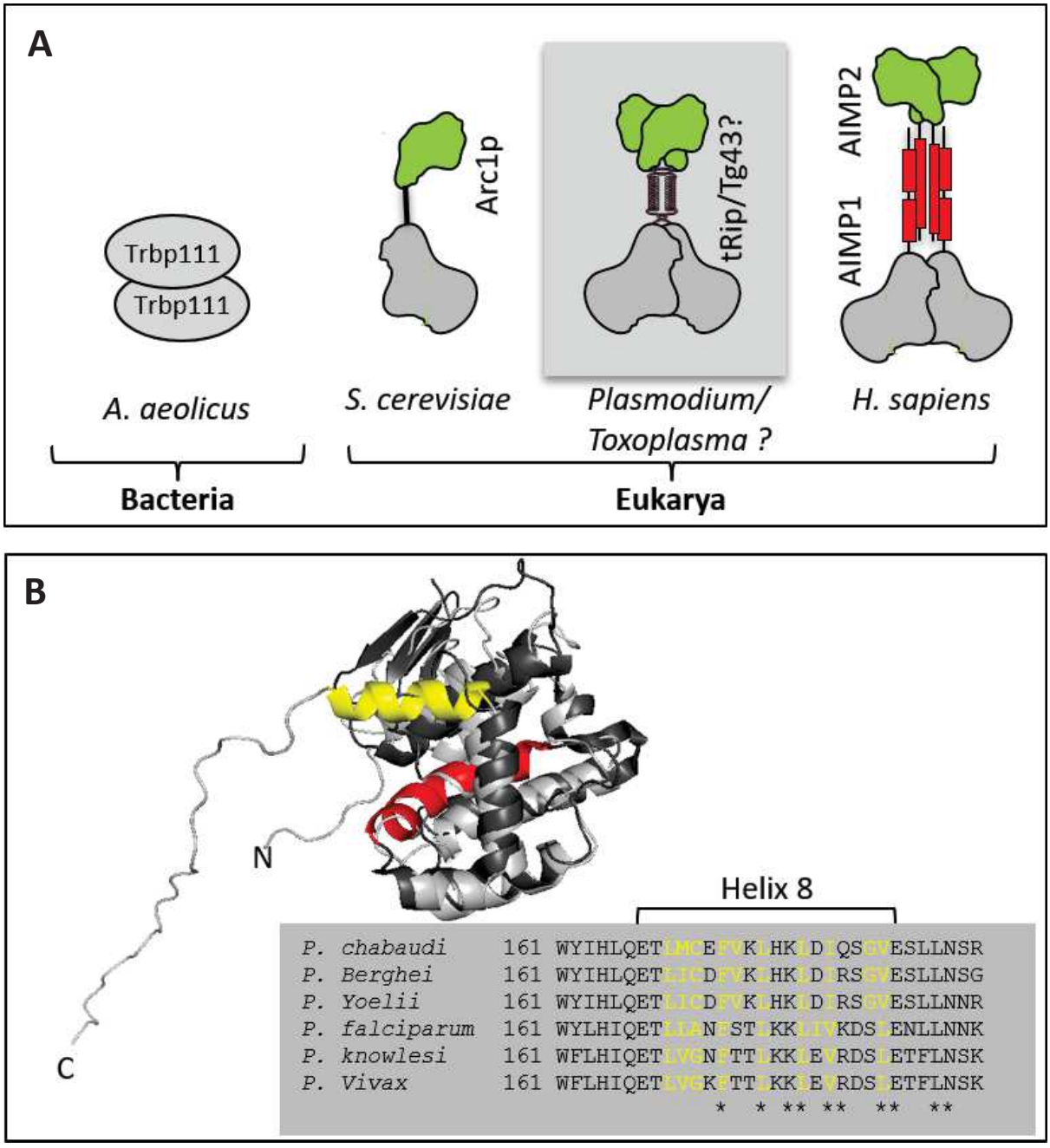


Figure 56. Looking for a trans-membrane domain in *Pb*tRip. A. Structural diversity of AIMPs containing an EMAPII-like domain. A large group of bacteria contains a free homodimeric Trbp111-like domain (Morales et al., 1999). In yeast, the cytosolic and monomeric Arc1p possesses a C-terminal GST domain and a N-terminal EMAPII-like domain. Homologous proteins in apicomplexan parasites (*Toxoplasma* and *Plasmodium*) are dimeric and the plasmodial protein has been shown to be localized at the plasma membrane. In human, the interaction between the two leucine-zippers of AIMP1 and AIMP2 allows to reconstitute a split protein with the same topology than *Plasmodium* tRip since it has the capacities to homodimerize *via* the GST domain of AIMP2. **B. Modeling of helix α8 in *Pb*tRip.** *Pb*tRip-N (residues 1-200) was modeled (light grey) and superimposed to the crystal structure of the N-terminal domain of *P. vivax* tRip-N (residues 1-174, PDB: 5ZKF) (dark grey). Only one monomer is shown. Helix α6 (red) was previously considered as a putative transmembrane helix (Bour et al., 2016). Helix α8 (yellow) stands out from the structure and contains substantial numbers of conserved hydrophobic residues to be a new transmembrane helix candidate.

3. A membrane-bound *Pb*MARS complex?

*Pb*tRip was clearly characterized as an integral membrane protein. All the tRip protein was localized at the plasma membrane of the parasite at both the sporozoite stage (the infectious stage injected by mosquitoes into vertebrate hosts) and the blood stage (Bour et al., 2016). This localization was determined not only by immunolocalization experiments (on sporozoites using a purified specific antibody raised against the *Pf*tRip-EMAPII-like domain), but also by biochemical approaches (differential solubility of membrane proteins of blood stage-parasites). Moreover, shaving experiments (Bour et al., 2016), and immunolocalization experiments performed in native conditions (unpublished data) showed that the C-terminal domain of *Pb*tRip is present outside the parasite.

To date, none other EMAPII-like protein was described with such features (Figure 56A). Moreover, it appears now, with the publication of the structure of the tRip N-terminal domain from *P. vivax*, that the transmembrane helix predicted in Bour et al., 2016 is in fact deeply buried in the GST structure (S. Gupta et al., 2020) (Figure 56B). As defined by Ganapathiraju et al., (2008): "All the transmembrane helix prediction methods make use of two fundamental characteristics: (i) the length of the transmembrane helix being at least 19 residues so that it is long enough to cross the 30 Å thick hydrophobic core of the bilayer, and (ii) the transmembrane residues being hydrophobic for reasons of thermodynamic stability in the membrane environment". Thus, the length and the high hydrophobicity of helix α6 explain the "mistake" of the predict-protein program in identifying this helix as a transmembrane helix. Moreover, our results demonstrate that *Pb*tRip and *Pb*ERS-N interact *via* their respective helix α7. This observation implies that if helix α6 is the transmembrane helix, helix α7 is outside and the complexes cannot form anymore inside the cell. *Pb*tRip contains an additional helix in its N-terminal domain, which was not included in the crystallized sequence (PDB 5ZKF). Helix α8 was modeled in the context of the GST domains using Raptor X software. The model of the monomer was generated before the publication of the crystal structure and superimpose well with it (Figure 56B). If the sequence of this helix α8 is not strictly conserved among *Plasmodium* tRip, it contains a bunch of conserved hydrophobic residues. In the hypothesis that helix α8 is the transmembrane domain, the N-terminal GST domains of *Pb*tRip and *Pb*ERS can then interact *via* their respective helix α7 inside the parasite. Moreover, the unique symmetry identified in the crystal structure of the *Pv*tRip-N dimer allows to position helices α8 of each monomer in the same direction. This is essential to orient both EMAPII-like domains also in the same direction, towards the outside of the cell, while the rest of the complex is located inside the cell.

Moreover, there are scattered elements in the literature that suggest the possibility for a GST-like domain to traverse/interact/bind cellular membranes. Among them, chloride intracellular channels (CLIC) are a unique class of ion channels, which exist as both soluble and membrane forms (reviewed in Argenzio & Moolenaar, 2016). Crystal structures of soluble CLICs show structural homology to the GST family and they can auto insert into membranes to form channels (and function as ions channels *in vitro*), even if they miss a signal sequence. Biophysical studies (FRET, Goodchild et al., 2011; Hare et al., 2016) indicate that, upon oxidation, CLIC1 forms large oligomeric complexes containing six to eight subunits, and propose a model in which the N-terminal domain inserts into the bilayer as an extended α-helix. If, the ion channel hypothesis remains speculative, CLIC proteins still have roles in diverse biological processes associated with membrane trafficking. They are often found associated with the actin cytoskeleton and to intracellular membranes, where they may participate in the formation and the maintenance of vesicular compartments.

Although we still don't know how *Pb*tRip anchors to the plasma membrane of the parasite, it is interesting to note that tRip is not the only unexpected RNA binding protein localized at the surface of the parasite.

4. *Pb*tRip is not the only RNA/tRNA binding protein present at the surface of *Plasmodium*.

PolyA Binding proteins (PABPs) are essential proteins involved in the addition/deletion of poly(A) sequences at the 3' end of transcripts to stabilize mRNAs and control their translation (reviewed in Goss & Kleiman, 2013). Unicellular eukaryotes generally encode a single cytosolic PABP, whereas metazoans encode at least two PABPs (nuclear and cytosolic). A recent study has identified two PABPs in *Plasmodium* species (Minns et al., 2018). In *P. yoelii*, PABP1 has characteristics similar to cytoplasmic PABP, while PABP2 has characteristics similar to nuclear PABP. Using the recombinant proteins *Py*PABP1 and *Py*PABP2, authors have shown that both proteins bind specifically to poly(A) sequences and that, at most stages, these proteins are expressed and localized in the cytosol and the nucleus, respectively. But, surprisingly, almost all *Py*PABP1 is localized on the surface of mosquito salivary glands sporozoites and is deposited in trails when the parasite slides on a substrate.

In all organisms, GAPDHs play a key role in glycolysis, but GAPDH is also a multifunctional protein involved in a variety of other cellular processes beyond metabolism, including membrane fusion, cytoskeletal dynamics, DNA replication and repair, etc. (reviewed in Nicholls et al., 2012). Among all these extra-glycolytic cellular functions, the one that concerns us here is the interaction of GAPDH with nucleic acids. Indeed, GAPDH has been implicated in the nuclear export of tRNAs, the stability and translation of mRNAs, and the replication and expression of several single-stranded RNA viruses. However, it is still not known how GAPDH binds to its target RNAs (reviewed in White & Garcin, 2016). Moreover, recent studies have identified surface expression of GAPDH at several stages of the *Plasmodium* life cycle (Lindner et al., 2013; Sangolgi et al., 2016), especially, it plays an important role in liver infection by sporozoites in *P. falciparum* (Cha et al., 2016).

These are not just one but three RNA-binding proteins found on the surface of *Plasmodium* sporozoites, although all of these studies may have been the artifact of labeling procedures or other experimental design... However, understanding why the parasite, particularly at the sporozoite stage, uses RNA-binding proteins on its surface may provide additional clues as to how it interacts with other parasites and/or the host. Although the precise functions of GAPDH, PABP1 and tRip on the *Plasmodium* surface are unknown, it can be argued that the RNA-binding capacity of these proteins could provide a powerful pathway for binding and importing different extracellular RNAs.

5. What is the evidence for an import of regulatory RNAs/tRNAs into the parasite?

In its complex life cycle, the *Plasmodium* parasite undergoes multiple stages in host vertebrates and mosquitoes. In order to complete its life cycle, the parasite must achieve highly controlled regulation of gene expression, but the genomes of *Plasmodium* show a relative scarcity of transcription factors compared to other eukaryotes (Coulson et al. 2004). Consequently, it has been suggested that post-transcriptional and post-translational gene regulation could play an important role in parasite development (reviewed in Bayer-Santos et al., 2017). However, a major pathway of post-transcriptional regulation, the miRNA pathway, is missing in *Plasmodium* (Baum et al. 2009). This does not prevent a number of miRNAs from being imported into *Plasmodium* falciparum from erythrocytes. Among them, it has been shown that two erythrocyte miRNAs, let-7i and miR-451, are imported, they bind to specific *Plasmodium* mRNAs and inhibit their translation (Lamonte et al. 2012).

In addition, it has also been shown very recently that the human miRNA-RISC complex is imported into *P. falciparum* (Dandewad et al, 2019), suggesting that this complex could act as a functional RISC complex and could thus specifically interact with *Plasmodium* mRNA and regulate their stability and translation. Thus, some of the host cell RNAs could provide indications to the parasite regarding the status of the host cell to properly adapt its gene expression. In addition, it has been suggested that some parasite non-coding RNAs, such as snoRNAs and tRNAs, may participate in such alternative RNA-silencing pathways in parasites (Garcia-Silva et al., 2010, Wang et al., 2019).

6. So much left to do…

So far, the team's results demonstrate that sporozoites, isolated from mosquito salivary glands, import exogenous (host) tRNAs. A knock-out parasite, deleted for the TRIP gene (tRip-KO), showed that in the absence of tRip, the parasite does not import tRNAs, its protein biosynthesis is significantly reduced and its infectivity is diminished at the blood stage as compared to the wild-type parasite. This phenotype suggests that host tRNAs are required for a robust translation. My PhD work belongs to a broader project which is to clarify how host tRNAs are imported into the parasite and how they affect *Plasmodium* growth and virulence. In this context, we still need to answer many questions:

- How do host tRNAs enter the parasite?
- What is the role of the *Plasmodium* MARS complex in tRNA import?
- Which tRNAs are the most susceptible to enter the parasite?
- Do imported tRNAs alter parasite gene expression?
- Can we use tRip to block the parasite infection?

MATERIAL

&

METHODS

I. Material

1. SDS-PAGE

10X TGS	250 mM Tris, 1.92M glycine, 1% SDS
1X TGS running buffer	25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3
2X loading buffer	100 mM Tris-HCl, 4% (w/v) SDS, 20% glycerol (v/v), 200 mM DTT, 0.2% (w/v) bromophenol blue
Resolving gel	12% acrylamide/bisacrylamide (19:1), 375 mM Tris-HCl, 0.1% SDS, pH 8.8.
Stacking gel	5% acrylamide/bisacrylamide (37.5:1), 150 mM Tris-HCl, 0.1% SDS, pH 6.8.
PageRuler Plus Prestained Protein Ladder	Thermo Fisher Scientific

2. Agarose gel electrophoresis

10X TBE running buffer	1 M Tris, 1 M boric acid, 0.02 M EDTA
1X TBE running buffer	100 mM Tris, 100 mM boric acid, 2 mM EDTA, pH 8.3
Agarose gel	1% (w/v) low-melting agarose in 1X TBE

3. Bacterial cultures

LB-agar plates	1% (w/v) tryptone/peptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract, 0.5% agar
LB medium	1% (w/v) tryptone/peptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract
10X PBS	1.37 M NaCl, 27 mM KCl, 100 mM Na2HPO4, 18 mM KH2PO4, pH 7.4
Ampicillin	100 $mg.mL^{-1}$ in H2O mQ
IPTG	1M in H2O mQ

4. Instruments

Centrifuge 1 L	Beckman-Coulter
Centrifuge conc	Beckman-Coulter
Centrifuge Falcons	Sorvall
Centrifuge Eppendorf	Eppendorf
Sonicator	VibraCell 75022
Sonicator	Annemasse
Optima XE-90 ultracentrifuge	Beckman Coulter
Type 70.1 Ti Fixed-Angle Titanium Rotor	Beckman Coulter
Discovery M150SE micro-ultracentrifuge	Sorvall Hitachi
S45-A Fixed-Angle Rotor	Thermo Fisher Scientific
Nanodrop ND-1000 spectrophotometer	Thermo Fisher Scientific
Biologic DuoFlow® Chromatography System	Bio-Rad
DynaPro Nanostar DLS instrument	Wyatt Technology
Quartz cuvette JC-164	Wyatt Technology
Gel Doc EZ Gel Documentation System	Bio-Rad
Mosquito nanolitre pipetting robot	TTP Labtech

5. Microorganisms

Electrocompetent E. cloni	Lucigen
Electrocompetent BL21(DE3)	NEB

6. Biomolecules

Thrombin Protease	GE Healthcare
His-TEV protease	Homemade
Total yeast tRNA	Homemade
Yeast $tRNA^{Phe}$	Homemade

7. Plasmids

pET15b	Novagen
Plasmid TEV	Homemade
pUC57-tRip	ProteoGenix
pUC57-ERS	GenScript
pUC57-QRS-C	GenScript
pUC57-QRS-N	GenScript
pUC57-MRS	ProteoGenix

8. Protein purification

Ni-IDA IMAC resin	Biorad
His-Select ® HF Ni-NTA resin	Sigma-Aldrich
1 ml Disposable SepFast™ Column	BioToolomics
Superdex™ 75 10/300 column	GE Healthcare
Superdex™ 200 increase 10/300 column	GE Healthcare
Superdex™ 200 increase 3.2/300 column	GE Healthcare
SepFast SEC 6-5000 kDa column	BioToolomics
HiTrap Benzamidine FF column	GE Healthcare
Amicon® Ultra Centrifugal Filters 10 K (0.5, 4, 15 mL)	Merck Millipore
Amicon® Ultra Centrifugal Filters 30 K (0.5, 4, 15 mL)	Merck Millipore
Amicon® Ultra Centrifugal Filters 50 K (0.5, 4, 15 mL)	Merck Millipore
Amicon® Ultra Centrifugal Filters 100 K (0.5, 4, 15 mL)	Merck Millipore
Spectra/Por™ Dialysis tubing (MWCO 12-14 kDa)	Spectrum
Filter 0.22 µm	Merck Millipore

9. Crystallization

CrystalQuick® X plates	Greiner bio-one
Index HT® Crystallization screen	Hampton research
Crystal Screen HT® Crystallization screen	Hampton research
PEG/Ion HT® Crystallization screen	Hampton research
MembFac HT® Crystallization screen	Hampton research
Natrix HT® Crystallization screen	Hampton research
JBScreen JCSG++ HTS	Jena Bioscience
Libro boxes	
Clear Seal Film	Hampton Reasearch
Vacuum grease	
22 mm siliconized glass coverslip	Hampton research
22 mm glass coverslip	
Tb-Xo4	Polyvalan

10. Softwares

ImageLab	Bio-Rad
ImageJ	NIH
DYNAMICS 7.8.1.3	Wyatt Technology
Zetasizer software	Malvern Paranalytical
Foxtrot 3.5	Xenocs
UltraScan Solution Modeller (US-SOMO) 4.0	Brookes et al. (2016)
BioXTAS RAW 2.0.3	Hopkins et al. (2017)
ATSAS 3.0.2	EMBL

II. Methods

1. Bioinformatics

The sequences (Uniprot) of *Pb*tRip (Q4Z3W3_PLABA), *Pb*ERS (A0A077XJI3_PLABA), *Pb*QRS (A0A077XKG1_PLABA), and *Pb*MRS (Q4YVD4_PLABA) obtained from co-immunoprecipitation experiments were used as BLAST query to retrieve the corresponding sequences of all *Plasmodium* strains available in PlasmoDB database (https://plasmodb.org). For each protein, 45 sequences were identified, one per parasite strain. Sequences of proteins from other organisms were manually retrieved from Uniprot database (https://www.uniprot.org/).

Multiple sequence alignments (MSA) were performed using T-Coffee software (http://tcoffee.crg.cat/) and results were visualized, adjusted, and analyzed with Jalview version 2 (http://www.jalview.org/). Detection of structural and functional domains was carried out using the Conserved Domain Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) service from NCBI in batch mode. The identification of related proteins from other organisms was performed using BLASTp and PSI-BLAST.

Sequences of *Plasmodium berghei* tRip (PBANKA_1306200), ERS (PBANKA_1362000), QRS (PBANKA_1346600) and MRS (PBANKA_0518700) were further analyzed. The sequences of the full-length protein and of individual domains were used to calculate physical and chemical parameters using the ProtParam tool from ExPASy (https://web.expasy.org/protparam/). Several structural and functional properties were predicted using PredictProtein server (https://predictprotein.org/). Secondary structure was predicted using the Quick2D tool available on the MPI Bioinformatics Toolkit (https://toolkit.tuebingen.mpg.de/tools/quick2d). Three-dimensional models were predicted using the Raptor X server (http://raptorx.uchicago.edu/).

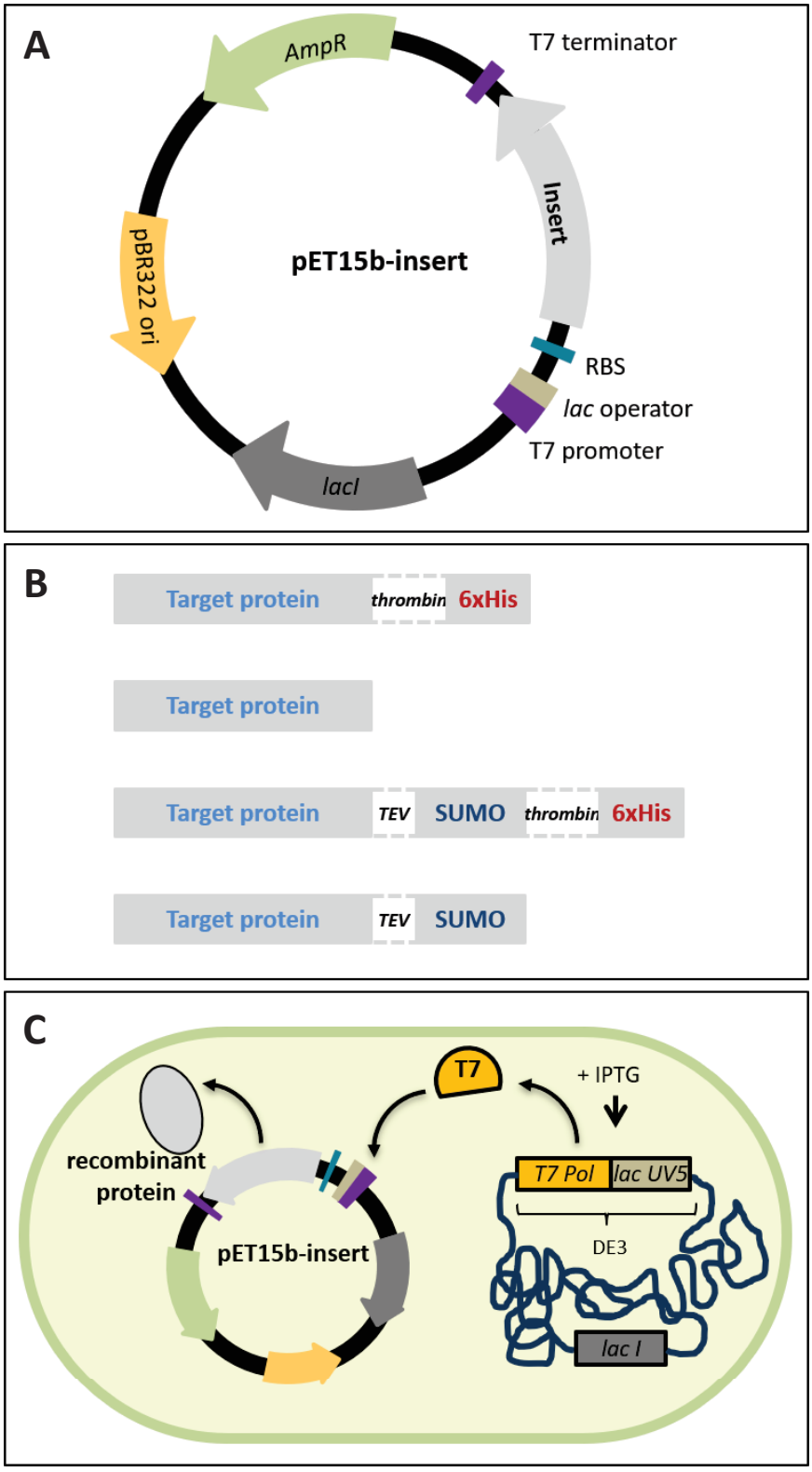


Figure 57. Recombinant expression used. A. pET15b map. The pET15b vector was used to express several target proteins (insert) cloned downstream the T7 promoter. The plasmid contains the gene for ampicillin resistance (*AmpR*) which facilitates selection of bacteria carrying the plasmid. **B. Schematic representation of the constructs.** Several types of inserts were introduced in the pET15b vector. The sequence of interest was fused to either a –His tag or a –SUMO-His tag or a –SUMO tag. Proteins without any tag were also cloned. **C. Principle of overexpression in BL21 bacteria.** IPTG induces the expression of T7 RNA polymerase in *E. coli* BL21(DE3) bacteria, which then transcribes the insert cloned in pET15b.

2. Recombinant plasmids

Several constructs of *pb*tRip, *pb*ERS, *pb*QRS, and *pb*MRS were engineered and cloned into the pET15b vector (Figure 57A). This plasmid allows IPTG-inducible overexpression of recombinant proteins in bacteria. Initially, the pET-15b vector carries an N-terminal 6xHis tag sequence followed by a thrombin site. Here, we modified this organization for most of the constructs, which display a C-terminal 6xHis tag preceded by the thrombin cleavage site. The target DNA sequences were PCR-amplified from synthetic genes (GenScript and ProteoGenix) and inserted downstream the T7 RNA polymerase promoter of pET15b. Cloning was performed using restriction enzyme-based methods (not detailed in the thesis). Recombinant plasmids were propagated using electrocompetent Top 10 *E. coli* cells (Lucigen E. cloni®). These bacteria were electroporated with the recombinant plasmid and selected on LB agar plates containing ampicillin (100 µg.mL^{-1}). Transformed bacteria were grown in LB medium (overnight at 37°C and 180 rpm) and cultures were used for plasmid extraction using the. NucleoSpin Plasmid, Mini kit for plasmid DNA (Macherey-Nagel). Plasmids containing correct sequences were verified by Sanger sequencing (Eurofins Genomics), amplified with GeneElute (TM) HP Plasmid Midiprep Kit (Sigma-Aldrich), and stored at -20°C.

Recombinant plasmids are grouped in different categories: i) full-length proteins, ii) N-terminal GST domains, iii) C-terminal RNA binding domains, iv) tRip-ERS chimeras, and v) point mutants. All these constructs are summarized in Table 9. Most of them contain a 6xHis tag (His) fused to the C-terminal extremity to ensure their purification by Ni-NTA affinity chromatography. In the case of GST domains, we also produced constructs fused to a Small Ubiquitin-like Modifier with a 6xHis tag at the C-terminus (SUMO-His) as well as constructs without any tag (Figure 57B). Protease cleavage sites were added to allow removal of 6xHis and SUMO-6xHis tags when necessary. Thrombin is used to remove 6xHis tags while the TEV protease is used to remove SUMO-6xHis tags. Additionally, point mutations were introduced in the GST domains using the QuickChange Site-Directed Mutagenesis Kit (Agilent).

Table 9. Recombinant Plasmids

ID	Plasmid name	Construct and restriction sites	Fusion tag
		Full-length proteins	
1720	His-tRip	NcoI/6xHis/thrombin/NdeI/tRip(1-403)/Stop/BamHI	His
1707	ERS-His	NcoI/ERS(1-803)/thrombin/6xHis/Stop/BamHI	His
1706	QRS-His (ins)	NheI/QRS(1-184-KL-185-852)/ClaI/thrombin/6xHis/Stop/BamHI	His
1708	MRS-His	NcoI/MRS(1-898)/ClaI/thrombin/6xHis/Stop/BamHI	His
1745	tRip	NcoI/tRipPb/Stop/BamHI	None
		N-terminal GST domains	
1721	tRip-N-His	NcoI/tRip(1-200)/KpnI/thrombin/6xHis/ Stop/BamHI	His
1760	tRip-N_{180}-His	NcoI/tRip(1-180)/KpnI/thrombin/6xHis/ Stop/BamHI	His
1701	ERS-N-His	NcoI/ERS(1-228)/ClaI/thrombin/6xHis/Stop/BamHI	His
1702	QRS-N_{208}-His	NcoI/QRS(1-208)/ClaI/thrombin/6xHis/Stop/BamHI	His
1703	MRS-N-His	NcoI/MRS(1-228)/ClaI/thrombin/6xHis/Stop/BamHI	His
1709	pET15b-SUMO	NcoI/Insert/SpeI/SUMO/KpnI/thrombin/6xHis/Stop/BamHI	SUMO-His
1717	ERS-N-SUMO-His	NcoI/ERS(1-228)/NheI/TEV/SpeI/SUMO/KpnI/thrombin/6xHis/Stop/BamHI	SUMO-His
1715	MRS-N-SUMO-His	NcoI/MRS(1-228)/NheI/TEV/SpeI/SUMO/KpnI/thrombin/6xHis/Stop/BamHI	SUMO-His
1716	QRS-N_{208}-SUMO-His	NcoI/QRS(1-208)/NheI/TEV/SpeI/SUMO/KpnI/thrombin/6xHis/Stop/BamHI	SUMO-His
1738	QRS-N-SUMO-His	NcoI/QRS(1-179)/NheI/TEV/SpeI/SUMO/KpnI/thrombin/6xHis/Stop/BamHI	SUMO-His
1739	tRip-N	NcoI/tRip(1-200)/KpnI/Stop/BamHI	None
1754	tRip-N_{180}	NcoI/tRip(1-180)/Stop/KpnI	None
1755	tRip-N_{170}	NcoI/tRip(1-170)/Stop/KpnI	None
1740	ERS-N	NcoI/ERS(1-228)/ClaI/Stop/BamHI	None
1749	QRS-N_{208}	NcoI/QRS(1-208)/NheI/Stop/KpnI	None
1744	MRS-N	NcoI/MRS(1-228)/ClaI/Stop/BamHI	None
1782	ERS-N-SUMO	NcoI/ERS(1-228)/NheI/TEV/SpeI/SUMO/KpnI/Stop/BamHI	SUMO
1741	MRS-N-SUMO	NcoI/MRS(1-228)/NheI/TEV/SpeI/SUMO/KpnI/Stop/BamHI	SUMO
1742	QRS-N_{208}-SUMO	NcoI/QRS(1-208)/NheI/TEV/SpeI/SUMO/KpnI/Stop/BamHI	SUMO
1743	QRS-N-SUMO	NcoI/QRS(1-180)/NheI/TEV/SpeI/SUMO/KpnI/Stop/BamHI	SUMO
		C-terminal RNA binding domains	
1762	His-tRip-C	NcoI/6xHis/thrombin/NdeI/tRip(201-403)/Stop/BamHI	His
1770	His-QRS-ABD-C	NcoI/6xHis/thrombin/NdeI/QRS(503-852)/Stop/BamHI	His
1771	His-QRS-ABD	NcoI/6xHis/thrombin/NdeI/QRS(503-803)/Stop/BamHI	His
1769	His-MRS-ABD-C	NcoI/6xHis/thrombin/NdeI/MRS(543-898)/Stop/BamHI	His
1768	His-MRS-ABD	NcoI/6xHis/thrombin/NdeI/MRS(543-730)/Stop/BamHI	His
1735	His-MRS-C	NcoI/6xHis/thrombin/NdeI/MRS(730-898)/Stop/BamHI	His
		tRip-ERS fusions	
1746	tRip-N-21aa-ERS-N-His	NcoI/tRip(1-200)/NheI/21aa-TEV/SpeI/ERS(1-228)/KpnI/thrombin/6xHis/Stop/BamHI	His
1751	tRip-N-14aa-ERS-N-His	NcoI/tRip(1-200)/NheI/14aa-TEV/SpeI/ERS(1-228)/KpnI/thrombin/6xHis/Stop/BamHI	His
1752	tRip-N-8aa-ERS-N-His	NcoI/tRip(1-200)/NheI/8aa-TEV/SpeI/ERS(1-228)/KpnI/thrombin/6xHis/Stop/BamHI	His
1753	tRip-N-ERS-N-His	NcoI/tRip(1-200)/SpeI/ERS(1-228)/KpnI/thrombin/6xHis/Stop/BamHI	His
1767	tRip-N_{191}-TEV-ERS-N-His	NcoI/tRiPb(1-200)/SpeI/ERS(1-228)/KpnI/thrombin/6xHis/Stop/BamHI	His
1757	tRip-N_{190}-ERS-N-His	NcoI/tRip(1-190)/SpeI/ERS(1-228)/KpnI/thrombin/6xHis/Stop/BamHI	His
1758	tRip-N_{180}-ERS-N-His	NcoI/tRip(1-180)/SpeI/ERS(1-228)/KpnI/thrombin/6xHis/Stop/BamHI	His
1759	tRip-N_{170}-ERS-N-His	NcoI/tRip(1-170)/SpeI/ERS(1-228)/KpnI/thrombin/6xHis/StopBamHI	His
1761	tRip-N-ERS-N	NcoI/tRip(1-200)/SpeI/ERS(1-228)/Stop/BamHI	None
1766	tRip-N-8aa-ERS-N	NcoI/tRip(1-200)/8aa-TEV/SpeI/ERS(1-228)/STOP/BamHI	None
1727	ERS-N-49aa-tRip-N-His	NcoI/ERS(1-228)/49aa/tRip(1-200)/KpnI/thrombin/6xHis/Stop/BamHI	His
1726	ERS-N-26aa-tRip-N-His	NcoI/ERS(1-228)/26aa/tRip(1-200)/KpnI/thrombin/6xHis/Stop/BamHI	His
1728	ERS-N-26aa-tRip-N-His	NcoI/ERS(1-228)/26aa-TEV/tRip(1-200)/KpnI/thrombin/6xHis/Stop/BamHI	His
1725	ERS-N-16aa-tRip-N-His	NcoI/ERS(1-228)/NheI/16aa-TEV/tRip(1-200)/KpnI/thrombin/6xHis/Stop/BamHI	His
1722	ERS-N-12aa-tRip-N-His	NcoI/ERS(1-228)/NheI/12aa-TEV/SpeI/tRip(1-200)/KpnI/thrombin/6xHis/Stop/BamHI	His
1764	ERS-N-tRip-N-His	NcoI/ERS(1-228)/SpeI/tRip(1-200)/KpnI/thrombin/6xHis/Stop/BamHI	His
1765	ERS-N_{206}-tRip-N-His	NcoI/ERS(1-206)/SpeI/tRip(1-200)/KpnI/thrombin/6xHis/Stop/BamHI	His
		Mutants	
1776	His-tRip(F58A)	NcoI/6xHis/thrombin/NdeI/tRip(1-403)/Stop/BamHI	His
1777	His-tRip(F90A)	NcoI/6xHis/thrombin/NdeI/tRip(1-403)/Stop/BamHI	His
1778	His-tRip(R154A)	NcoI/6xHis/thrombin/NdeI/tRip(1-403)/Stop/BamHI	His
1779	ERS-N(D95A)-SUMO-His	NcoI/ERS(1-228)/NheI/TEV/SpeI/SUMO/KpnI/thrombin/6xHis/Stop/BamHI	His
1794	ERS-N(V120R)-SUMO-His	NcoI/ERS(1-228)/NheI/TEV/SpeI/SUMO/KpnI/thrombin/6xHis/Stop/BamHI	His
1780	ERS-N(A124R)-SUMO-His	NcoI/ERS(1-228)/NheI/TEV/SpeI/SUMO/KpnI/thrombin/6xHis/Stop/BamHI	His
1795	ERS-N(F132A)-SUMO-His	NcoI/ERS(1-228)/NheI/TEV/SpeI/SUMO/KpnI/thrombin/6xHis/Stop/BamHI	His
1796	ERS-N(H148A)-SUMO-His	NcoI/ERS(1-228)/NheI/TEV/SpeI/SUMO/KpnI/thrombin/6xHis/Stop/BamHI	His
1781	ERS-N(R198A)-SUMO-His	NcoI/ERS(1-228)/NheI/TEV/SpeI/SUMO/KpnI/thrombin/6xHis/Stop/BamHI	His
1789	ERS-N(TNLY/GSGS)-SUMO-His	NcoI/ERS(1-228)/NheI/TEV/SpeI/SUMO/KpnI/thrombin/6xHis/Stop/BamHI	His
1790	ERSN(D95A-A124R)SUMO-His	NcoI/ERS(1-228)/NheI/TEV/SpeI/SUMO/KpnI/thrombin/6xHis/Stop/BamHI	His
1783	QRS-N_{208}(V66R)	NcoI/QRS(1-208)/NheI/Stop/KpnI	None
1784	QRS-N_{208}(K97A)	NcoI/QRS(1-208)/NheI/Stop/KpnI	None
1785	QRS-N_{208}(E102A)	NcoI/QRS(1-208)/NheI/Stop/KpnI	None
1791	QRS-N_{208}(V66R-K97A)	NcoI/QRS(1-208)/NheI/Stop/KpnI	None
1792	QRS-N_{208}(V66R-E102A)	NcoI/QRS(1-208)/NheI/Stop/KpnI	None
1797	MRS-N(N48A)	NcoI/MRS(1-228)/ClaI/Stop/BamHI	None
1798	MRS-N(K50A)	NcoI/MRS(1-228)/ClaI/Stop/BamHI	None
1786	MRS-N(A65R)	NcoI/MRS(1-228)/ClaI/Stop/BamHI	None
1787	MRS-N(E95A)	NcoI/MRS(1-228)/ClaI/Stop/BamHI	None
1788	MRS-N(R191A)	NcoI/MRS(1-228)/ClaI/Stop/BamHI	None
1793	MRS-N(A65R-E95A)	NcoI/MRS(1-228)/ClaI/Stop/BamHI	None

3. Expression of recombinant proteins

E.coli BL21(DE3) bacteria (NEB) were used to produce the recombinant proteins cloned in pET15b (Figure 57C). This strain contains the lysogen DE3 that carries the gene for T7 RNA polymerase under the control of a *lac*UV5 promoter, which is inducible by isopropyl-β-D-thiogalactopyranoside (IPTG). Adding IPTG to a growing culture of BL21(DE3) induces the production of T7 RNA polymerase, which in turn transcribes the target DNA cloned in pET15b.

Cultures for protein production were systematically started from freshly transformed bacteria. Competent BL21(DE3) cells were electroporated (Bio-Rad Micropulser Electroporator) with recombinant plasmid and spread on LB agar plates containing ampicillin (100 $\mu g.mL^{-1}$) for selection. After incubation overnight at 37°C, all colonies were scraped-off and suspended in 10 mL of LB medium. The OD_{600} was measured on a 1:10 dilution and the suspension was used to inoculate a given volume of LB medium with ampicillin (100 $\mu g.mL^{-1}$) to a starting OD_{600} of 0.015. The culture was incubated at 30°C and 180 rpm until an OD_{600} of 0.8 was reached (usually 4 to 5 hours). Then, IPTG was added to a final concentration of 0.5 mM and the culture was continued at 16°C and 180 rpm for 16-18 h. Bacteria were harvested in different ways depending on the application:

a) *Pull-down assays*: 100 mL cultures were distributed in several tubes and centrifuged 20 min at 4000 g and 4°C. Supernatant was discarded and the bacterial pellets were stored at -20°C until use. The size of the aliquots depends on the production and the solubility of the protein used. Usually, aliquots of 1 mL were used for *Pb*tRip or *Pb*ERS-N constructs (high expression and high solubility) while aliquots of 4 mL were needed for *Pb*QRS-N and *Pb*MRS-N constructs (high expression but poor solubility).

b) *Protein purification*: Bacteria from 1 liter of culture were pelleted by centrifugation (20 min at 4000 g and 4°C), washed in PBS, and stored at -20°C until use. Aliquots of 250 mL were used for *Pb*tRip and *Pb*ERS-N and aliquots of 500 mL were needed for *Pb*QRS-N and *Pb*MRS-N.

The expression of recombinant proteins was verified systematically by SDS-PAGE. Bacteria from 2 mL of culture were pelleted by centrifugation (5 min at 4000 g and 4°C) and suspended in 500 µL of A2A4 buffer. Cells were disrupted by sonication (20 s at amplitude 40 in sonicator Vibra Cell 75022) and the crude extract was centrifuged (15 min at 15000 g and 4°C) to remove insoluble material. Crude and centrifuged extracts (5 µL) were analyzed by SDS-PAGE.

Table 10. Content of R, H and K of different protein constructions

ID	Protein	Protease digestion	# R	# H	# K	R+H+K
tRip constructs						
1720	His-tRip	None	13	15	44	72
		Thrombin	12	9	44	65
1721	tRip-N-His	None	5	12	21	38
		Thrombin	5	6	21	32
1760	tRip-N_{180}-His	None	5	12	20	37
		Thrombin	5	6	20	31
1745	tRip	None	12	8	44	64
1739	tRip-N	None	4	6	21	31
1754	tRip-N_{180}	None	4	6	20	30
1755	tRip-N_{170}	None	3	5	19	27
ERS constructs						
1701	ERS-N-His	None	9	12	25	46
		Thrombin	9	6	25	40
1717	ERS-N-SUMO-His	None	5	14	34	63
		Thrombin	15	8	34	57
		TEV	8	6	25	39
1740	ERS-N	None	8	6	25	39
1782	ERS-N-SUMO	None	14	8	34	56
		TEV	8	6	25	39
QRS constructs						
1702	QRS-N-His	None	4	12	28	44
		Thrombin	4	6	28	38
1716	QRS-N-SUMO-His	None	10	14	37	61
		Thrombin	10	8	37	55
		TEV	3	6	28	37
1738	QRS-N_{179}-SUMO-His	None	10	12	32	54
		Thrombin	10	6	32	48
		TEV	3	4	23	30
1749	QRS-N	None	3	6	28	37
1742	QRS-N-SUMO	None	9	8	37	54
		TEV	3	6	28	37
1743	QRS-N_{179}-SUMO	None	9	6	32	47
		TEV	3	4	23	30
MRS constructs						
1703	MRS-N-His	None	6	11	26	43
		Thrombin	6	5	26	37
1715	MRS-N-SUMO-His	None	12	13	35	60
		Thrombin	12	7	35	54
		TEV	5	5	26	36
1744	MRS-N	None	5	5	26	36
1741	MRS-N-SUMO	None	11	7	35	53
		TEV	5	5	26	36
tRip-ERS constructs						
1746	tRip-N-21aa-ERS-N-His	None	13	18	46	77
		Thrombin	13	12	46	71
1752	tRip-N-8aa-ERS-N-His	None	13	18	46	77
		Thrombin	13	12	46	71
1753	tRip-N-ERS-N-His	None	13	18	46	77
		Thrombin	13	12	46	71
1766	tRip-N-8aa-ERS-N	None	12	12	46	70
1761	tRip-N-ERS-N	None	13	12	45	70

4. *In vitro* pull-down assays

Interaction between *Pb*tRip and the GST domains of *Pb*ERS, *Pb*QRS and *Pb*MRS were examined by *in vitro* pull-down assays using a Ni-NTA affinity resin. Several types of experiments were performed using this approach: a) pairwise interactions, b) formation of ternary complexes, c) interactions with four proteins, d) competition assays and e) effect of point mutations.

One aliquot of bacteria expressing a 6xHis-tagged protein (bait) together with one or more aliquots of bacteria expressing non-tagged proteins (preys) were suspended in a total volume of 1200 µL pull-down buffer (composition in Table 11) supplemented with 0.005% DDM. Cells were disrupted by sonication (2 pulses of 10 s at amplitude 40 in Sonicator VibraCell 75022) and the crude extract was centrifuged (15 min at 15000g and 4°C) to remove cell debris. The centrifuged extract was recovered in a 2.2 mL tube and incubated with 75 µL of a Ni-IDA resin (Bio-Rad Profinity® IMAC, washed and equilibrated with DDM-containing pull-down buffer) at room temperature under agitation in a tube rotator (VWR) for 30 min. The suspension was transferred to an empty 10 mL chromatography column (C2103-200EA, Sigma-Aldrich) and the resin was washed twice with 2.5 mL of buffer with DDM and once with 2.5 mL of buffer without DDM. Proteins specifically bound to the resin were eluted with 100 µL of pull-down buffer containing 250 mM imidazole. Finally, 2.5 µL of the centrifuged extract and of the eluted proteins were loaded and analyzed on a 12% (19:1) SDS-PAGE.

The relative amount of proteins found in the elution was determined using Image J software (Abràmoff et al., 2004). These values are proportional to the amount of protein but are also influenced by the capacity of the protein to bind the Coomassie blue (R-250). To account for this effect, the area of each band was divided by the sum of arginines (R), histidines (H), and lysine (K) present in the proteins (Table 10). Error bars were calculated from the results of at least 3 replicates.

Table 11. Composition of buffer for Pull Down experiments

Buffer name	Component	Concentration of stock	Volume of stock	Final concentration
Pull-down	HEPES-NaOH pH 8.0	1 M	25 mL	50 mM
	NaCl	5 M	30 mL	300 mM
	Glycerol	50 % (v/v)	100 mL	10 % (v/v)
	2-mercaptoethanol	14.3 M	180 µL	5.148 mM
	H_2O mQ	-	**qsp 500 mL**	-

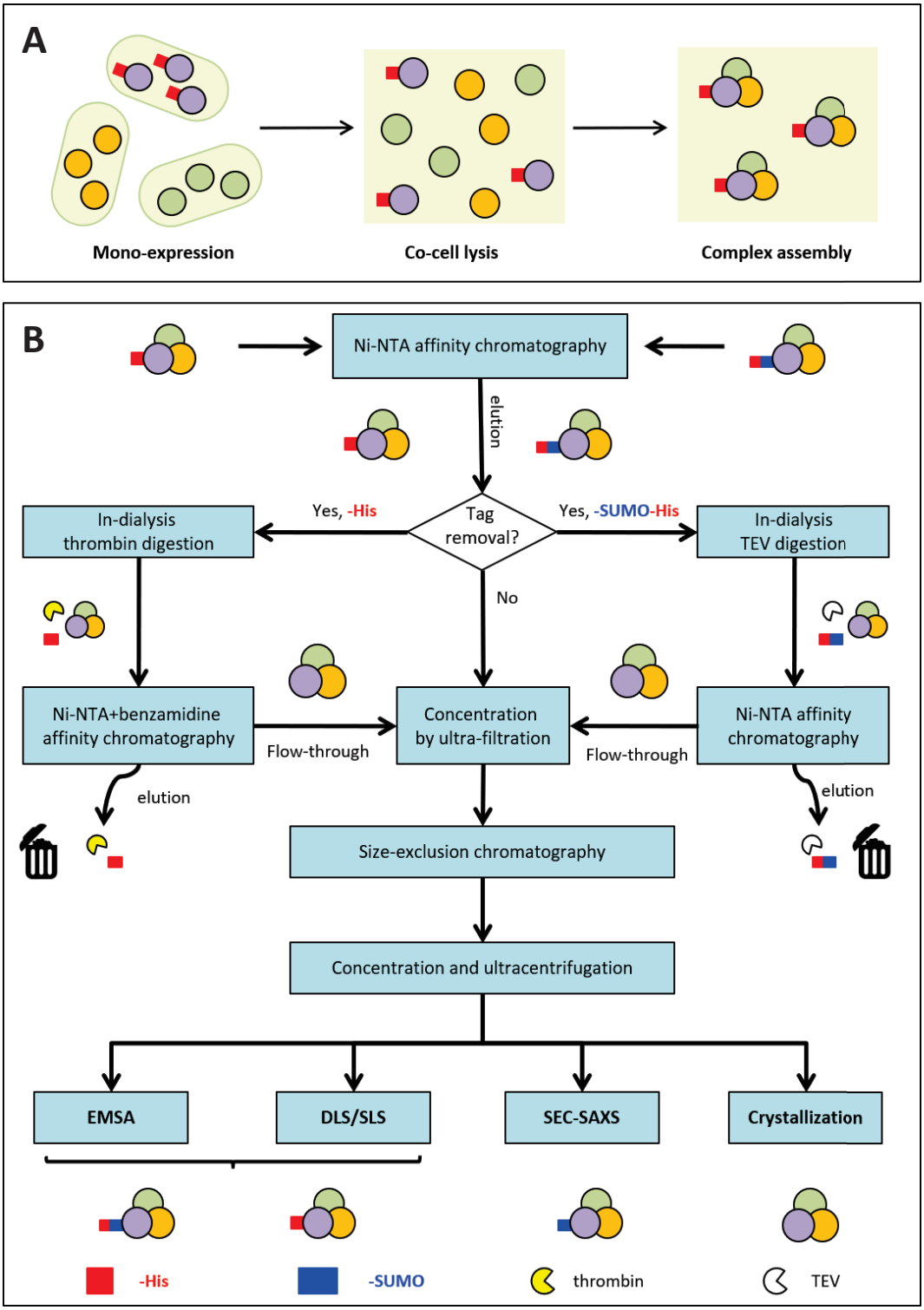


Figure 58. Reconstitution and purification of *Plasmodium* MARS complexes. A. Reconstitution of complexes by co-lysis. Protein partners were expressed individually in bacteria. The different bacteria were mixed and lyzed together in order to allow the spontaneous assembly of complexes **B. Purification workflow.** Ni-NTA column captures his-tagged complexes. When the presence of -His was compatible with the final application (e.g. EMSA or DLS/SLS), the complex was concentrated and injected directly into a size-exclusion column. When tag removal was preferred (e.g. crystallization), a treatment with thrombin or TEV was performed. A second step of affinity chromatography captured both cleaved tags and protease and the untagged complex are recovered in the flow-through. The complex is concentrated, subjected to gel filtration, concentrated again and ultracentrifuged before SEC-SAXS or crystallization assays.

5. Reconstitution and purification of *Pb*MARS complexes

6xHis-tagged domains and complexes were purified using Ni-affinity chromatography followed by additional purification steps depending on the purpose of the protein sample (Figure 58). All chromatography columns were performed at low temperature (4-10°C) using BioLogic DuoFlow® Chromatography Systems (Bio-Rad), each purification step was checked by SDS-PAGE and contamination of purified samples with nucleic acids was assessed using A_{260}/A_{280} ratio (Nanodrop).

5.1. Purification of individual proteins

Culture pellets were thawed in ice and suspended in 25 mL of buffer A1 (Table 12). Cells were disrupted by sonication 7 min at 120 V in ice (Ultrasons Annemasse) and the crude extract was ultracentrifuged 45 min at 45000 g and 4°C (Beckman-Coulter Optima XE-90). The clarified extract was loaded onto a 1 mL Ni-NTA (Sigma-Aldrich His-Select® HF) column equilibrated with buffer A1 and several washing steps were performed: (i) 15 mL buffer A1, (ii) 6 mL linear gradient from 0 to 100% of buffer B-NaCl (300 mM to 2 M NaCl), (iii) 6 mL inverted gradient back to 0% of buffer B-NaCl and (iv) 33 mL buffer A1. The washing step with NaCl reduced the nucleic acid contamination in the samples. Elution of His-tagged proteins bound to the column was performed with a 25 mL linear gradient from 4 to 50% buffer B-Imidazole (from 20 to 250 mM imidazole). Fractions containing the purified His-tagged proteins were pooled and concentrated to a final volume of 500-800 µL using an appropriate centrifugal filter (Merck MilliporeAmicon®). The protein was injected onto a Superdex® 200 Increase 10/300 column (GE Healthcare) or a Superdex® 75 10/300 column and eluted with SEC buffer (pH 8.0 or 7.0) at a flow rate of 0.4 $mL.min^{-1}$. Fractions containing pure proteins (SDS-PAGE and $A_{260}/A_{280} \approx 0.5$) were pooled and concentrated. C-terminal domains of MRS and QRS were purified according to the same protocol, except that the cell lysis was performed in the presence of a cocktail of protease inhibitors (BS380, Bio Basic) and that buffers did not contain DDM.

5.2. Purification of *Pb*tRip complexes

Complexes were reconstituted by cellular "co-lysis" (Figure 58A). One pellet of bacteria expressing a given 6xHis-tagged protein (bait) and several pellets of bacteria expressing non-tagged proteins (preys) were suspended in A2A4 buffer (25 mL for 500 mL culture pellet), mixed together and the cells were sonicated in ice (7 min at 120 V for 25 mL). Thereafter, the samples were treated in the same general manner as for the purification of individual proteins (see 5.1) (Figure 58B). Only two steps were modified: (i) the amplitude of the NaCl gradient was reduced

(300 mM to 1M) to avoid protein dissociation and (ii) complexes larger than 150 kDa were further purified on a SepFast® SEC 6-5000 kDa column (BioToolomics) and eluted with SEC buffer pH 7.0 at 0.2 mL.min^{-1}. As an additional quality criterion, the relative amounts of the different proteins present in the complex were determined by SDS-PAGE quantification as described in section 4. Sometimes, we observed that regardless of the size of the column, the resin did not always capture the total amount of complexes, thus the column flow-through was recovered and subjected to another round of chromatography when necessary.

5.3. Cleavage of 6xHis-tag and SUMO

If proteins or complexes without any tag are necessary (e.g. crystallization assays), additional purification steps were added. After elution from the Ni-NTA column, the protein fractions are pooled and dialyzed overnight at 4°C against buffer A2A4 in presence of the required protease: thrombin (GE Healthcare) was used to cleave the 6xHis tag (10 U for 100 µg protein) and 6xHis-tagged TEV (homemade) was used to remove the SUMO-His domain (1 µg for 25µg protein). The next day, the sample was rerun on a Ni-NTA column in buffer A2A4 to recover only the cleaved proteins in the flow-through and the washing fractions, the 6xHis-tagged TEV remaining on the resin. Alternatively, when using thrombin, a 1 mL HiTrap Benzamidine FF column (GE Healthcare) was placed downstream of the Ni-NTA column to remove the protease from the sample. The fractions were pooled, concentrated, and subjected to gel filtration as described previously.

5.4. Determination of protein concentration

The concentration of protein samples was determined using the NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific). The absorbance at 280 nm of 2 µL of sample was measured using the application "Protein A280" and the concentration in mg.mL^{-1} was provided. The extinction coefficient ε ($M^{-1}cm^{-1}$) × 1000 and the molecular weight M_w (in kDa) are required to calculate the accurate concentration of the protein sample:

$$C_{mg/ml} = \left(\frac{A_{280}}{\varepsilon \times 1000}\right) \times M_W$$

The extinction coefficients and molecular weights were obtained from the protein sequences using the ExPASy's ProtParam tool (Table 13 and 14). The quality of the sample was estimated using the A_{260}/A_{280} ratio, which should range around 0.5 for pure protein solutions. Higher values suggest nucleic acid contamination.

Table 12. Buffers for purification of complexes

Buffer name	Component	Concentration of stock	Volume of stock	Final concentration
A1	HEPES-NaOH pH 8.0	1 M	25 mL	50 mM
	NaCl	5 M	30 mL	300 mM
	Glycerol	50 % (v/v)	100 mL	10 % (v/v)
	Imidazole	2 M	5 mL	20 mM
	DDM	10 % (w/v)	250 µL	0.005 % (m/v)
	2-mercaptoethanol	14.3 M	180 µL	5.148 mM
	H_2O mQ	-	**qsp 500 mL**	-
A2A4	HEPES-NaOH pH 8.0	1 M	25 mL	50 mM
	NaCl	5 M	30 mL	300 mM
	Glycerol	50 % (v/v)	100 mL	10 % (v/v)
	DDM	10 % (w/v)	250 µL	0.005 % (m/v)
	2-mercaptoethanol	14.3 M	180 µL	5.148 mM
	H_2O mQ	-	**qsp 500 mL**	-
A3	HEPES-NaOH pH 8.0	1 M	12.5 mL	50 mM
	NaCl	5 M	15 ml	300 mM
	Glycerol	50 % (v/v)	50 mL	10 % (v/v)
	Imidazole	2 M	31.25 mL	250 mM
	DDM	10 % (w/v)	125 µL	0.005 % (m/v)
	2-mercaptoethanol	14.3 M	90 µL	5.148 mM
	H_2O mQ	-	**qsp 250 mL**	-
B-NaCl	HEPES-NaOH pH 8.0	1 M	12.5 mL	50 mM
	NaCl	5 M	100 mL	2 M
	Glycerol	50 % (v/v)	50 mL	10 % (v/v)
	Imidazole	2 M	2.5 mL	20 mM
	DDM	10 % (w/v)	250 µL	0.005 % (m/v)
	2-mercaptoethanol	14.3 M	180 µL	5.148 mM
	H_2O mQ	-	**qsp 250 mL**	-
B-imidazole	HEPES-NaOH pH 8.0	1 M	12.5 mL	50 mM
	NaCl	5 M	15 mL	300 mM
	Glycerol	50 % (v/v)	50 mL	10 % (v/v)
	Imidazole	2 M	62.5 mL	500 mM
	DDM	10 % (w/v)	125 µL	0.005 % (m/v)
	2-mercaptoethanol	14.3 M	90 µL	5.148 mM
	H_2O mQ	-	**qsp 250 mL**	-
SEC pH 8.0	HEPES-NaOH pH 8.0	1 M	6.25 mL	25 mM
	NaCl	5 M	15 mL	300 mM
	Glycerol	50 % (v/v)	25 mL	5 % (v/v)
	DDM	10 % (w/v)	125 µL	0.005 % (m/v)
	2-mercaptoethanol	14.3 M	90 µL	5.148 mM
	H_2O mQ	-	**qsp 250 mL**	-
SEC pH 7.0	HEPES-NaOH pH 7.0	1 M	6.25 mL	25 mM
	NaCl	5 M	15 mL	300 mM
	Glycerol	50 % (v/v)	25 mL	5 % (v/v)
	DDM	10 % (w/v)	125 µL	0.005 % (m/v)
	2-mercaptoethanol	14.3 M	90 µL	5.148 mM
	H_2O mQ	-	**qsp 250 mL**	-

Table 13. ProtParam – Individual proteins

Bacteria Pellet	Culture volume (mL)	Protein name	Protease digestion	MW (kDa)	ε × 1000 ($M^{-1}cm^{-1}$)	pI	Centrifugal filer MWCO
Full-length proteins							
1720	250	His-tRip	None	48.33715	34.840	7.20	30 K
			Thrombin	46.4559	34.840	6.83	
1707	250	ERS-his	None	96.05478	113.570	8.91	50 K
			Thrombin	95.00073	113.570	8.91	
1706	500	QRS-his	None	103.04297	119.640	7.80	50 K
			Thrombin	101.98892	119.640	7.78	
1708	500	MRS-his	None	106.85231	123.080	7.40	50 K
			Thrombin	105.79826	123.080	7.36	
N-terminal GST domains							
1721	250	tRip-N-his	None	25.60727	23.380	6.17	10 K
			Thrombin	24.55322	23.380	5.72	
1760	250	tRip-N180-his	None	23.37384	23.380	6.30	10 K
			Thrombin	22.31978	23.380	5.88	
1701	250	ERS-N-his	None	29.60183	34.840	9.38	10 K
			Thrombin	28.54778	34.840	9.38	
1702	500	QRS-N_{208}-his	None	27.26394	23.950	6.64	10 K
			Thrombin	26.20989	23.950	6.34	
1703	500	MRS-N-his	None	28.70702	32.890	8.07	10 K
			Thrombin	27.65297	32.890	8.05	
1717	250	ERS-N-SUMO-his	None	42.22992	37.820	8.77	10 K
			Thrombin	41.17587	37.820	8.77	
			TEV	28.80498	36.330	9.30	
1715	500	MRS-N-SUMO-his	None	40.99177	35.870	6.37	10 K
			Thrombin	39.93772	35.870	6.06	
			TEV	27.56683	34.380	7.55	
1716	500	QRS-N_{208}-SUMO-his	None	39.66185	26.930	6.00	10 K
			Thrombin	38.60780	26.930	5.66	
			TEV	26.23691	25.440	6.12	
1738	500	QRS-N-SUMO-his	None	36.31805	26.930	5.89	10 K
			Thrombin	35.26399	26.930	5.49	
			TEV	22.89311	25.440	5.90	
C-terminal RNA binding domains							
1762		His-tRip-C	None	24.90220	11.460	9.07	10 K
			Thrombin	23.02015	11.460	8.94	
1770		His-QRS-ABD-C	None	43.21464	33.350	6.90	
			Thrombin	41.33258	33.350	6.47	
1771		His-QRS-ABD	None	37.28759	33.350	5.93	10 K
			Thrombin	35.40554	33.350	5.43	
1769		His-MRS-ABD-C	None	43.38332	31.860	8.35	10 K
			Thrombin	41.50127	31.860	8.12	
1768		His-MRS-ABD	None	24.35820	22.920	6.91	10 K
			Thrombin	22.47615	22.920	6.47	
1735		His-MRS-C	None	21.48482	8.940	9.22	10 K
			Thrombin	19.60277	8.940	9.06	
tRip-ERS constructs							
1746	500	tRip-N-21aa-ERS-N-His	None	55.42933	59.710	8.23	30 K
			Thrombin	54.37528	59.710	8.23	
1752	500	tRip-N-8aa-ERS-N-His	None	54.34628	59.710	8.25	30 K
			Thrombin	53.29223	59.710	8.25	
1753	500	tRip-N-ERS-N-His	None	53.62151	58.220	8.39	30 K
			Thrombin	52.56745	58.220	8.38	
1766	500	tRip-N-8aa-ERS-N	None	52.26710	59.710	8.25	30 K
1761	500	tRip-N-ERS-N	None	51.57034	58.220	8.39	30 K

Table 14. ProtParam – Complexes

Bacteria pellets	Culture volume (mL)	Complex name	Protease digestion	MW (kDa)	ε × 1000 ($M^{-1}cm^{-1}$)	pI	Centrifugal filer MWCO
1703	500	MRS-N-His-His:ERS-N:tRip	None	102.63874	102.570	8.52	100 K
1740	250		Thrombin	101.58468	102.570	8.52	
1745	250		TEV	-	-	-	
1703	500	MRS-N-His:ERS-N:tRip-N	None	80.35135	91.110	8.28	50 K
1740	250		Thrombin	79.29730	91.110	8.28	
1739	250		TEV	-	-	-	
1715	500	MRS-N-SUMO-His:ERS-N:tRip	None	114.92349	105.550	8.09	100 K
1740	250		Thrombin	113.86943	105.550	8.08	
1745	250		TEV	101.49855	104.060	8.46	
1715	500	MRS-N-SUMO-His:ERS-N:tRip-N	None	92.63610	94.090	7.29	50 K
1740	250		Thrombin	91.58205	94.090	7.23	
1739	250		TEV	79.21116	92.600	8.18	
1702	500	QRS-N_{208}-His:ERS-N:tRip	None	101.19566	93.630	8.34	100 K
1740	250		Thrombin	100.14160	93.630	8.34	
1745	250		TEV	-	-	-	
1702	500	QRS-N_{208}-His:ERS-N:tRip-N	None	78.90827	82.170	7.80	50K
1740	250		Thrombin	77.85422	82.170	7.78	
1739	250		TEV	-	-	-	
1716	500	QRS-N_{208}-SUMO-His:ERS-N:tRip	None	113.59357	96.610	7.53	100 K
1740	250		Thrombin	112.53951	96.610	7.49	
1745	250		TEV	100.16862	95.120	8.26	
1716	500	QRS-N_{208}-SUMO-His:ERS-N:tRip-N	None	91.30618	85.150	6.72	50 K
1740	250		Thrombin	90.25213	85.150	6.61	
1739	250		TEV	77.88124	83.660	7.52	
1738	500	QRS-N-SUMO-His:ERS-N:tRip	None	110.24976	96.610	7.52	100 K
1740	250		Thrombin	109.19571	96.610	7.48	
1745	250		TEV	96.82482	95.120	8.27	
1738	500	QRS-N-SUMO-His:ERS-N:tRip-N	None	87.96238	85.150	6.68	50 K
1740	250		Thrombin	86.90833	85.150	6.56	
1739	250		TEV	74.53744	83.660	7.50	
1717	250	ERS-N-SUMO-His:tRip:MRS-N	None	115.26633	105.550	8.09	100 K
1745	250		Thrombin	114.21277	105.550	8.08	
1744	500		TEV	101.84189	104.060	8.46	
1717	250	ERS-N-SUMO-His:tRip:QRS-N_{208}	None	113.82375	96.610	7.53	100 K
1745	250		Thrombin	112.76969	96.610	7.49	
1749	1000		TEV	100.39881	95.120	8.26	
1717	250	ERS-N-SUMO-His:tRip:MRS-N + ERS-N-SUMO-His:tRip:QRS-N_{208}	None	114.54505	101.080	7.81	100 K
1745	250		Thrombin	113.49123	101.080	7.785	
1744	500		TEV	101.12035	99.55	8.36	
1749	1000		-	-	-	-	
1717	250	ERS-N-SUMO-His:tRip-N:MRS-N	None	92.97944	94.090	7.29	50 K
1739	250		Thrombin	91.92539	94.090	7.23	
1744	500		TEV	79.55450	92.600	8.18	
1717	250	ERS-N-SUMO-His:tRip-N:QRS-N_{208}	None	91.53636	85.150	6.72	50 K
1739	250		Thrombin	90.48231	85.150	6.61	
1749	1000		TEV	78.11142	83.660	7.52	
1717	250	ERS-N-SUMO-His:tRip-N:MRS-N + ERS-N-SUMO-His:tRip-N:QRS-N_{208}	None	92.2579	89.62	7.405	50 K
1739	250		Thrombin	91.20385	89.62	6.92	
1744	500		TEV	78.83296	88.13	7.85	
1749	1000		-	-	-	-	
1717	250	ERS-N-SUMO-His:tRip:MRS-N:QRS-N_{208} (Caution! Complex does not exist)	None	140.64779	129.500	7.66	100 K
1745	250		Thrombin	139.59374	129.500	7.64	
1744	500		TEV	127.22285	128.010	8.21	
1749	1000		-	-	-	-	
1717	250	ERS-N-SUMO-His:tRip-N:MRS-N:QRS-N_{208} (Caution! Complex does not exist)	None	118.36041	118.040	6.90	50 K
1739	250		Thrombin	117.30635	118.040	6.82	
1744	500		TEV	104.93547	116.550	7.68	
1749	1000		-	-	-	-	

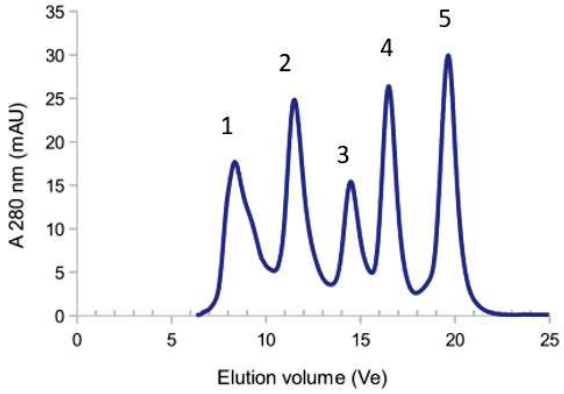
A

Column: Superdex 200 Increase 10/300

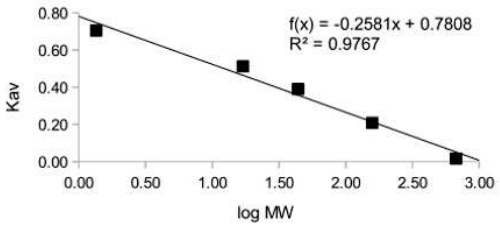
Sample:

1. Thyroglobulin (670 kDa), 0.2 mg.mL^{-1}
2. γ-globlin (158 kDa), 0.2 mg.mL^{-1}
3. Ovalbumin (44 kDa), 0.2 mg.mL^{-1}
4. Myoglobin (17 kDa), 0.1 mg.mL^{-1}
5. Vitamin B12 (1.35 kDa), 0.02 mg.mL^{-1}

Sample volume: 500 µL
Flow rate: 0.4 mL.min^{-1}
Buffer: SEC pH 7.0
System: BioLogic DuoFlow



Protein	MW (kDa)	Log MW	Ve (mL)	Kav
Thyroglobulin	670	2.83	8.35	0.02
γ-globulin	158	2.20	11.49	0.21
Ovalbumin	44	1.64	14.48	0.39
Myoglobin	17	1.23	16.49	0.51
Vitamin B12	1.35	0.13	19.65	0.70
Blue dextran	-	-	8.073	0.00
Imidazole	-	-	24.5	1.00



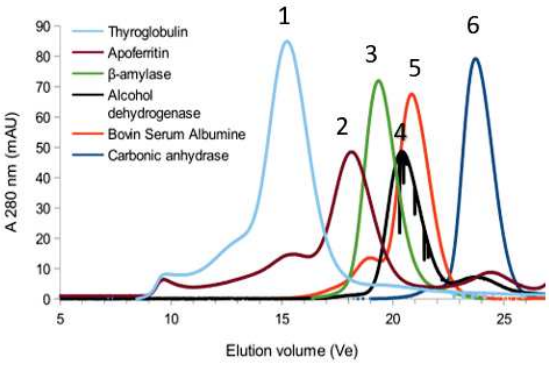
B

Column: SepFast SEC 11/300 6-5000 kDa

Sample:

1. Thyroglobulin (669 kDa), 0.8 mg.mL^{-1}
2. Apoferritin (443 kDa), 1 mg.mL^{-1}
3. β-amylase (200 kDa), 0.4 mg.mL^{-1}
4. Alcohol dehydrogenase (150 kDa), 0.5 mg.mL^{-1}
5. Bovine serum albumin (66 kDa), 1 mg.mL^{-1}
6. Carbonic anhydrase (29 kDa), 0.3 mg.mL^{-1}

Sample volume: 500 µL
Flow rate: 0.2 mL.min^{-1}
Buffer: SEC pH 7.0
System: BioLogic DuoFlow



Protein	MW (kDa)	Log MW	Ve (mL)	Kav
Thyroglobulin	669	2.83	15.23	0.24
Apoferritin	443	2.65	18.14	0.36
β-amylase	200	2.30	19.36	0.42
Alcohol dehydrogenase	150	2.18	20.38	0.46
Bovine serum albumin	66	1.82	20.85	0.48
Carbonic anhydrase	29	1.46	23.73	0.60
Blue dextran	-	-	9.56	0.00
Imidazole			33.06	1.00

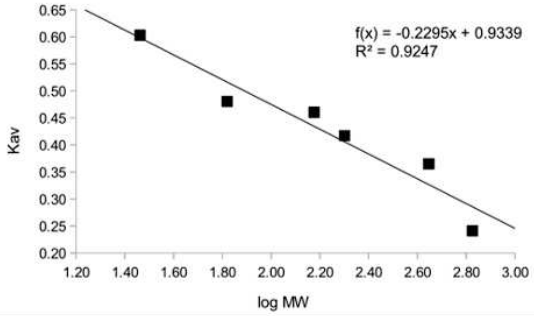


Figure 59. Calibration of size exclusion columns. **A. Calibration of Superdex 200 increased 10/300.** Calibration was performed with Bio-Rad's gel filtration standard containing a mixture of 5 proteins of different molecular weights. **B. Calibration of SepFast SEC 11/300 6-5000 kDa.** It was calibrated using individual protein markers purchased from Sigma-Aldrich. For of both columns, the void volume and total volume was determined using a mixture of blue dextran and imidazole. A calibration curve was obtained by plotting the distribution coefficient Kav of each marker *vs* its logarithmic molecular weight.

6. Characterization of biomolecules in solution

6.1. Size-exclusion chromatography (SEC)

6.1.1. Molecular weight estimation

SEC analyses were performed on the BioLogic DuoFlow® Chromatography System (Bio-Rad). In general, a Superdex® 200 increase 10/300 column (GE Healthcare) was used for individual proteins and a SepFast® SEC 11/300 6-5000 kDa column (Bio Toolomics Ltd.) was used for complexes, as its separation range was more appropriate. Both columns were periodically calibrated to determine molecular weight (MW) estimates from the SEC chromatograms.

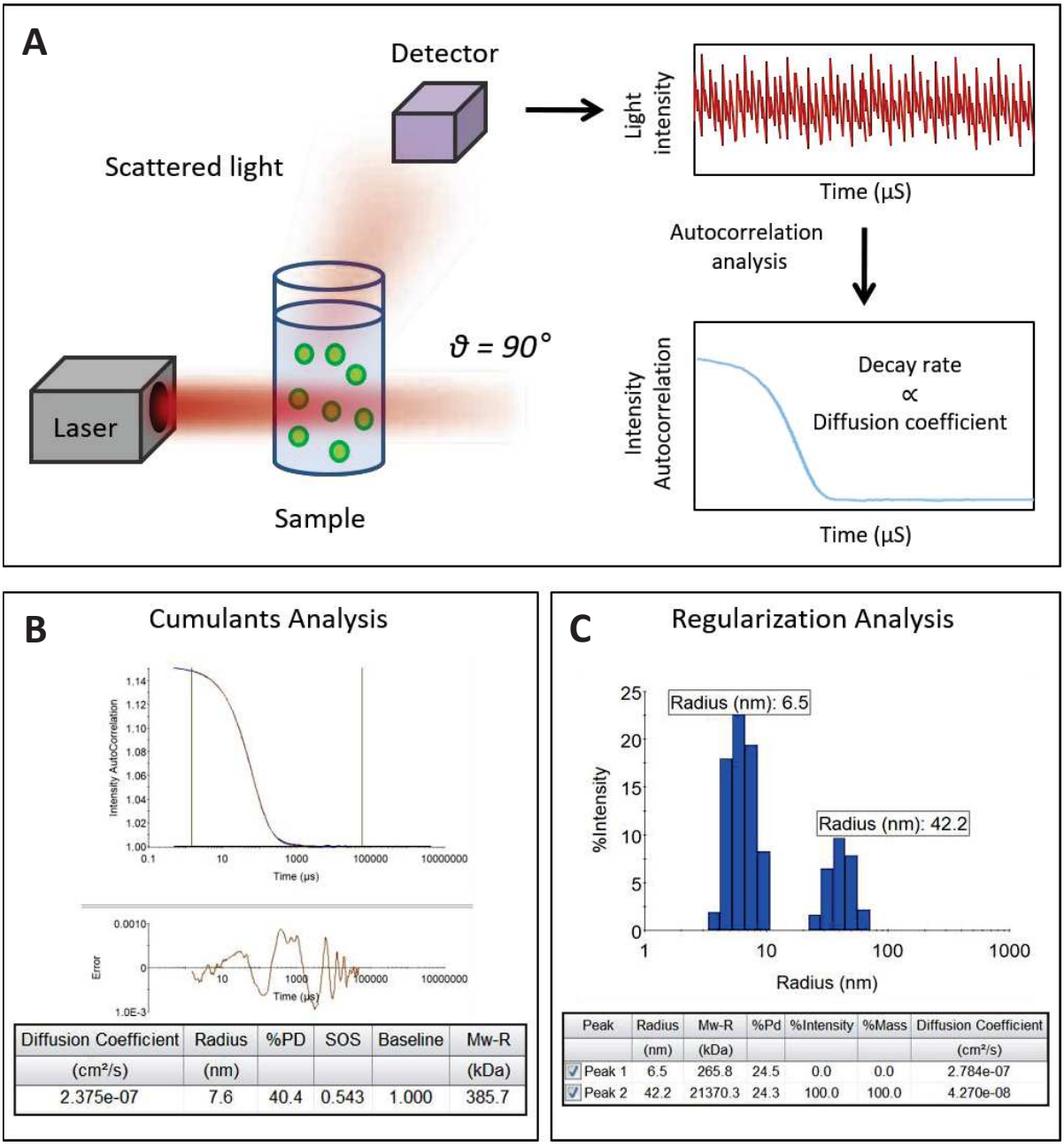
The Superdex 200 column was calibrated with 10 µL of Bio-Rad's gel filtration standard (#1511901). This solution contains 10 $mg.mL^{-1}$ thyroglobulin (670 kDa), 10 $mg.mL^{-1}$ γ-globulin (158 kDa), 10 $mg.mL^{-1}$ ovalbumin (44 kDa), 5 $mg.mL^{-1}$ myoglobin (17 kDa) and 1 $mg.mL^{-1}$ vitamin B12 (1.35kDa). In addition, a solution of blue dextran (about 500 µL at 1 $mg.mL^{-1}$) and imidazole (200 mM) was used to determine the column's void volume (V_0) and total bed volume (V_t).

The SepFast column was calibrated using the MWGF1000 kit (Sigma-Aldrich), which includes individual molecular weight protein markers. They were dissolved in buffer 50 mM Tris-HCl pH 7.5 and 100 mM KCl to achieve the following concentrations: 8 $mg.mL^{-1}$ thyroglobulin (669 kDa), 4 $mg.mL^{-1}$ β-amylase (200 kDa), 5 $mg.mL^{-1}$ alcohol dehydrogenase (150 kDa), 10 $mg.mL^{-1}$ BSA (66 kDa) and 3 $mg.mL^{-1}$ carbonic anhydrase (29 kDa). Only apoferritin (443 kDa) was supplied as a solution (25 $mg.mL^{-1}$ in 50% glycerol and 75 mM NaCl). Each protein marker (50 µL) was injected individually on the column. V_0 and V_t were determined as described previously.

The elution volume (V_e) of each protein marker was used to calculate its distribution coefficient (K_{av}):

$$K_{av} = \frac{V_e - V_0}{V_t - V_0}$$

The K_{av} value of each marker was plotted against the corresponding logarithmic M_W and a linear regression were performed to obtain the calibration curve. This curve was used to estimate the M_W of a protein sample from its elution volume. Figure 59 shows an example of calibration curve and the associated chromatograms for each column.



Diffusion Coefficient	Radius	%PD	SOS	Baseline	Mw-R
(cm²/s)	(nm)				(kDa)
2.375e-07	7.6	40.4	0.543	1.000	385.7

Peak	Radius	Mw-R	%Pd	%Intensity	%Mass	Diffusion Coefficient
	(nm)	(kDa)				(cm²/s)
☑ Peak 1	6.5	265.8	24.5	0.0	0.0	2.784e-07
☑ Peak 2	42.2	21370.3	24.3	100.0	100.0	4.270e-08

Figure 60. DLS/SLS measurements with the DynaPro Nanostar instrument. A. Principle. The sample is illuminated with a laser and the fluctuations in the intensity of the scattered light at 90° are recorded over a time period. An autocorrelation curve is obtained from the analysis of these fluctuations and used to calculate the diffusion coefficient of the particles in the sample and subsequently their hydrodynamic radius. Two methods are used to analyze the experimental autocorrelation curve. **B. Example of cumulants analysis.** The cumulant method yields the mean radius, the polydispersity (%PD) and the quality of fit (SOS) assuming that the sample is monodisperse. **C. Example of regularization analysis**. The regularization analysis does not make any assumption of the sample monodispersity and gives the size distribution of the different species in the sample that best fits the data.

6.1.2. Recovery of *Pb*MARS complexes

SEC was used to reconstitute complexes from individually purified domains. In these experiments, the Superdex 200 10/300 was used to analyze individual proteins and mixtures of two or more partners. The concentration of protein in these samples ranged from 4 to 8 µM (2000 - 4000 pmol of protein in 500 µL of SEC buffer pH 8.0). At this concentration, the UV signal is high enough to be interpreted (about 40 mAU). Proteins eluted from the column were concentrated by precipitation with TCA/acetone and analyzed by SDS-PAGE.

The binding of tRNA on the complexes was examined using similar small-scale SEC experiments. 500 µL of sample containing the protein complexes (2 - 4 µM) and different amounts of tRNA were injected on a Superdex 200 increased 10/300 or a Superose 6 increase 10/300 column under different salt ad pH conditions.

6.2. Dynamic and static light scattering

6.2.1. Theory

Dynamic light scattering (DLS) was applied to analyze the size distribution of particles in our protein or protein/tRNA samples and the size of homogeneous macromolecules in solution. In this non-invasive method the sample is irradiated with a monochromatic visible light produced by a laser (Figure 60A). The fluctuations of the intensity of the scattered light due to Brownian motion are analyzed by a correlator to generate an auto-correlation function from which a diffusion coefficient (D) is extracted. Assuming that macromolecules are spheres, a hydrodynamic radius (R_h) can be derived from the Stokes-Einstein relationship:

$$R_h = \frac{kT}{6\pi\eta D}$$

where k is the Boltzmann's constant, T the temperature in Kelvin and η is the solvent viscosity. R_h is defined as the radius of a solid sphere diffusing in the solution at the same speed as the particle of interest. A M_W can be estimated from R_h but the value is highly influenced by the shape of the molecule.

Static light scattering (SLS) derives the mass of the particles composing a homogeneous population from the relationship between the intensity of the scattered light and particle mass and concentration. For small molecules (R_h < 15 nm), it is:

$$\frac{Kc}{R(\theta)} = \frac{1}{M_W} - 2A_2c$$

where c is the particle concentration and M_W its mass (Zimm, 1948). $R(\theta)$ is the excess Rayleigh ratio at angle θ. K is an optical constant which includes the wavelength of the laser (λ), the refractive index of the solvent (n) and the increment of refractive index of the particle per concentration unit (dn/dc). A_2 is the second virial coefficient, a corrective factor for non-ideal solutions, which accounts for the strength of the interactions between particles and solvent molecules. If $A_2 > 0$, they have an affinity with the solvent and the solution is stable. If $A_2 < 0$, molecules have good affinity with themselves and tend to aggregate. In practice, samples at various concentrations are measured and A_2 and M_W are respectively obtained from the slope and the intersection with the y-axis of the Debye plot representing $Kc/R(\theta)$ *vs* concentration.

6.2.2. Instrument and calibration process

All light scattering measurements were performed on a DynaPro Nanostar instrument from Wyatt Technology (Santa Barbara, CA, USA) operating with a 658 nm wavelength laser and a scattering angle of 90°. The instrument is equipped with two detectors (a dynamic one and a proportional one) to perform simultaneously DLS and SLS measurements, respectively. All measurements were performed in batch using a single 1 µl quartz cuvette.

Prior to SLS analyzes, the cuvette was first filled with pure toluene to determine its calibration constant (here 8.387×10^{-5} $V^{-1}cm^{-1}$) at 25°C to convert the measured voltages to light intensities (performed by Dr. B. Lorber). The intensity of the solvent was subtracted from that of the sample. To do so, the offset of every buffer was measured after filtration on a 0.1 µm Millex membrane the day before the measurement to allow microscopic air bubbles to escape and any phase separation to settle. Results are summarized in Table 15.

Solvent refractive index and absolute viscosity required for DLS and SLS calculations was estimated using the version 8 calculator of the DLS software provided with the Zetasizer ZS light scattering instrument commercialized by Malvern Pananalytical. A database contains the properties of a list of chemicals. The estimated refractive indexes and viscosities of our buffers

are listed in Table 15. The increment of refractive index dn/dc of the particles used in SLS calculations was the average value 0.185 $mL.g^{-1}$ generally accepted (Zhao et al., 2011). For RNA the accepted value is 0.17 - 0.19 $mL.g^{-1}$. Consequently, we have used the same value as for the protein alone in the case of our complexes containing tRNA.

Table 15. Calibration of solvents in the quartz cuvette JC-164

Solvent	Refractive index	Viscosity (cP)	Offset measurements		
			Mean offset (V)	Standard Deviation	% of Mean
SEC pH 7.0	1.341	1.1913	0.0268862	4.77367e-05	0.177551
tRNA binding 150 mMNaCl	1.339	1.1606	0.0175128	5.70302e-05	0.32565
tRNA binding 120 mMNaCl	1.3373	1.0203	0.0152909	4.45081e-05	0.291077
tRNA binding 10% glycerol	1.347	1.3489	0.0185218	1.30715e-05	0.0705739

6.2.3. Data analysis

a) Background

In our experiments, every DLS/SLS measurement was composed of 10 acquisitions of 5 seconds that produced as many autocorrelation functions (ACFs). For every sample at least ten measurements were performed. The DYNAMICS software from Nanostar manufacturer uses two methods to extract information from the ACFs: i) Cumulants and ii) Regularization Analysis.

The Cumulants Analysis (Koppel, 1972) fits the data assuming that there is only one type of particles in the solution (the sample is said to be monomodal) and derives the averaged radius and the spread of radii (polydispersity), reported as “Radius (nm)” and “%PD” in the Datalog grid (Figure 60B). The sum of squares (SOS) assesses the difference between the measured autocorrelation curve and the cumulants-calculated curve and informs about data accuracy (i.e. how much the data depart from the fit). Low SOS values (<20) indicate reasonable agreement between experimental and fitted curves, suggesting that sample is likely monomodal with a low polydispersity. For monomodal samples, the autocorrelation curve is a smooth exponential with a maximal amplitude ranging between 1.1 and 2.0 and a baseline of 1.0.

The Regularization Analysis (Provencher, 1982) does not make assumptions on the number of size populations and estimates the radii and relative abundance of all species present in the sample (Figure 60C). The results are displayed as a Regularization Graph that is a plot of the scattering intensity distribution of each particle size. This plot does not show the abundance of the different populations because the intensity depends on particle size. The relative abundance of a population is reported as "% Mass" in the Results Summary table next to the estimated values for radius, polydispersity, M_w, and so on. Caution is required when interpreting regularization data as many different particle distributions can fit the data equally well. DYNAMICS gives a "degree of smoothness" as criterion when choosing among equivalent solutions.

b) Analysis of protein samples

Our DLS/SLS analysis were performed immediately after the SEC purification step. Fractions containing pure complexes (as assessed by SDS-PAGE and absorption spectra) were concentrated to at least 3 $mg.mL^{-1}$ using Amicon centrifugal filters with the appropriate M_w cutoff. Usually, five dilutions were prepared for light scattering measurements on the same sample at different concentrations. Measurements were performed as described in Lorber et al. (2012). First, the 20-30 µL samples were ultracentrifuged at 4°C during 1 hour at 40,000 rpm (or 99,000 x g) using a S45A rotor in a Sorvall Hitachi Discovery M150SE micro-ultracentrifuge to eliminate buffer impurities and protein aggregates formed during sample concentration. The supernatants were stored on ice until measurements. Then, 10 µL of supernatant were carefully transferred in the quartz cuvette and centrifuged 10 min at 2400 rpm (or 1500 x g) in a Sigma 1-6P bench centrifuge to removed air bubbles and dust particles. Once the sample placed in the laser beam was at rest at 20°C, the data were collected and processed using the version 7.8.1.3 of the DYNAMICS software. Sample concentration was always determined after the DLS/SLS measurements using a Thermo Fisher Scientific NanoDrop® ND 1000 spectrophotometer because ultracentrifugation caused some minor protein loss.

c) Analysis of *Pb*MARS complexes with tRNA

Protein complexes required at least 300 mM of NaCl to be homogeneous. This salt concentration and the resulting ionic strength significantly reduced tRNA binding. As a consequence, the formation of aggregates was minimized by adjusting the salt concentration in the protein/tRNA sample as late as possible. To do so, the tRNA was diluted in a low-salt buffer, volumes and concentrations were calculated to achieve a final concentrations of 150 mM NaCl, 0.8 $mg.mL^{-1}$ of protein complexes and enough tRNA to yield a complex/tRNA ratio of either 1:1, 1:2 or 1:4.

Experiments were performed using either total yeast tRNA or pure yeast $tRNA^{Phe}$ (homemade). All buffers were filtered on 0.22 µm membranes and protein samples were ultracentrifuged at 4°C (1 hour at 99,000 x g) before mixing. The binding reactions occurred during a 10 min incubation on ice before DLS/SLS measurements.

6.3. SEC-SAXS

6.3.1. Principle and requirements

Small-angle X-ray scattering (SAXS) provides low-resolution (12-30 Å) information about the size and shape of proteins in solution (Svergun et al., 2010). The sample is illuminated with a monochromatic X-ray beam and the intensity of the scattered X-rays is recorded on an area detector (Figure 61A). The scattering of the solvent is also collected and subtracted from that of the sample solution to obtain only the signal from the protein. As the molecules in the sample are randomly oriented, the scattering pattern is isotropic (i.e. scattering is the same in all directions), and can be radially averaged. Then, the scattering intensity I is represented as a function of momentum transfer $q = 4\pi sin\theta/\lambda$, where λ is the beam wavelength and 2θ is the scattering angle. Several parameters can be obtained from this 1D scattering curve including radius of gyration R_g, maximum particle size D_{max}, molecular weight M_w, particle volume V and the globularity/unfoldedness degree.

Meaningful SAXS experiments require pure samples containing single molecular species (monodisperse) without aggregates (Figure 61B). As the scattering is proportional to the square of the particle volume (V^2), the presence of small concentrations of aggregates may lead to severe alterations of the scattering curve, especially at low angles. Moreover, the measured buffer must exactly match the composition of the sample solvent. Indeed, the scattering signal of the protein is so low that even small differences between the measured buffer and the sample solvent could lead to wrong results. Size-exclusion chromatography directly coupled with SAXS (SEC-SAXS) helps to overcome these difficulties. In this set-up, the different oligomeric species in the sample are separated through a SEC column immediately before flowing through the capillary for X-ray exposure. The scattering of the elution buffer is collected before the void volume of the column and provides a well-matched solvent curve for accurate background subtraction.

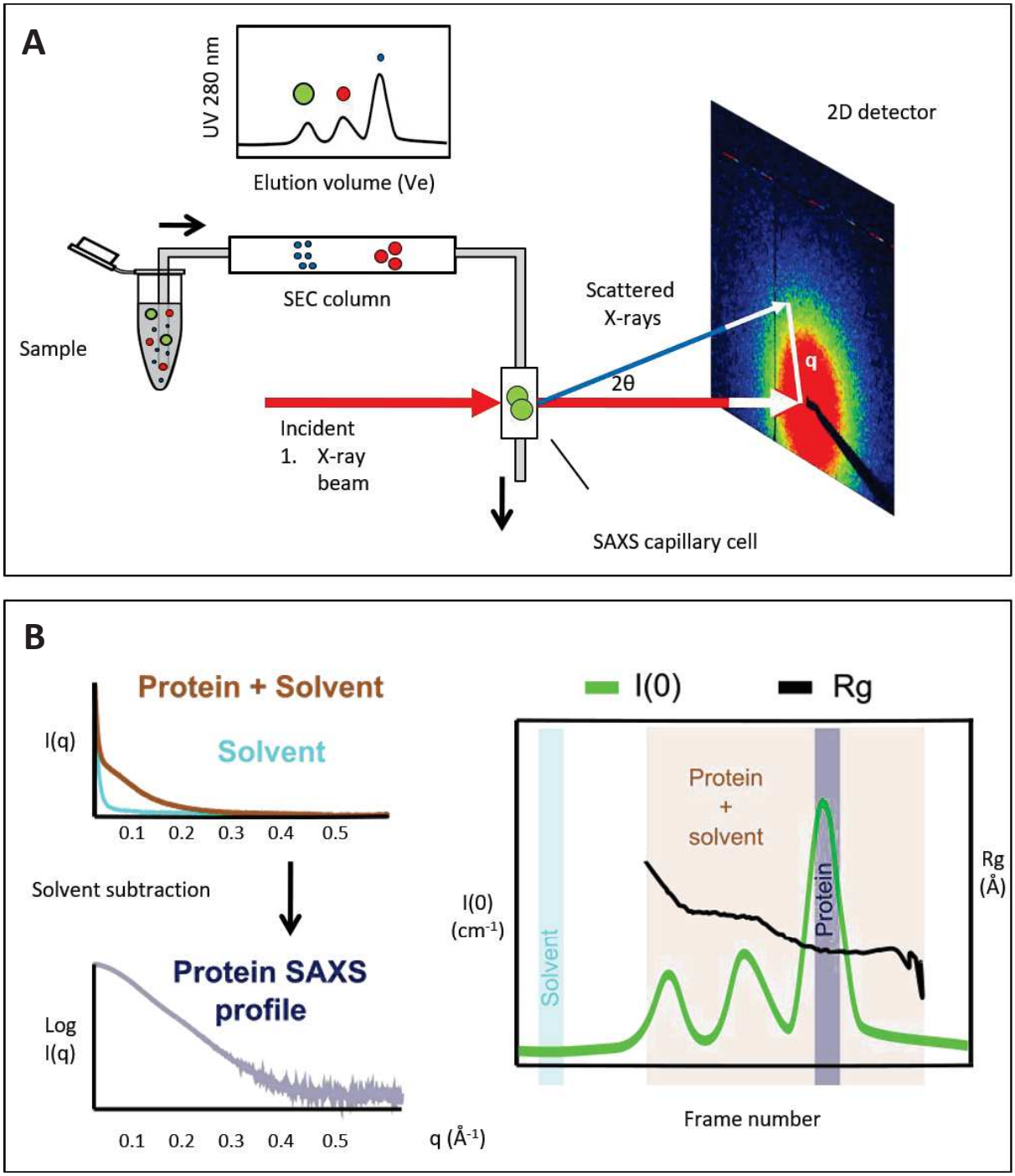


Figure 61. SEC-SAXS measurements. A. Principle of SEC-SAXS experiments. The sample is loaded onto a HPLC-SEC column to separate the protein from aggregates and higher oligomers. The eluate is injected directly in the SAXS capillary cell and the scattering patterns are collected on an area detector. **B. SAXS data processing.** The 2D images are radially averaged and the intensity at zero angle I(0) is plotted against the image number. Buffer images are selected, averaged and subtracted from all other images in order to obtain the SAXS profiles of the protein species. The R_g of each curve is obtained and plotted *vs* the image number. Consecutive equivalent images within the peak of interest are averaged to obtain the final SAXS profile.

SEC-SAXS experiments were conducted at SOLEIL synchrotron on the SWING beamline. An Agilent 1200 HPLC system equipped with a Superdex® 200 increase 3.2/300 (GE Healthcare) operating at a flow rate of 0.1 $mL.min^{-1}$ was used to separate the samples before X-ray exposure. Different types of samples were analyzed under different chromatographic conditions (Table 16).

Table 16. Details of SAXS data collection of individual proteins and complexes.

Date	Sample	Concentration $mg.mL^{-1}$ (µM)	Injection volume (µL)	Elution buffer	Exposure time (s)	Time for buffer collection (#frames)	Time for sample collection (#frames)
Protein samples							
11/2018	ERS-N-His	4.3 (145)	50	SEC 1M NaCl	1	2 min (180)	8 min (1320)
03/2019	tRip-N-His	4.7 (183)	50	SEC pH 8.0	1	3 min (180)	9.5 min (930)
06/2019	MRS-N:ERS-N:tRip-N	13.5 (170)	20	SEC pH 8.0	1	3 min (180)	8 min (1620)
06/2019	MRS-N-SUMO:ERS-N:tRip-N	9.4 (102)	20	SEC pH 8.0	1	3 min (180)	8 min (1620)
06/2019	MRS-N-SUMO:ERS-N:tRip	7.9 (69)	20	SEC pH 8.0	1	3 min (180)	8 min (1620)
06/2019	QRS-N-SUMO:ERS-N:tRip	11.4 (131)	20	SEC pH 8.0	1	3 min (180)	8 min (1620)
06/2019	tRip-N-ERS-N-His	7 (130)	20	SEC pH 8.0	1	3 min (180)	8 min (1620)
06/2019	tRip-N-ERS-N-His	11 (205)	20	SEC pH 8.0	1	3 min (180)	8 min (1620)
06/2020	MRS-N:ERS-N:tRip-N	13 (160)	70	SEC tRNA binding	2	2 min (180)	11 min (570)
06/2020	MRS-N:ERS-N:tRip	9 (90)	46	SEC tRNA binding	2	2 min (180)	11 min (570
06/2020	QRS-N:ERS-N:tRip-N	12 (160)	46	SEC tRNA binding	2	2 min (180)	11 min (570)
06/2020	QRS-N:ERS-N:tRip	8 (80)	36	SEC tRNA binding	2	2 min (180)	11 min (570)
Complexes protein:tRNA							
06/2020	MRS-N:ERS-N:tRip + tRNA (2:1):	Protein: 5.45 (54.5) tRNA: 0.7 (28) Oligo-dT: 0.2 (55)	50	SEC tRNA binding	2	2 min (180)	11 min (570)
06/2020	MRS-N:ERS-N:tRip + tRNA (4:1)	Protein: 5.45 (54.5) tRNA: 0.35 (14) Oligo-dT: 0.2 (55)	50	SEC tRNA binding	2	2 min (180)	11 min (570)
06/2020	QRS-N:ERS-N:tRip + tRNA (2:1)	Protein: 4 (40) tRNA: 0.52 (20.8) Oligo-dT: 0.14 (40.6)	50	SEC tRNA binding	2	2 min (180)	11 min (570)
06/2020	QRS-N:ERS-N:tRip + tRNA (4:1)	Protein: 4 (40) tRNA: 0.26 (10.4) Oligo-dT: 0.14 (40.6)	50	SEC tRNA binding	2	2 min (180)	11 min (570)

6.3.2. Data Processing

The scattering images were processed with Foxtrot 3.5.2-3645 (Xenocs) at the beamline. Buffer and sample frames were radially averaged using the operation "azimuthal integration" called from a macro containing a user-defined detector mask. Statistically, similar buffer frames were averaged using the operation "Clever Average HPLC Buffer". The resulting averaged buffer was then subtracted from all sample frames. Data were exported as files containing three columns of data: the momentum transfer *q* ($Å^{-1}$), the scattering intensity *I(q)* and the error *Sig(q)*.

Selection of sample frames was performed using BioXTAS RAW (Hopkins et al. 2017). Buffer-subtracted frames were loaded and plotted as Series. The intensity of each frame was plotted as function of frame number to obtain a "SAXS chromatogram". The LC Series analysis module was then used to set the buffer and calculate the R_g value of each frame, which were also plotted *vs* frame number. A monodisperse peak should display a region of flat R_g near the center. Sample frames were selected automatically. When RAW did not succeed to select a sample range, the HPLC-SAXS module of US-SOMO (Brookes et al. 2016) was used to perform some corrections in the data. Capillary fouling issues were mitigated by performing an integral baseline correction and non-baseline resolved peaks were separated using Gaussian decomposition. After these corrections, RAW normally managed to select a good sample range.

SAXS curves were analyzed with RAW and some ATSAS programs called from the same interface. Guinier analysis was performed using data within a range of $q_{max.}Rg < 1.2$. Normalized Kratky plots were obtained to assess the globularity/unfoldedness degree. The M_W was determined using different concentration-independent methods: i) the volume of correlation (Rambo & Tainer, 2013), ii) the adjusted Porod volume (Piiadov et al. 2019), iii) Bayesian inference (Hajizadeh et al. 2018) and iv) comparison to known structures (Franke et al. 2018). The pair-distance distribution *P(r)* was computed with the GNOM program (Svergun, 1992). *Ab initio* bead models were generated using DAMMIF, DAMAVER and DAMCLUST (Svergun, 1999; Franke & Svergun, 2009). The ambiguity of 3D shape reconstructions was estimated with AMBIMETER (Petoukhov & Svergun, 2015).

7. Crystallization and X-ray analysis

7.1. Choice of crystallization method

The vapour diffusion method was used to crystallize all macromolecules. In this method, a volume of protein solution is mixed with a volume of crystallizing solution and the resultant droplet is equilibrated against a larger volume of crystallizing solution. Diffusion of water (and any other volatile species) proceeds from the droplet to the reservoir until equilibration occurs. At this point, the concentration of the components of the crystallizing solution is the same in the droplet and the reservoir while the concentration of protein is increased. If the protein reaches the supersaturated states, crystallization may arise with any "luck" (Ducruix & Giegé, 1999). Two experimental set-ups of vapour diffusion, sitting and hanging drops were used in these experiments.

7.2. Crystallization screening

All crystallization trials were performed using fresh protein samples. Fractions of pure proteins eluted from the gel filtration were pooled and concentrated up to 5-10 $mg.mL^{-1}$ using an appropriate centrifugal filter (Merck Millipore Amicon®). Impurities and large protein aggregates were removed by ultracentrifugation at 99,000 g, 4°C for 1 hour (S45A rotor, Sorvall Hitachi Discovery M150SE micro-ultracentrifuge), and the protein concentration and $A_{260}/_{280}$ ratio were determined (Thermo Fisher Scientific NanoDrop® ND-1000) before crystallization assays.

Sitting drop experiments were performed in 96-well CrystalQuickX® plates (Greiner bio-one) using a Mosquito nanoliter pipetting robot (TTP Labtech). Before the experiment, the plate was blown with clean dry compressed air in order to remove any dust particle from the wells. Reservoirs were filled with 40 µL crystallization solution and the Mosquito robot dispensed 200 nL drops (100 nL of protein sample and 100 nL of reservoir solution) in the crystallization wells. The plate was carefully removed from the robot, sealed with a film (Hampton Research ClearSeal Film) and incubated at a specific temperature (4°C, 20°C, 25°C or 30°C). Drops were regularly inspected using a stereo microscope. Several commercial crystallization screens were tested: Index HT®, Crystal Screen HT®, PEG/Ion HT®, MembFac HT®, Natrix HT® (Hampton Research), JBScreen JCSG++ HTS (Jena Bioscience) as well as several customized screens prepared in 96 well plates.

Hanging drop experiments were performed in Linbro boxes. Several crystallizing agents (one per box) were tested at different pHs (rows A, B, C, D) and different concentrations (columns 1 to 6). Wells were filled with 1 mL of crystallizing solution (reservoir). To prepare the hanging drops, 1 µL of protein sample was pipetted onto a 22 mm siliconized glass coverslip (Hampton Research) was mixed with 1 µL reservoir. The coverslip was turned over, set on the greased rim and gently pressed to seal the well. The boxes were incubated at 20°C and the drops were monitored regularly. The crystallization agents tested in these experiments were MPD, PEG400, PEG3350, PEG6K, ammonium sulfate, sodium nitrate and lithium chloride.

Table 17. Crystallization screen tested for different samples of proteins and complexes.

Sample	Screens										
	Index	**Crystal Screen**	**PEG /Ion**	**Mem bFac**	**Natrix**	**JCSG++**	**Linbro boxes**	**AS0**	**AS1**	**AS2**	**AS 3**
tRip-N	X					X					
ERS-N	X					X	X	X	X		X
ERS-N-His		X		X		X	X	X	X		X
QRS-N-SUMO-His	X					X					
QRS-N_{208}:ERS-N									X		
tRip-His + ERS-N-SUMO-His									X		
tRip-His + ERS-N-His									X		
QRS-N_{208}:ERS-N:tRip-N	X		X			X					
MRS-N:ERS-N:tRip-N	X		X	X		X					
MRS-N-His: tRip-N-ERS-N						X					
MRS-N:ERS-N:tRip-N_{180}	X	X	X		X	X	X	X	X		
QRS-N:ERS-N:tRip-N_{180}		X			X		X	X	X		
MRS-N:ERS-N:tRip-N_{170}		X				X				X	
QRS-N:ERS-N:tRip-N_{170}		X				X				X	
MRS-N:ERS-N:tRip	X	X	X	X		X	X				
QRS-N:ERS-N:tRip	X	X	X			X	X				
MRS-N-His:ERS-N:tRip + tRNA (4:1)				X							
ERS-N:tRip-N:QRS-N-SUMO:MRS-N-SUMO	X					X					

7.3. Optimization of *Pb*ERS-N crystals

Initially, crystallization screens were performed with 10 mg.ml^{-1} of ERS-N in SEC buffer pH 8.0. Spherulites were obtained in 2 M ammonium sulfate, 100 mM Bis-Tris pH 5.5, (condition G11 of JBScreen JCSG++). Then, the customized crystallization screen "AS-0" was tested (Annex 1). It allowed the screening of different pH (4.5 – 8.5) and different PEGs (1 - 10% of low- and high-MW PEGs) and yielded small crystals in 2 M ammonium sulfate, 100 mM HEPES-NaOH pH 7.5 and 1% (v/v) of low-MW PEGs. With the customized crystallization screen "AS-1" (Annex 1) different concentrations of ammonium sulfate were assayed (0.9 - 2 M) resulting in crystals of about 50 µm in 1.5 M ammonium sulfate, 100 mM HEPES-NaOH pH 7.5 and 0.5%(v/v) PEG at 25°C or 30°C. These crystals were cryoprotected with paraffin oil, frozen in liquid nitrogen and subjected to X-ray analysis at the synchrotron source (beamline PXIII of Swiss Light Source and beamline Proxima-2A of SOLEIL). Protein diffraction was observed but only at low-resolution (about 5 Å).

Larger crystals were grown using seeding techniques. A seed stock was prepared using crystals from AS-1 screens (columns 6 and 7). They were transferred to a seed bead tube (Hampton Research) with 200 µL of 1.5 M ammonium sulfate, 0.5% PEG400 and 100 mM HEPES-NaOH pH 7.5 (condition A7 in AS-1) and crushed by vortexing for 3 min. Seeds were used to grow new crystals in the customized AS-3 screen (Annex 1) which has a narrow concentration range of ammonium sulfate (1.3 – 1.7 M) and includes conditions with glycerol to facilitate cryoprotection. To perform the experiment, 3 - 5 µL of seeds were mixed with 55 µL of protein solution (ERS-N purified in SEC buffer pH 7.5 and concentrated to 6–9 mg.ml^{-1}) and was used immediately to dispense 800 nL drops (400 nL protein/seeds + 400 nL reservoir) with the Mosquito robot. Crystals of about 100 µm formed in 1.4-1.5 M ammonium sulfate, 100 mM HEPES-NaOH pH 7.5 and 0.5%(v/v) PEG (400, 1500 or 8K) or 10-20% glycerol after 2 - 3 weeks at 25°C.

7.4. X-ray data collection of *Pb*ERS-N crystals

Crystals were soaked in different cryo-protectant solutions (lithium sulfate, Xylitol, Paratone-N, Glucose, Proline) prior to freezing in liquid nitrogen. However, best diffraction was obtained with crystals directly grown in 10 or 20% glycerol. A native data set was collected at a λ = 0.9786 Å using an EIGER-X 16M detector (Dectris Ltd.) at the PROXIMA-1 beamline (SOLEIL synchrotron, France). 360° data were collected with an oscillation range of 0.1°. Data were processed using XDS *via* the command-line interface xdsme (Legrand, 2017). The crystal diffracted to 2.7 Å and belonged to the space group C121 (a = 129.8 Å, b= 88.6 Å, c=169.1 Å, α= 90°, β= 106°, γ=90°).

As molecular replacement on the native data set using homology models based on known crystal structures of other aaRS-related GST domains (Simader, Hothorn, Köhler et al., 2006; Simader, Hothorn & Suck, 2006; Kim K.J. et al., 2008; Cho et al., 2015) was unsuccessful, crystals derivatized with the crystallophore Tb-Xo4™ (Engilberge et al. 2017) were prepared in order to perform experimental phasing. A tube containing 0.6 mg Tb-Xo4 (M_W = 556 Da) (Polyvalan) was suspended in 10 µL of solution 1.6 M ammonium sulfate, 100 mM HEPES-NaOH pH 7.5, 20% glycerol to a final concentration of 100 mM. 1 µL of this solution was added to the 400 nL drop containing the crystals and soaking was performed during 1 - 2 minutes. Crystals were mounted in cryo-loops and frozen in liquid nitrogen. Single-wavelength Anomalous Diffraction (SAD) experiments were performed at the terbium L_{III} absorption edge by setting the incident beam energy to 7.6 keV (λ = 1.6314 Å). Several 720° data sets were collected with an oscillation range of 0.1° and were merged to increase the signal-to-noise ratio of the anomalous signal

7.5. Structure determination of ERS-N

The X-ray diffraction data have been processed and reduced with the XDS package. The Tb derivative dataset has been scaled and merged from 3 different crystals within the same package (XSCALE). The resulting statistics for both the native and the Tb derivative datasets are reported in the table 18.

Not shown in these table, the $< I/\sigma(I) >$ is around 12 and the R-merge has a value of 30% in the 4.2 - 4.0 Å resolution shell for the Tb derivative. The values for these two indicators are 5.8 and 32% respectively for the native dataset but in the 3.31-3.06 Å shell. The conclusion is that the datasets are statistically useful up to 4.0 Å for the Tb derivative and 3.1 Å for the native using old refinement programs that doesn't model the errors. More recent statistical indicators such R-pim or the Pearson's statistics correlation coefficient (CC1/2 and CC*) are more accurate to extract signal from noise and hence to determine the maximum resolution for modern phasing or refinement programs and justify our choice to use reflections up 3.09 Å for the derivative and 2.7 Å for the native dataset.

Although our data do not extend that much between 3.5 Å and higher resolution and hence the Wilson statistic might be somewhat inaccurate, this analysis indicates that coordinates in the final model might have very high B factors.

Table 18. Data collection statistics for the TbXo4 derivative and the native datasets

Data set code	Tb derivative ERS8bcd [a]	Native ERS2-1
Wavelength	1.63137	0.978565
Resolution range	46.98 - 3.093	44.34 - 2.703
Highest-resolution shell	3.170 - 3.093	2.799 - 2.703
Space group	C2	C2
a,b, c (Å)	130.11, 88.51, 168.64	129.98, 88.68, 169.28
β(°)	105.70	106.13
Total reflections	1363717 (61763)	356985 (33970)
Unique reflections	66007 (4790)	50722 (4911)
Multiplicity	20.7 (12.9)	7.0 (6.9)
Completeness (%)	99.68 (96.68)	99.69 (97.73)
Mean I/σ(I)	14.38 (0.7)	14.81 (1.47)
Wilson B-factor	136.90	82.29
R-merge	0.1328 (2.418)	0.07188 (1.181)
R-meas	0.1345 (2.464)	0.07767 (1.277)
R-pim	0.0210 (0.4627)	0.0292 (0.4809)
CC1/2	1.000 (0.515)	0.999 (0.759)
CC*	1.000 (0.912)	1.000 (0.929)
SigAno	1.200 (0.598)	
N_ano	32050 (2279)	

[a] : statistics reported assume Friedel's law is not satisfied.

$$R_{merge} = \frac{\sum_{hkl}\sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl}\sum_{i=1}^{n} I_i(hkl)}$$

$$R_{meas} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl}\sum_{i=1}^{n} I_i(hkl)}$$

$$R_{pim} = \frac{\sum_{hkl}\sqrt{1/n-1}\sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl}\sum_{i=1}^{n} I_i(hkl)}$$

$$r = \frac{\sum_{i=1}^{n}\left((x_i - \bar{x})(y_i - \bar{y})\right)}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

CC1/2 corresponds to the r value when splitting randomly the dataset in two equal parts, x and y.

$$CC^* = \sqrt{\frac{2\,CC_{1/2}}{1 + CC_{1/2}}}$$

a) Anomalous signal

The previous statistics indicate that the ERS8bcd dataset is complete and shows a high multiplicity which means that errors on the anomalous differences (Dano/Sigdano) will be statistically reduced due to this high multiplicity. A further analysis of the < I/σ(I) > and Dano/Sigdano *versus* resolution shows that the reflections are also strong and that there is a significant anomalous contribution both in favour of a potential solution structure by SAD on the Tb (Figure 62 and 63, respectively).

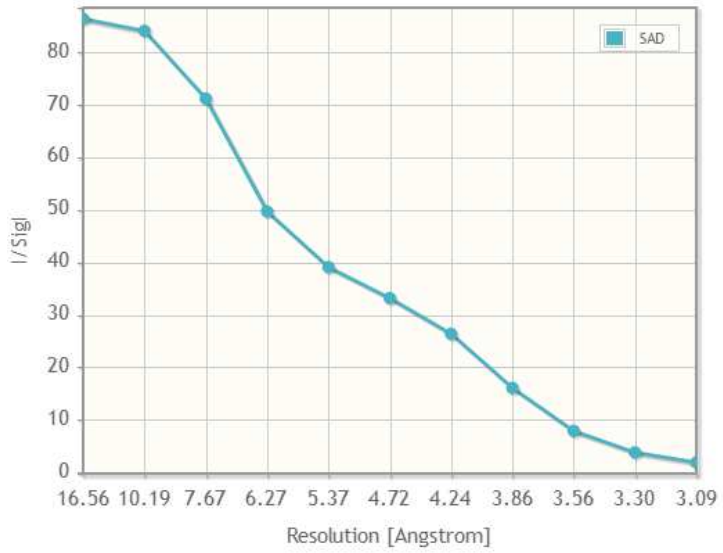


Figure 62. Analysis of intensities *vs* resolution for the Tb derivative dataset

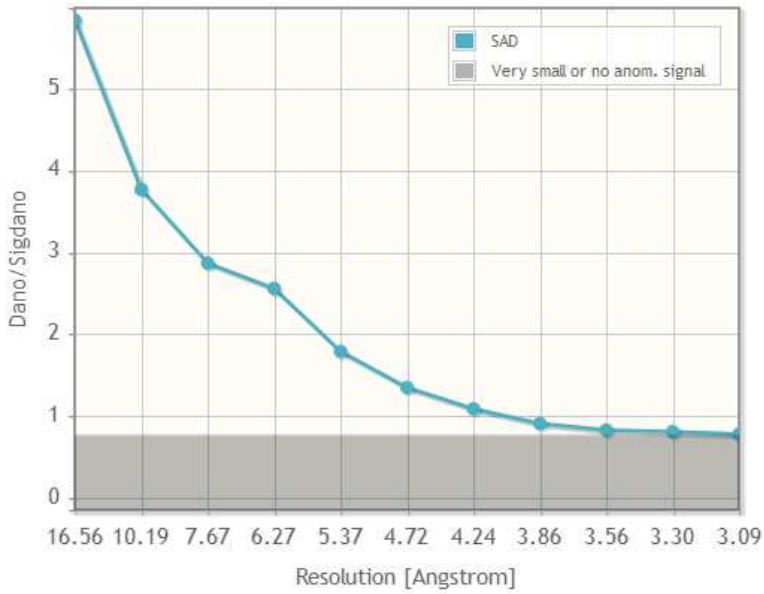


Figure 63. Analysis of the anomalous signal *vs* resolution for the Tb derivative dataset

b) Self-rotation function

The examination of the output (Figure 64) produced by *molrep* from *CCP4 package* for the self-rotation function of the Tb derivative dataset indicates two non-crystallographic twofold axis (peaks for χ = 180°) and a fivefold non crystallographic axis (or at least pseudo-axis). The presence of 5-fold axis or pseudo-axis led us to start the SAD phasing on the basis of 5 protein monomers per asymmetric unit with the *ShelX suite* (*shelxC, shelxD*). 5 monomers per asymmetric unit correspond to a solvent content of 61.2% considering the volume of the present unit cell. This observation of a high solvent content correlates well with the high B factor resulting from the Wilson's analysis.

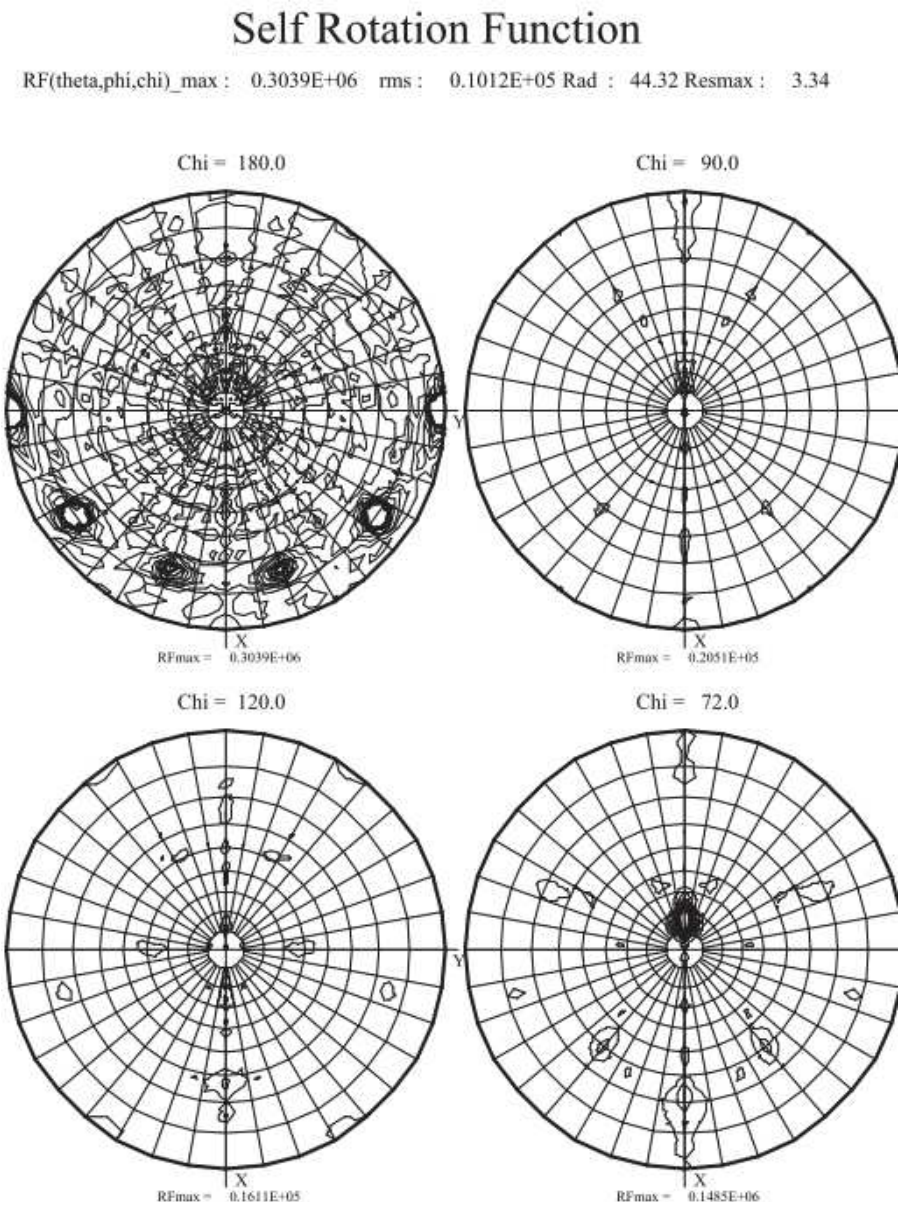


Figure 64: Self-rotation function plots for various χ values and in particular 180° and 72°

c) Experimental phasing

The *crank2* pipeline of the *CCP4 package* has been used for experimental phasing. Briefly, *shelxC* and *shelxD* were used for the identification of anomalous sites estimation and detection. Using a high-resolution cut-off of 4.17 Å, 28 anomalous sites have been found by this procedure, 21 of which having an occupancy equal or higher than 0.25 (maximum Combined FOM of 0.52 and CC around 0.35) (Figure 65). The substructure has been refined using *REFMAC5* and gave a mean FOM of 0.1866.

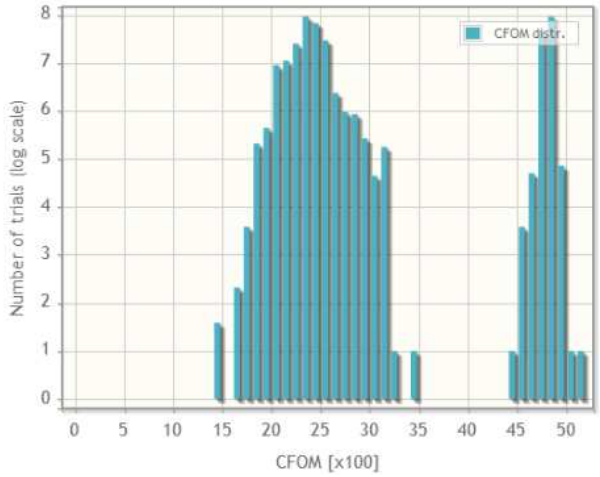


Figure 65. Histogram of number of trials *vs* Combined Figure Of Merit for both hands

A clear hand solution shows up (Figure 66). The correct hand was selected by the combination of MAPRO, Solomon for the Density Modification (DM) at this step, Multicomb and REFMAC5 for phase combinations in the hand selection process. The combination of these programs leads to a clear hand solution (combined DM FOM and phasing CLD score of 0.0 for hand #1 and 16.22 for hand #2).

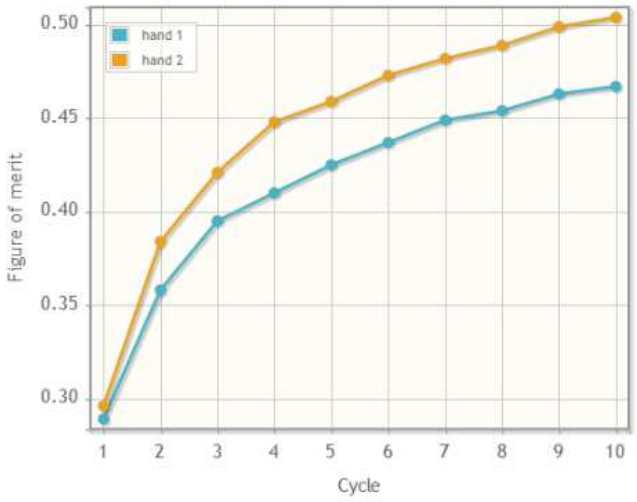


Figure 66: Figure of merit for each hand vs Density Modification cycles indicating a clear hand selection

Starting with the phases from hand #2 a further phase refinement process by density modification was conducted using *Parrot* (Figure 67) (which allows search and use of Non-crystallographic Symmetries) in combination with *REFMAC5*.

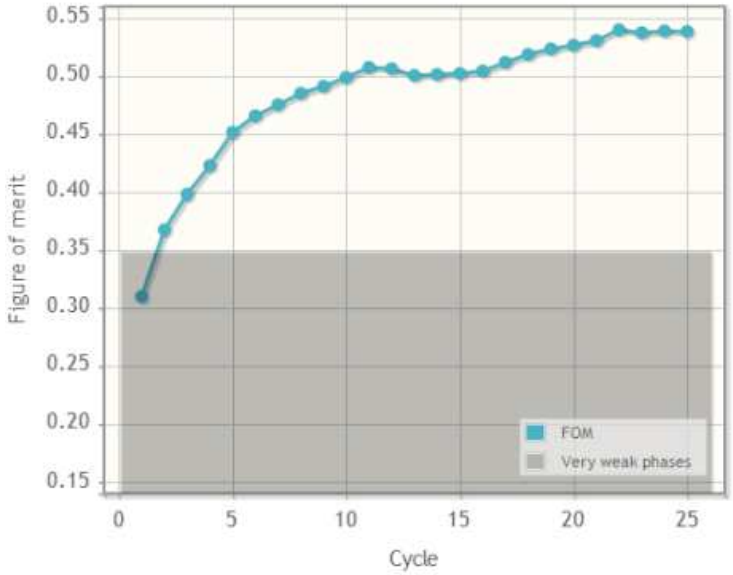


Figure 67: *Parrot* Density Modification cycles for hand #2

Finally, a last step made used of the combination of Parrot for density modification, REFMAC5 for coordinates refinement and Buccaneer for automated model building. This step led to a model built from 862 amino acids (over a total of 1240) in 37 fragments (i.e. around 7 per monomer). The corresponding built coordinate file has a R factor of 0.3677 and a R-free factor of 0.4054 (Figure 68).

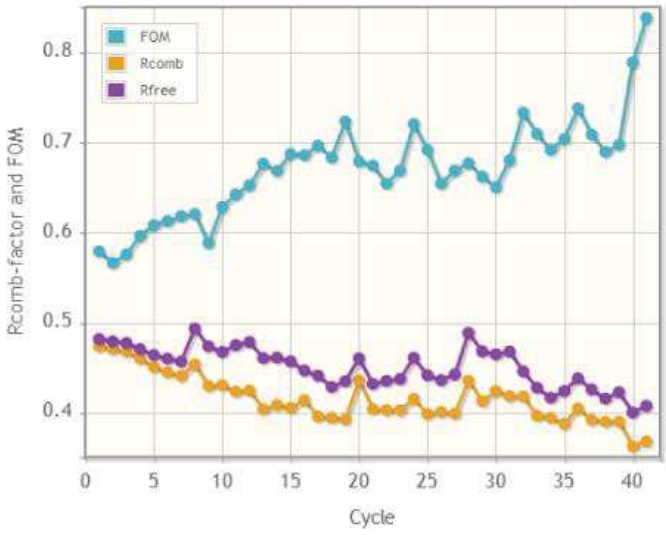


Figure 68. R factors and FOM improvements during the automated model building process

At this stage manual corrections and model building took place followed by refinement in the Phenix package. The statistics of the current best model (R factor of 20.8% and Free-R factor of 25%) are given in the Table 4 in Results. As expected from the X-ray data collection and the Wilson statistics, coordinates have very high thermal agitation B factors. Although B factors are correlated with resolution, the current model has higher B factors than the average PDB structures (mean B higher than 110 $Å^2$). This is not the final model yet and this is the reason why very few ligand molecules (ions, water molecules) have been added to the model at this stage and several Ramachandran outliers are still present in the model. Indeed, our current goal is to improve locally the electronic density, in particular in regions were loops linking secondary structure elements are still missing, while reducing as much as possible any bias due to misplaced atoms.

8. Electrophoretic mobility shift assay (EMSA)

8.1. Non-radioactive EMSA in agarose gels

As the size of complexes was not compatible with polyacrylamide gels, an electrophoretic mobility shift assay (EMSA) in agarose gel had to be optimized to study their interactions with tRNA. In these experiments, the protein complex (3.3 µM) was incubated with different concentrations of total yeast tRNA (0 - 6.7 µM) in 25 mM HEPES-NaOH pH 7.0, 150 mM NaCl, 10% glycerol, 5 mM $MgCl_2$, 0.005% DDM and 5 mM 2-ME, containing 6.7 µM oligo-dT (*5'-TTTTTTTTTTTT-3'*) to avoid non-specific binding. After 20 min of incubation in ice, samples were analyzed by electrophoresis on a low-melting agarose gel (Quantum Biotechnologies, #AGAL0050) 1% (w/v) in TBE buffer at 75 V for 1h30 at 4°C. The gel was first stained with ethidium bromide to reveal tRNA molecules, then with InstantBlue®Coomasie Protein Stain (Expedeon Ltd.) to expose the proteins. The protein content of each band was analyzed by SDS-PAGE. Thus, protein bands were excised from the agarose gel, transferred into 1.5 mL tubes, and incubated 10 min at 95°C to melt the agarose. A volume of SDS-PAGE loading buffer (2X) was added. As samples became yellow, the pH had to be adjusted by the addition of a few drops of concentrated NaOH until the solution became blue again. Finally, samples were heated again 10 min at 95°C and 40µL were loaded on a 12% (w/v) acrylamide/bisacrylamide (19:1) 1.5 mm thick SDS gel.

Binding assays were always prepared by adding the three ingredients in a specific order. First, one volume of tRNA/oligo-dT mixture (diluted in water) then, one volume of EMSA buffer (50 mM HEPES-NaOH pH 7.0, 150 mM NaCl, 25% glycerol, 15 mM $MgCl_2$, 0.01% DDM and 10 mM 2-ME) and one volume of the protein complex (in buffer SEC pH 7.0 containing 300 mM NaCl) was added last. This procedure lowers the salt concentration to 150 mM in the protein sample only at the very last step. Most experiments were performed in a final volume of 15 µL containing 50 pmol of protein complexes, 100 pmol oligo-dT, and variable amounts (100, 50, 25, 12.5 pmol) of total yeast tRNA.

8.2. Polyacrylamide affinity co-electrophoresis

Increasing concentrations of peptides (62.5 nM to 1 µM) were embedded in a 1.5 x 80 x 100 mm^3 polyacrylamide gel with 10 stripes of 200 µL (Figure 69). The gel framework, the stripes containing the different concentrations of peptides and the wells are made of 6% (19:1) polyacrylamide gel in buffer TBE. Cascade dilutions of peptides and gel solutions were prepared to obtain 250 µl of gel 6% (19:1) in TBE 1X containing the desired concentration of each peptide and 3 times the amount of polyT to neutralize unspecific interactions. 5 µL samples (about 5 nM [3'-^{32}P] RNA in 25 mM HEPES-NaOH pH 7.0, 5 mM $MgCl_2$, 10% glycerol and bromophenol blue) of radiolabeled tRNA transcripts (elongator and initiator Met_{CAU}, Gln_{UUG}, and Gln_{CUG}) or total yeast tRNA were electrophoresed for 90 min 70 V at 4°C. The gels were dried and analyzed on a PhosphorImager (Typhoon FLA 7000).

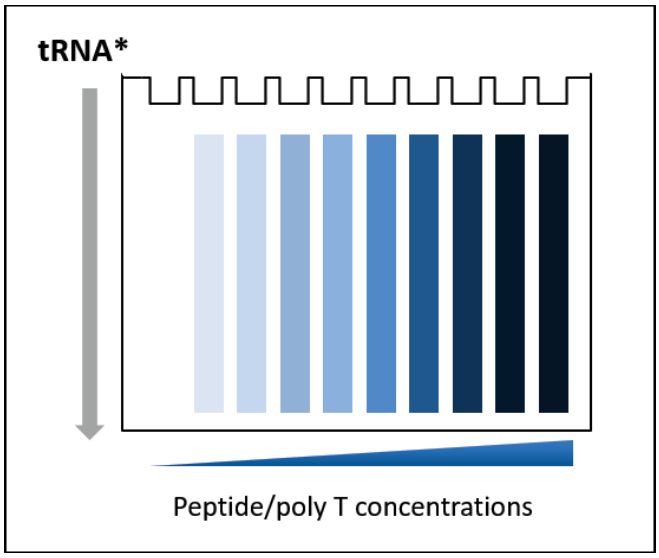


Figure 69. Schematic representation of a polyacrylamide affinity co-electrophoresis gel. Each lane contains different concentrations of the protein to be tested. 3 times more polyT was added as a competitor.

Annex 1. Customized crystallization screens "AS"

Screen "AS0"

			1	2	3	4	5	6	7	8	9	10	11	12			
		pH	**4.5**	**4.5**	**4.5**	**6**	**6**	**6**	**7.5**	**7.5**	**7.5**	**8.5**	**8.5**	**8.5**			
		Buffer	**NaOAc**			**Citrate Na**			**HEPES-NaOH**			**Tris-HCl**					
		Conc. Stock (M)	2			2			1.12			1.9					
		Vol to add (uL)	50			50			90			50					
																Conc stock (M)	**Vol to add (uL)**
	Conc stock		**1%**	**5%**	**10%**	**1%**	**5%**	**10%**	**1%**	**5%**	**10%**	**1%**	**5%**	**10%**			
MPD	35 %	**A**	**29**	**143**	**286**	**29**	**143**	**286**	**29**	**143**	**286**	**29**	**143**	**286**			
		H2O	350	236	93	350	236	93	310	196	53	350	236	93			
PEG400	100 %	**B**	**10**	**50**	**100**	**10**	**50**	**100**	**10**	**50**	**100**	**10**	**50**	**100**			
		H2O	369	329	279	369	329	279	329	289	239	369	329	279			
PEG3350	40 %	**C**	**25**	**125**	**250**	**25**	**125**	**250**	**25**	**125**	**250**	**25**	**125**	**250**	**AS 2 M**	3.5	571
		H2O	354	254	129	354	254	129	314	214	89	354	254	129			
PEG6K	40 %	**D**	**25**	**125**	**250**	**25**	**125**	**250**	**25**	**125**	**250**	**25**	**125**	**250**			
		H2O	354	254	129	354	254	129	314	214	89	354	254	129			
MPD	35 %	**E**	**29**	**143**	**286**	**29**	**143**	**286**	**29**	**143**	**286**	**29**	**143**	**286**			
		H2O	635	521	379	635	521	379	595	481	339	635	521	379			
PEG400	100 %	**F**	**10**	**50**	**100**	**10**	**50**	**100**	**10**	**50**	**100**	**10**	**50**	**100**			
		H2O	654	614	564	654	614	564	614	574	524	654	614	564			
PEG3350	40 %	**G**	**25**	**125**	**250**	**25**	**125**	**250**	**25**	**125**	**250**	**25**	**125**	**250**	**AS 1 M**	3.5	286
		H2O	639	539	414	639	539	414	599	499	374	639	539	414			
PEG6K	40 %	**H**	**25**	**125**	**250**	**25**	**125**	**250**	**25**	**125**	**250**	**25**	**125**	**250**			
		H2O	639	539	414	639	539	414	599	499	374	639	539	414			

						Concentration buffer												
						0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	M
						Concentration de AS												
						0.9	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2	M
Additif	Concentration				Stock	1	2	3	4	5	6	7	8	9	10	11	12	
PEG400	0.5		A	PEG400	10 %	100	100	100	100	100	100	100	100	100	100	100	100	
				HEPES-NaOH pH 7,5	1 M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5 M	514	571	629	686	743	800	857	914	971	1029	1086	1143	
				H2O		1186	1129	1071	1014	957	900	843	786	729	671	614	557	
PEG1500	0.5	%	B	PEG1500	10 %	100	100	100	100	100	100	100	100	100	100	100	100	
				HEPES-NaOH pH 7,5	1 M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5 M	514	571	629	686	743	800	857	914	971	1029	1086	1143	
				H2O		1186	1129	1071	1014	957	900	843	786	729	671	614	557	
PEG3350	0.5	%	C	PEG3350	10 %	100	100	100	100	100	100	100	100	100	100	100	100	
				HEPES-NaOH pH 7,5	1 M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5 M	514	571	629	686	743	800	857	914	971	1029	1086	1143	
				H2O		1186	1129	1071	1014	957	900	843	786	729	671	614	557	
PEG4K	0.5	%	D	PEG4K	10 %	100	100	100	100	100	100	100	100	100	100	100	100	
				HEPES-NaOH pH 7,5	1 M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5 M	514	571	629	686	743	800	857	914	971	1029	1086	1143	
				H2O		1186	1129	1071	1014	957	900	843	786	729	671	614	557	
PEG6K	0.5	%	E	PEG6K	10 %	100	100	100	100	100	100	100	100	100	100	100	100	
				HEPES-NaOH pH 7,5	1 M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5 M	514	571	629	686	743	800	857	914	971	1029	1086	1143	
				H2O		1186	1129	1071	1014	957	900	843	786	729	671	614	557	
PEG8K	0.5	%	F	PEG8K	10 %	100	100	100	100	100	100	100	100	100	100	100	100	
				HEPES-NaOH pH 7,5	1 M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5 M	514	571	629	686	743	800	857	914	971	1029	1086	1143	
				H2O		1186	1129	1071	1014	957	900	843	786	729	671	614	557	
PEG20K	0.5	%	G	PEG20K	10 %	100	100	100	100	100	100	100	100	100	100	100	100	
				HEPES-NaOH pH 7,5	1 M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5 M	514	571	629	686	743	800	857	914	971	1029	1086	1143	
				H2O		1186	1129	1071	1014	957	900	843	786	729	671	614	557	
Glycerol	10	%	H	Glycerol	50 %	400	400	400	400	400	400	400	400	400	400	400	400	
				HEPES-NaOH pH 7,5	1 M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5 M	514	571	629	686	743	800	857	914	971	1029	1086	1143	
				H2O		886	829	771	714	657	600	543	486	429	371	314	257	

PEGs	1.2 mL
HEPES	19.2 mL
(NH4)2SO4	79.5 mL
H2O	80.1 mL
	192

Conc Stock		
PEG400	100	%
PEG1500	50	%
PEG3350	40	%
PEG4K	50	%
PEG6K	40	%
PEG8K	40	%
PEG20K	30	%
Glycerol	50	%

Solutions 10 mL à 10%

Vol stock (ml	Vol H2O (mL	Vol final (mL)	Conc final (%)
1.00	9.00	10	10
2.00	8.00	10	10
2.50	7.50	10	10
2.00	8.00	10	10
2.50	7.50	10	10
2.50	7.50	10	10
3.33	6.67	10	10
2.00	8.00	10	10

							Concentration buffer												
							0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	M
							Concentration de AS												
							1.3	1.35	1.4	1.425	1.45	1.475	1.5	1.525	1.55	1.575	1.6	1.7	M
Additif	Concentration				Stock		1	2	3	4	5	6	7	8	9	10	11	12	
PEG400	0.5	%	A	PEG400	10	%	100	100	100	100	100	100	100	100	100	100	100	100	
				HEPES-NaOH pH 7,5	1	M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5	M	743	771	800	814	829	843	857	871	886	900	914	971	
				H2O			957	929	900	886	871	857	843	829	814	800	786	729	
PEG1500	0.5	%	B	PEG1500	10	%	100	100	100	100	100	100	100	100	100	100	100	100	
				HEPES-NaOH pH 7,5	1	M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5	M	743	771	800	814	829	843	857	871	886	900	914	971	
				H2O			957	929	900	886	871	857	843	829	814	800	786	729	
PEG8K	0.5	%	C	PEG8K	10	%	100	100	100	100	100	100	100	100	100	100	100	100	
				HEPES-NaOH pH 7,5	1	M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5	M	743	771	800	814	829	843	857	871	886	900	914	971	
				H2O			957	929	900	886	871	857	843	829	814	800	786	729	
Glycerol	10	%	D	Glycerol	50	%	400	400	400	400	400	400	400	400	400	400	400	400	
				HEPES-NaOH pH 7,5	1	M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5	M	743	771	800	814	829	843	857	871	886	900	914	971	
				H2O			657	629	600	586	571	557	543	529	514	500	486	429	
Glycerol	20	%	E	Glycerol	50	%	800	800	800	800	800	800	800	800	800	800	800	800	
				HEPES-NaOH pH 7,5	1	M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5	M	743	771	800	814	829	843	857	871	886	900	914	971	
				H2O			257	229	200	186	171	157	143	129	114	100	86	29	
Glycerol	30	%	F	Glycerol	100	%	600	600	600	600	600	600	600	600	600	600	600	600	
				HEPES-NaOH pH 7,5	1	M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5	M	743	771	800	814	829	843	857	871	886	900	914	971	
				H2O			457	429	400	386	371	357	343	329	314	300	286	229	
NiSO4	10	mM	G	NiSO4	200	mM	100	100	100	100	100	100	100	100	100	100	100	100	
				HEPES-NaOH pH 7,5	1	M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5	M	743	771	800	814	829	843	857	871	886	900	914	971	
				H2O			957	929	900	886	871	857	843	829	814	800	786	729	
none		%	H	none	50	%	0	0	0	0	0	0	0	0	0	0	0	0	
				HEPES-NaOH pH 7,5	1	M	200	200	200	200	200	200	200	200	200	200	200	200	
				(NH4)2SO4	3.5	M	743	771	800	814	829	843	857	871	886	900	914	971	
				H2O			1057	1029	1000	986	971	957	943	929	914	900	886	829	

PEGs	1.2 mL
HEPES	19.2 mL
(NH4)2SO4	81.6 mL
H2O	64.8 mL
	192

	Conc Stock	
PEG400	100	%
PEG1500	50	%
PEG3350	40	%
PEG4K	50	%
PEG6K	40	%
PEG8K	40	%
PEG20K	30	%
Glycerol	50	%

Solutions 10 mL à 10%

Vol stock (m	Vol H2O (mL)	Vol final (mL)	Conc final (%)
1.00	9.00	10	10
2.00	8.00	10	10
2.50	7.50	10	10
2.00	8.00	10	10
2.50	7.50	10	10
2.50	7.50	10	10
3.33	6.67	10	10
2.00	8.00	10	10

REFERENCES

A

Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. (2004). Image processing with imageJ. *Biophotonics International*, *11*(7), 36–41. https://doi.org/10.1201/9781420005615.ax4

Aceto, A., Dragani, B., Melino, S., Allocati, N., Masulli, M., Di Ilio, C., & Petruzzelli, R. (1997). Identification of an N-capping box that affects the α6-helix propensity in glutathione S-transferase superfamily proteins: A role for an invariant aspartic residue. *Biochemical Journal*, *322*(1), 229–234. https://doi.org/10.1042/bj3220229

Adio, S., Sharma, H., Senyushkina, T., Karki, P., Maracci, C., Wohlgemuth, I., Holtkamp, W., Peske, F., & Rodnina, M. V. (2018). Dynamics of ribosomes and release factors during translation termination in E. Coli. *ELife*, *7*, e34252. https://doi.org/10.7554/eLife.34252

Afonso, A., Neto, Z., Castro, H., Lopes, D., Alves, A. C., Tomás, A. M., & Rosário, V. D. (2010). Plasmodium chabaudi chabaudi malaria parasites can develop stable resistance to atovaquone with a mutation in the cytochrome b gene. *Malaria Journal*, *9*, 135. https://doi.org/10.1186/1475-2875-9-135

Ahn, H. C., Kim, S., & Lee, B. J. (2003). Solution structure and p43 binding of the p38 leucine zipper motif: Coiled-coil interactions mediate the association between p38 and p43. *FEBS Letters*, *542*(1–3), 119–124. https://doi.org/10.1016/S0014-5793(03)00362-4

AhYoung, A. P., Koehl, A., Cascio, D., & Egea, P. F. (2015). Structural mapping of the ClpB ATPases of Plasmodium falciparum: Targeting protein folding and secretion for antimalarial drug design. *Protein Science*, *24*(9), 1508–1520. https://doi.org/10.1002/pro.2739

Alexander, R. W., & Schimmel, P. (2001). Domain-domain communication in aminoacyl-tRNA synthetases. *Progress in Nucleic Acid Research and Molecular Biology*, *69*, 317–349. https://doi.org/10.1016/s0079-6603(01)69050-0

Aly, A. S. I., Vaughan, A. M., & Kappe, S. H. I. (2009). Malaria Parasite Development in the Mosquito and Infection of the Mammalian Host. *Annual Review of Microbiology*, *63*, 195–221. https://doi.org/10.1146/annurev.micro.091208.073403.Malaria

Andersen, G. R., Nissen, P., & Nyborg, J. (2003). Elongation factors in protein biosynthesis. *Trends in Biochemical Sciences*, *28*(8), 434–441. https://doi.org/10.1016/S0968-0004(03)00162-2

Aravind, L., Iyer, L. M., Wellems, T. E., & Miller, L. H. (2003). Plasmodium Biology: Genomic Gleanings. *Cell*, *115*(7), 771–785. https://doi.org/10.1016/S0092-8674(03)01023-7

Argenzio, E., & Moolenaar, W. H. (2016). Emerging biological roles of Cl- intracellular channel proteins. *Journal of Cell Science*, *129*(22), 4165–4174. https://doi.org/10.1242/jcs.189795

Arif, A., Jia, J., Moodt, R. A., DiCorleto, P. E., & Fox, P. L. (2011). Phosphorylation of glutamyl-prolyl tRNA synthetase by cyclin-dependent kinase 5 dictates transcriptselective translational control. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1415–1420. https://doi.org/10.1073/pnas.1011275108

Arthur, L. L., Pavlovic-Djuranovic, S., Koutmou, K. S., Green, R., Szczesny, P., & Djuranovic, S. (2015). Translational control by lysine-encoding A-rich sequences. *Science Advances*, *1*(6), e1500154. https://doi.org/10.1126/sciadv.1500154

B

Baragaña, B., Hallyburton, I., Lee, M. C. S., Norcross, N. R., Grimaldi, R., Otto, T. D., Proto, W. R., Blagborough, A. M., Meister, S., Wirjanata, G., Ruecker, A., Upton, L. M., Abraham, T. S., Almeida, M. J., Pradhan, A., Porzelle, A., Martínez, M. S., Bolscher, J. M., Woodland, A., … Gilbert, I. H. (2015). A novel multiple-stage antimalarial agent that inhibits protein synthesis. *Nature*, *522*(7556), 315–320. https://doi.org/10.1038/nature14451

Bastien, O., Lespinats, S., Roy, S., Métayer, K., Fertil, B., Codani, J. J., & Maréchal, E. (2004). Analysis of the compositional biases in Plasmodium falciparum genome and proteome using Arabidopsis thaliana as a reference. *Gene*, *336*(2), 163–173. https://doi.org/10.1016/j.gene.2004.04.029

Baum, J., Papenfuss, A. T., Mair, G. R., Janse, C. J., Vlachou, D., Waters, A. P., Cowman, A. F., Crabb, B. S., & De Koning-Ward, T. F. (2009). Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. *Nucleic Acids Research*, *37*(11), 3788–3798. https://doi.org/10.1093/nar/gkp239

Bayer-Santos, E., Marini, M. M., & da Silveira, J. F. (2017). Non-coding RNAs in host-pathogen interactions: Subversion of mammalian cell functions by protozoan parasites. *Frontiers in Microbiology*, *8*, 474. https://doi.org/10.3389/fmicb.2017.00474

Bec, G., Kerjan, P., & Waller, J. P. (1994). Reconstitution in vitro of the valyl-tRNA synthetase-elongation factor (EF) 1βγδ complex. Essential roles of the NH2-terminal extension of valyl-tRNA synthetase and of the EF-1δ subunit in complex formation. *Journal of Biological Chemistry*, *269*(3), 2086–2092.

Bec, G., Kerjan, P., Zha, X. D., & Waller, J.-P. (1989). Valyl-tRNA Synthetase from Rabbit Liver. *Journal of Biological Chemistry*, *264*(35), 21138–21143.

Becker, T., Franckenberg, S., Wickles, S., Shoemaker, C. J., Anger, A. M., Armache, J. P., Sieber, H., Ungewickell, C., Berninghausen, O., Daberkow, I., Karcher, A., Thomm, M., Hopfner, K. P., Green, R., & Beckmann, R. (2012). Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature*, *482*(7386), 501–506. https://doi.org/10.1038/nature10829

Belachew, E. B. (2018). Immune Response and Evasion Mechanisms of Plasmodium falciparum Parasites. *Journal of Immunology Research*, *2018*(6529681). https://doi.org/10.1155/2018/6529681

Beri, D., Balan, B., & Tatu, U. (2018). Commit, hide and escape: The story of Plasmodium gametocytes. *Parasitology*, *145*(13), 1772–1782. https://doi.org/10.1017/S0031182018000926

Bernabeu, M., Danziger, S. A., Avril, M., Vaz, M., Babar, P. H., Brazier, A. J., Herricks, T., Maki, J. N., Pereira, L., Mascarenhas, A., Gomes, E., Chery, L., Aitchison, J. D., Rathod, P. K., & Smith, J. D. (2016). Severe adult malaria is associated with specific PfEMP1 adhesion types and high parasite biomass. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(23), E3270–E3279. https://doi.org/10.1073/pnas.1524294113

Bhatt, T. K., Kapil, C., Khan, S., Jairajpuri, M. A., Sharma, V., Santoni, D., Silvestrini, F., Pizzi, E., & Sharma, A. (2009). A genomic glimpse of aminoacyl-tRNA synthetases in malaria parasite Plasmodium falciparum. *BMC Genomics*, *10*, 644. https://doi.org/10.1186/1471-2164-10-644

Bhatt, T. K., Sharma, A., Khan, S., Dwivedi, V. P., Banday, M. M., Sharma, A., Chandele, A., Camacho, N., De Pouplana, L. S. R., Wu, Y., Craig, A. G., Mikkonen, A. T., Maier, A. G., & Yogavel, M. (2011). Malaria parasite tyrosyl-tRNA synthetase secretion triggers pro-inflammatory responses. *Nature Communications*, *2*, 530. https://doi.org/10.1038/ncomms1522

Biswas, S., Lim, E. E., Gupta, A., Saqib, U., Mir, S. S., Siddiqi, M. I., Ralph, S. A., & Habib, S. (2011). Interaction of apicoplast-encoded elongation factor (EF) EF-Tu with nuclear-encoded EF-Ts mediates translation in the Plasmodium falciparum plastid. *International Journal for Parasitology*, *41*(3), 417–427. https://doi.org/10.1016/j.ijpara.2010.11.003

Boddey, J. A., & Cowman, A. F. (2013). Plasmodium nesting: Remaking the erythrocyte from the inside out. *Annual Review of Microbiology*, *67*, 243–269. https://doi.org/10.1146/annurev-micro-092412-155730

Boddey, J. A., Moritz, R. L., Simpson, R. J., & Cowman, A. F. (2009). Role of the Plasmodium export element in trafficking parasite proteins to the infected erythrocyte. *Traffic*, *10*(3), 285–299. https://doi.org/10.1111/j.1600-0854.2008.00864.x

Boucher, M. J., Ghosh, S., Zhang, L., Lal, A., Jang, S. W., Ju, A., Zhang, S., Wang, X., Ralph, S. A., Zou, J., Elias, J. E., & Yeh, E. (2018). Integrative proteomics and bioinformatic prediction enable a high-confidence apicoplast proteome in malaria parasites. *PLoS Biology*, *16*(9), e2005895. https://doi.org/10.1371/journal.pbio.2005895

Bour, T., Akaddar, A., Lorber, B., Blais, S., Balg, C., Candolfi, E., & Frugier, M. (2009). Plasmodial aspartyl-tRNA synthetases and peculiarities in plasmodium falciparum. *Journal of Biological Chemistry*, *284*(28), 18893–18903. https://doi.org/10.1074/jbc.M109.015297

Bour, T., Mahmoudi, N., Kapps, D., Thiberge, S., Bargieri, D., Ménard, R., & Frugier, M. (2016). Apicomplexa-specific tRip facilitates import of exogenous tRNAs into malaria parasites. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(17), 4717–4722. https://doi.org/10.1073/pnas.1600476113

Brick, P., Bhat, T. N., & Blow, D. M. (1989). Structure of tyrosyl-tRNA synthetase refined at 2.3 Å resolution. Interaction of the enzyme with the tyrosyl adenylate intermediate. *Journal of Molecular Biology*, *208*(1), 83–98. https://doi.org/10.1016/0022-2836(89)90090-9

Brookes, E., Vachette, P., Rocco, M., & Pérez, J. (2016). US-SOMO HPLC-SAXS module: Dealing with capillary fouling and extraction of pure component patterns from poorly resolved SEC-SAXS data. *Journal of Applied Crystallography*, *49*(5), 1827–1841. https://doi.org/10.1107/S1600576716011201

Brown, J. R., & Doolittle, W. F. (1999). Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. *Journal of Molecular Evolution*, *49*(4), 485–495. https://doi.org/10.1007/PL00006571

Buckland, R., & Wild, F. (1989). Leucine zipper motif extends. *Nature*, *338*(6216), 547.

Bullen, H. E., Charnaud, S. C., Kalanon, M., Riglar, D. T., Dekiwadia, C., Kangwanrangsan, N., Torii, M., Tsuboi, T., Baum, J., Ralph, S. A., Cowman, A. F., De Koning-Ward, T. F., Crabb, B. S., & Gilsona, P. R. (2012). Biosynthesis, localization, and macromolecular arrangement of the Plasmodium falciparum translocon of exported proteins (PTEX). *Journal of Biological Chemistry*, *287*(11), 7871–7884. https://doi.org/10.1074/jbc.M111.328591

C

Cahuzac, B., Berthonneau, E., Birlirakis, N., Guittet, E., & Mirande, M. (2000). A recurrent RNA-binding domain is appended to eukaryotic aminoacyl-tRNA synthetases. *The EMBO Journal*, *19*(3), 445–452.

Carlton, J. M., Adams, J. H., Silva, J. C., Bidwell, S. L., Lorenzi, H., Caler, E., Crabtree, J., Angiuoli, S. V., Merino, E. F., Amedeo, P., Cheng, Q., Coulson, R. M. R., Crabb, B. S., Del Portillo, H. A., Essien, K., Feldblyum, T. V., Fernandez-Becerra, C., Gilson, P. R., Gueye, A. H., … Fraser-Liggett, C. M. (2008). Comparative genomics of the neglected human malaria parasite Plasmodium vivax. *Nature*, *455*(7214), 757–763. https://doi.org/10.1038/nature07327

Casares, S., & Richie, T. L. (2009). Immune evasion by malaria parasites: a challenge for vaccine development. *Current Opinion in Immunology*, *21*(3), 321–330. https://doi.org/10.1016/j.coi.2009.05.015

Castro de Moura, M., Miro, F., Han, J. M., Kim, S., Celada, A., & Ribas de Pouplana, L. (2011). Entamoeba lysyl-tRNA synthetase contains a cytokine-like domain with chemokine activity towards human endothelial cells. *PLoS Neglected Tropical Diseases*, *5*(11), e1398. https://doi.org/10.1371/journal.pntd.0001398

Cavarelli, J., Delagoutte, B., Eriani, G., Gangloff, J., & Moras, D. (1998). L-arginine recognition by yeast arginyl-tRNA synthetase. *EMBO Journal*, *17*(18), 5438–5448. https://doi.org/10.1093/emboj/17.18.5438

Cavarelli, J., Eriani, G., Rees, B., Ruff, M., Boeglin, M., Mitschler, A., Martin, F., Gangloff, J., Thierry, J. C., & Moras, D. (1994). The active site of yeast aspartyl-tRNA synthetase: Structural and functional aspects of the aminoacylation reaction. *EMBO Journal*, *13*(2), 327–337. https://doi.org/10.1002/j.1460-2075.1994.tb06265.x

Cela, M., Paulus, C., Santos, M. A. S., Moura, G. R., Frugier, M., & Rudinger-Thirion, J. (2018). Plasmodium apicoplast tyrosyl-tRNA synthetase recognizes an unusual, simplified identity set in cognate tRNA Tyr. *PLoS ONE*, *13*(12), e0209805. https://doi.org/10.1371/journal.pone.0209805

Cen, S., Javanbakht, H., Niu, M., & Kleiman, L. (2004). Ability of Wild-Type and Mutant Lysyl-tRNA Synthetase To Facilitate tRNALys Incorporation into Human Immunodeficiency Virus Type 1. *Journal of Virology*, *78*(3), 1595–1601. https://doi.org/10.1128/jvi.78.3.1595-1601.2004

Cerini, C., Kerjan, P., Astler, M., Gratecos, D., Mirande, M., & Semeriva, M. (1991). A component of the multisynthetase complex is a multifunctional aminoacyl-tRNA synthetase. *EMBO Journal*, *10*(13), 4267–4277. https://doi.org/10.1002/j.1460-2075.1991.tb05005.x

Cestari, I., Kalidas, S., Monnerat, S., Anupama, A., Phillips, M. A., & Stuart, K. (2013). A Multiple Aminoacyl-tRNA Synthetase Complex That Enhances tRNA-Aminoacylation in African Trypanosomes. *Molecular and Cellular Biology*, *33*(24), 4872–4888. https://doi.org/10.1128/mcb.00711-13

Cha, S. J., Kim, M. S., Pandey, A., & Jacobs-Lorena, M. (2016). Identification of GAP DH on the surface of Plasmodium sporozoites as a new candidate for targeting malaria liver invasion. *Journal of Experimental Medicine*, *213*(10), 2099–2112. https://doi.org/10.1084/jem.20160059

Chaliotis, A., Vlastaridis, P., Mossialos, D., Ibba, M., Becker, H. D., Stathopoulos, C., & Amoutzias, G. D. (2017). The complex evolutionary history of aminoacyl-tRNA synthetases. *Nucleic Acids Research*, *45*(3), 1059–1068. https://doi.org/10.1093/nar/gkw1182

Chan, J. A., Fowkes, F. J. I., & Beeson, J. G. (2014). Surface antigens of Plasmodium falciparum-infected erythrocytes as immune targets and malaria vaccine candidates. *Cellular and Molecular Life Sciences : CMLS*, *71*, 3633–3657. https://doi.org/10.1007/s00018-014-1614-3

Chatterjee, K., Nostramo, R. T., Wan, Y., & Hopper, A. K. (2018). tRNA dynamics between the nucleus, cytoplasm and mitochondrial surface: Location, location, location. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1861*(4), 373–386. https://doi.org/10.1016/j.bbagrm.2017.11.007

Chaubey, S., Kumar, A., Singh, D., & Habib, S. (2005). The apicoplast of Plasmodium falciparum is translationally active. *Molecular Microbiology*, *56*(1), 81–89. https://doi.org/10.1111/j.1365-2958.2005.04538.x

Chisholm, S. A., Kalanon, M., Nebl, T., Sanders, P. R., Matthews, K. M., Dickerman, B. K., Gilson, P. R., & de Koning-Ward, T. F. (2018). The malaria PTEX component PTEX88 interacts most closely with HSP101 at the host–parasite interface. *FEBS Journal*, *285*(11), 2037–2055. https://doi.org/10.1111/febs.14463

Cho, H. Y., Lee, H. J., Choi, Y. S., Kim, D. K., Jin, K. S., Kim, S., & Kang, B. S. (2019). Symmetric Assembly of a Decameric Subcomplex in Human Multi-tRNA Synthetase Complex Via Interactions between Glutathione Transferase-Homology Domains and Aspartyl-tRNA Synthetase. *Journal of Molecular Biology*, *431*(22), 4475–4496. https://doi.org/10.1016/j.jmb.2019.08.013

Cho, H. Y., Maeng, S. J., Cho, H. J., Choi, Y. S., Chung, J. M., Lee, S., Kim, H. K., Kim, J. H., Eom, C. Y., Kim, Y. G., Guo, M., Jung, H. S., Kang, B. S., & Kim, S. (2015). Assembly of multi-tRNA synthetase complex via heterotetrameric glutathione transferase-homology domains. *Journal of Biological Chemistry*, *290*(49), 29313–29328. https://doi.org/10.1074/jbc.M115.690867

Coelho, C. H., Doritchamou, J. Y. A., Zaidi, I., & Duffy, P. E. (2017). Advances in malaria vaccine development: Report from the 2017 malaria vaccine symposium. *Npj Vaccines*, *2*(34). https://doi.org/10.1038/s41541-017-0035-3

Cooke, B. M., Lingelbach, K., Bannister, L. H., & Tilley, L. (2004). Protein trafficking in Plasmodium falciparum-infected red blood cells. *Trends in Parasitology*, *20*(12), 581–589. https://doi.org/10.1016/j.pt.2004.09.008

Coppi, A., Tewari, R., Bishop, J. R., Bennett, B. L., Lawrence, R., Esko, J. D., Billker, O., & Sinnis, P. (2007). Heparan Sulfate Proteoglycans Provide a Signal to Plasmodium Sporozoites to Stop Migrating and Productively Invade Cells. *Cell Host Microbe*, *2*(5), 316–327.

Coulson, R. M. R., Hall, N., & Ouzounis, C. A. (2004). Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum. *Genome Research*, *14*(8), 1548–1554. https://doi.org/10.1101/gr.2218604

Cowell, A. N., & Winzeler, E. A. (2019). The genomic architecture of antimalarial drug resistance. *Briefings in Functional Genomics*, *18*(5), 314–328. https://doi.org/10.1093/bfgp/elz008

Cowman, A. F., Healer, J., Marapana, D., & Marsh, K. (2016). Malaria: Biology and Disease. *Cell*, *167*(3), 610–624. https://doi.org/10.1016/j.cell.2016.07.055

Cowman, A. F., Tonkin, C. J., Tham, W. H., & Duraisingh, M. T. (2017). The Molecular Basis of Erythrocyte Invasion by Malaria Parasites. *Cell Host and Microbe*, *22*(2), 232–245. https://doi.org/10.1016/j.chom.2017.07.003

Crepin, T., Peterson, F., Haertlein, M., Jensen, D., Wang, C., Cusack, S., & Kron, M. (2011). A hybrid structural model of the complete Brugia malayi cytoplasmic asparaginyl-tRNA synthetase. *Journal of Molecular Biology*, *405*(4), 1056–1069. https://doi.org/10.1016/j.jmb.2010.11.049

Crepin, T., Schmitt, E., Blanquet, S., & Mechulam, Y. (2002). Structure and function of the C-terminal domain of methionyl-tRNA synthetase. *Biochemistry*, *41*(43), 13003–13011. https://doi.org/10.1021/bi026343m

Crepin, T., Schmitt, E., Blanquet, S., & Mechulam, Y. (2004). Three-Dimensional Structure of Methionyl-tRNA Synthetase from Pyrococcus abyssi. *Biochemistry*, *43*(9), 2635–2644. https://doi.org/10.1021/bi0356247

Crepin, T., Schmitt, E., Mechulam, Y., Sampson, P. B., Vaughan, M. D., Honek, J. F., & Blanquet, S. (2003). Use of analogues of methionine and methionyl adenylate to sample conformational changes during catalysis in Escherichia coli methionyl-tRNA synthetase. *Journal of Molecular Biology*, *332*(1), 59–72. https://doi.org/10.1016/S0022-2836(03)00917-3

Cromer, D., Evans, K. J., Schofield, L., & Davenport, M. P. (2006). Preferential invasion of reticulocytes during late-stage Plasmodium berghei infection accounts for reduced circulating reticulocyte levels. *International Journal for Parasitology*, *36*(13), 1389–1397. https://doi.org/10.1016/j.ijpara.2006.07.009

Cukras, A. R., & Green, R. (2005). Multiple effects of S13 in modulating the strength of intersubunit interactions in the ribosome during translation. *Journal of Molecular Biology*, *349*(1), 47–59. https://doi.org/10.1016/j.jmb.2005.03.075

Cusack, S. (1993). Sequence, structure and evolutionary relationships between class 2 aminoacyl-tRNA synthetases: An update. *Biochimie*, *75*(12), 1077–1081. https://doi.org/10.1016/0300-9084(93)90006-E

Cusack, S. (1995). Eleven down and nine to go. *Nature Structural Biology*, *2*(10), 824–831. https://doi.org/10.1038/nsb1095-824

Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N., & Leberman, R. (1990). A second class of synthetase structure revealed by X-ray analysis of Escherichia coli seryl-tRNA synthetase at 2.5 Å. *Nature*, *347*(6290), 249–255. https://doi.org/10.1038/347249a0

D

Dalby, A. R. (2009). A comparative proteomic analysis of the simple amino acid repeat distributions in Plasmodia reveals lineage specific amino acid selection. *PLoS ONE*, *4*(7), e6231. https://doi.org/10.1371/journal.pone.0006231

Dandewad, V., Vindu, A., Joseph, J., & Seshadri, V. (2019). Import of human miRNA-RISC complex into Plasmodium falciparum and regulation of the parasite gene expression. *Journal of Biosciences*, *44*(2), 50. https://doi.org/10.1007/s12038-019-9870-x

Davies, H. M., Nofal, S. D., McLaughlin, E. J., & Osborne, A. R. (2017). Repetitive sequences in malaria parasite proteins. *FEMS Microbiology Reviews*, *41*(6), 923–940. https://doi.org/10.1093/femsre/fux046

De Koning-Ward, T. F., Dixon, M. W. A., Tilley, L., & Gilson, P. R. (2016). Plasmodium species: Master renovators of their host cells. *Nature Reviews Microbiology*, *14*(8), 494–507. https://doi.org/10.1038/nrmicro.2016.79

De Koning-Ward, T. F., Gilson, P. R., Boddey, J. A., Rug, M., Smith, B. J., Papenfuss, A. T., Sanders, P. R., Lundie, R. J., Maier, A. G., Cowman, A. F., & Crabb, B. S. (2009). A newly discovered protein export machine in malaria parasites. *Nature*, *459*(7249), 945–949. https://doi.org/10.1038/nature08104

De Niz, M., & Heussler, V. T. (2018). Rodent malaria models: insights into human disease and parasite biology. *Current Opinion in Microbiology*, *46*, 93–101. https://doi.org/10.1016/j.mib.2018.09.003

Deinert, K., Fasiolo, F., Hurt, E. C., & Simos, G. (2001). Arc1p Organizes the Yeast Aminoacyl-tRNA Synthetase Complex and Stabilizes Its Interaction with the Cognate tRNAs. *Journal of Biological Chemistry*, *276*(8), 6000–6008. https://doi.org/10.1074/jbc.M008682200

Delarue, M., & Moras, D. (1993). The Aminoacyl-tRNA Synthethase Family: Molecules at Work. *BioEssays*, *15*(9), 675–687.

Deniziak, M., & Barciszewski, J. (2001). Methionyl-tRNA synthetase. *Acta Biochimica Polonica*, *48*(2), 337–350.

Deniziak, M., Sauter, C., Becker, H. D., Paulus, C. A., Giegé, R., & Kern, D. (2007). Deinococcus glutaminyl-tRNA synthetase is a chimer between proteins from an ancient and the modern pathways of aminoacyl-tRNA formation. *Nucleic Acids Research*, *35*(5), 1421–1431. https://doi.org/10.1093/nar/gkl1164

Deponte, M., Hoppe, H. C., Lee, M. C. S., Maier, A. G., Richard, D., Rug, M., Spielmann, T., & Przyborski, J. M. (2012). Wherever i may roam: Protein and membrane trafficking in P. falciparum-infected red blood cells. *Molecular and Biochemical Parasitology*, *186*(2), 95–116. https://doi.org/10.1016/j.molbiopara.2012.09.007

DePristo, M. A., Zilversmit, M. M., & Hartl, D. L. (2006). On the abundance, amino acid composition, and evolutionary dynamics of low-complexity regions in proteins. *Gene*, *378*, 19–30. https://doi.org/10.1016/j.gene.2006.03.023

Desai, M., Hill, J., Fernandes, S., Walker, P., Pell, C., Gutman, J., Kayentao, K., Gonzalez, R., Webster, J., Greenwood, B., Cot, M., & ter Kuile, F. O. (2018). Prevention of malaria in pregnancy. *The Lancet Infectious Diseases*, *18*(4), e119–e132. https://doi.org/10.1016/S1473-3099(18)30064-1

Dhangadamajhi, G., Kar, S. K., & Ranjit, M. (2010). The Survival Strategies of Malaria Parasite in the Red Blood Cell and Host Cell Polymorphisms. *Malaria Research and Treatment*, *2010*, 973094. https://doi.org/10.4061/2010/973094

Dias, J., Renault, L., Pérez, J., & Mirande, M. (2013). Small-angle X-ray solution scattering study of the multi-aminoacyl-tRNA synthetase complex reveals an elongated and multi-armed particle. *Journal of Biological Chemistry*, *288*(33), 23979–23989. https://doi.org/10.1074/jbc.M113.489922

Dirr, H., Reinemer, P., & Huber, R. (1994). X-ray crystal structures of cytosolic glutathione S-transferases: Implications for protein architecture, substrate recognition and catalytic function. *European Journal of Biochemistry*, *220*(3), 645–661. https://doi.org/10.1111/j.1432-1033.1994.tb18666.x

Djuranovic, S. P., Erath, J., Andrews, R., Bayguinov, P., Chung, J., Chalker, D., Fitzpatrick, J. A., Moss, W., Szczesny, P., & Djuranovic, S. (2018). PolyA tracks and poly-lysine repeats are the Achilles heel of Plasmodium falciparum. *BioRxiv*, 420109. https://doi.org/10.1101/420109

Djuranovic, S. P., Erath, J., Andrews, R. J., Bayguinov, P. O., Chung, J. J., Chalker, D. L., Fitzpatrick, J. A. J., Moss, W. N., Szczesny, P., & Djuranovic, S. (2020). Plasmodium falciparum translational machinery condones polyadenosine repeats. *ELife*, *9*, e57799. https://doi.org/10.7554/eLife.57799

Ducruix, A., & Giegé, R. (1999). *Crystallization of Nucleic Acids and Proteins: A Practical Approach* (2nd Ed). Oxford University Press.

Duffy, P. E., & Gorres, J. P. (2020). Malaria vaccines since 2000: progress, priorities, products. *Npj Vaccines*, *5*(48). https://doi.org/10.1038/s41541-020-0196-3

Duret, L. (2000). tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. *Trends in Genetics*, *16*(7), 287–289. https://doi.org/10.1016/S0168-9525(00)02041-2

Dvorak, J. A., Miller, L. H., Whitehouse, W. C., & Shiroishi, T. (1975). Invasion of erythrocytes by malaria merozoites. *Science*, *187*(4178), 748–750. https://doi.org/10.1126/science.803712

E

Engilberge, S., Riobé, F., Di Pietro, S., Lassalle, L., Coquelle, N., Arnaud, C. A., Pitrat, D., Mulatier, J. C., Madern, D., Breyton, C., Maury, O., & Girard, E. (2017). Crystallophore: A versatile lanthanide complex for protein crystallography combining nucleating effects, phasing properties, and luminescence. *Chemical Science*, *8*(9), 5909–5917. https://doi.org/10.1039/c7sc00758b

Erath, J., Djuranovic, S., & Djuranovic, S. P. (2019). Adaptation of Translational Machinery in Malaria Parasites to Accommodate Translation of Poly-Adenosine Stretches Throughout Its Life Cycle. *Frontiers in Microbiology*, *10*(2823). https://doi.org/10.3389/fmicb.2019.02823

Eriani, G., Delarue, M., Poch, O., Gangloff, J., & Moras, D. (1990). Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature*, *347*(6289), 203–206.

Erickson, H. P. (2009). Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biological Procedures Online*, *11*(1), 32–51. https://doi.org/10.1007/s12575-009-9008-x

Esseiva, A. C., Naguleswaran, A., Hemphill, A., & Schneider, A. (2004). Mitochondrial tRNA import in Toxoplasma gondii. *Journal of Biological Chemistry*, *279*(41), 42363–42368. https://doi.org/10.1074/jbc.M404519200

F

Feagin, J. E., Harrell, M. I., Lee, J. C., Coe, K. J., Sands, B. H., Cannone, J. J., Tami, G., Schnare, M. N., & Gutell, R. R. (2012). The fragmented mitochondrial ribosomal RNAs of plasmodium falciparum. *PLoS ONE*, *7*(6), e38320. https://doi.org/10.1371/journal.pone.0038320

Feagin, J. E., Mericle, B. L., Werner, E., & Morris, M. (1997). Identification of additional rRNA fragments encoded by the Plasmodium falciparum 6 kb element. *Nucleic Acids Research*, *25*(2), 438–446. https://doi.org/10.1093/nar/25.2.438

Filisetti, D., Théobald-Dietrich, A., Mahmoudi, N., Rudinger-Thirion, J., Candolfi, E., & Frugier, M. (2013). Aminoacylation of plasmodium falciparum tRNA-Asn and insights in the synthesis of asparagine repeats. *Journal of Biological Chemistry*, *288*(51), 36361–36371. https://doi.org/10.1074/jbc.M113.522896

Finarov, I., Moor, N., Kessler, N., Klipcan, L., & Safro, M. G. (2010). Structure of Human Cytosolic Phenylalanyl-tRNA Synthetase: Evidence for Kingdom-Specific Design of the Active Sites and tRNA Binding Patterns. *Structure*, *18*(3), 343–353. https://doi.org/10.1016/j.str.2010.01.002

Fourmy, D., Mechulam, Y., & Blanquet, S. (1995). Crucial Role of an Idiosyncratic Insertion in the Rossman Fold of Class 1 Aminoacyl-tRNA Synthetases: The Case of Methionyl-tRNA Synthetase. *Biochemistry*, *34*(48), 15681–15688. https://doi.org/10.1021/bi00048a012

Fourmy, D., Meinnel, T., Mechulam, Y., & Blanquet, S. (1993). Mapping of the zinc binding domain of Escherichia coli methionyl-tRNA synthetase. *Journal of Molecular Biology*, *231*(4), 1068–1077.

Francin, M., Kaminska, M., Kerjan, P., & Mirande, M. (2002). The N-terminal domain of mammalian Lysyl-tRNA synthetase is a functional tRNA-binding domain. *Journal of Biological Chemistry*, *277*(3), 1762–1769. https://doi.org/10.1074/jbc.M109759200

Francin, M., & Mirande, M. (2003). Functional dissection of the eukaryotic-specific tRNA-interacting factor of lysyl-tRNA synthetase. *Journal of Biological Chemistry*, *278*(3), 1472–1479. https://doi.org/10.1074/jbc.M208802200

Franke, D., Jeffries, C. M., & Svergun, D. I. (2018). Machine Learning Methods for X-Ray Scattering Data Analysis from Biomacromolecular Solutions. *Biophysical Journal*, *114*(11), 2485–2492. https://doi.org/10.1016/j.bpj.2018.04.018

Franke, D., & Svergun, D. I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography*, *42*(2), 342–346. https://doi.org/10.1107/S0021889809000338

Frechin, M., Enkler, L., Tetaud, E., Laporte, D., Senger, B., Blancard, C., Hammann, P., Bader, G., Clauder-Münster, S., Steinmetz, L. M., Martin, R. P., DiRago, J. P., & Becker, H. D. (2014). Expression of Nuclear and Mitochondrial Genes Encoding ATP Synthase Is Synchronized by Disassembly of a Multisynthetase Complex. *Molecular Cell*, *56*(6), 763–776. https://doi.org/10.1016/j.molcel.2014.10.015

Frechin, M., Senger, B., Braye, M., Kern, D., Martin, R. P., & Becker, H. D. (2009). Yeast mitochondrial Gln-tRNAGlnis generated by a GatFAB-mediated transamidation pathway involving Arc1p-controlled subcellular sorting of cytosolic GluRS. *Genes and Development*, *23*(9), 1119–1130. https://doi.org/10.1101/gad.518109

Frimpong, A., Kusi, K. A., Ofori, M. F., & Ndifon, W. (2018). Novel strategies for malaria vaccine design. *Frontiers in Immunology*, *9*(2769). https://doi.org/10.3389/fimmu.2018.02769

Frischknecht, F., & Matuschewski, K. (2017). Plasmodium sporozoite biology. *Cold Spring Harbor Perspectives in Medicine*, *7*, a025478. https://doi.org/10.1101/cshperspect.a025478

Fritz-Wolf, K., Becker, A., Rahlfs, S., Harwaldt, P., Schirmer, R. H., Kabsch, W., & Becker, K. (2003). X-ray structure of glutathione S-transferase from the malarial parasite Plasmodium falciparum. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(24), 13821–13826. https://doi.org/10.1073/pnas.2333763100

Frugier, M., Bour, T., Ayach, M., Santos, M. A. S., Rudinger-Thirion, J., Théobald-Dietrich, A., & Pizzi, E. (2010). Low Complexity Regions behave as tRNA sponges to help co-translational folding of plasmodial proteins. *FEBS Letters*, *584*(2), 448–454. https://doi.org/10.1016/j.febslet.2009.11.004

Frugier, M., Moulinier, L., & Giegé, R. (2000). A domain in the N-terminal extension of class IIb eukaryotic aminoacyl-tRNA synthetases is important for tRNA binding. *The EMBO Journal*, *19*(10), 2371–2380. https://doi.org/10.1093/emboj/19.10.2371

Frugier, M., Ryckelynck, M., & Giegé, R. (2005). tRNA-balanced expression of a eukaryal aminoacyl-tRNA synthetase by an mRNA-mediated pathway. *EMBO Reports*, *6*(9), 860–865. https://doi.org/10.1038/sj.embor.7400481

Fu, Y., Kim, Y., Jin, K. S., Kim, H. S., Kim, J. H., Wang, D. M., Park, M., Jo, C. H., Kwon, N. H., Kim, D., Kim, M. H., Jeon, Y. H., Hwang, K. Y., Kim, S., & Cho, Y. (2014). Structure of the ArgRS-GlnRS-AIMP1 complex and its implications for mammalian translation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(42), 15084–15089. https://doi.org/10.1073/pnas.1408836111

Fukai, S., Nureki, O., Sekine, S. ichi, Shimada, A., Tao, J., Vassylyev, D. G., & Yokoyama, S. (2000). Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of tRNA(Val) and valyl-tRNA synthetase. *Cell*, *103*(5), 793–803. https://doi.org/10.1016/S0092-8674(00)00182-3

Fukui, H., Hanaoka, R., & Kawahara, A. (2009). Noncanonical activity of seryl-tRNA synthetase is involved in vascular development. *Circulation Research*, *104*(11), 1253–1259. https://doi.org/10.1161/CIRCRESAHA.108.191189

G

Galani, K., Großhans, H., Deinert, K., Hurt, E. C., & Simos, G. (2001). The intracellular location of two aminoacyl-tRNA synthetases depends on complex formation with Arc1p. *EMBO Journal*, *20*(23), 6889–6898. https://doi.org/10.1093/emboj/20.23.6889

Ganapathiraju, M., Balakrishnan, N., Reddy, R., & Klein-Seetharaman, J. (2008). Transmembrane helix prediction using amino acid property features and latent semantic analysis. *BMC Bioinformatics*, *9*(1), S4.

Garcia, C. R. S., de Azevedo, M. F., Wunderlich, G., Budu, A., Young, J. A., & Bannister, L. (2008). Plasmodium in the Postgenomic Era: New Insights into the Molecular Cell Biology of Malaria Parasites. *International Review of Cell and Molecular Biology*, *266*, 85–156. https://doi.org/10.1016/S1937-6448(07)66003-1

Garcia Silva, M. R., Tosar, J. P., Frugier, M., Pantano, S., Bonilla, B., Esteban, L., Serra, E., Rovira, C., Robello, C., & Cayota, A. (2010). Cloning, characterization and subcellular localization of a Trypanosoma cruzi argonaute protein defining a new subfamily distinctive of trypanosomatids. *Gene*, *466*(1–2), 26–35. https://doi.org/10.1016/j.gene.2010.06.012

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M.-S., Nene, V., … Barrell, B. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, *419*(6906), 498–511. https://doi.org/10.1038/nature01097

Garten, M., Nasamu, A. S., Niles, J. C., Zimmerberg, J., Goldberg, D. E., & Beck, J. R. (2018). EXP2 is a nutrient-permeable channel in the vacuolar membrane of Plasmodium and is essential for protein export via PTEX. *Nature Microbiology*, *3*(10), 1090–1098. https://doi.org/10.1038/s41564-018-0222-7

Geary, T. G., Divo, A. A., Bonanni, L. C., & Jensen, J. B. (1985). Nutritional Requirements of Plasmodium falciparum in Culture. III. Further Observations on Essential Nutrients and Antimetabolites. *The Journal of Protozoology*, *32*(4), 608–613. https://doi.org/10.1111/j.1550-7408.1985.tb03087.x

Gehde, N., Hinrichs, C., Montilla, I., Charpian, S., Lingelbach, K., & Przyborski, J. M. (2009). Protein unfolding is an essential requirement for transport across the parasitophorous vacuolar membrane of Plasmodium falciparum. *Molecular Microbiology*, *71*(3), 613–628. https://doi.org/10.1111/j.1365-2958.2008.06552.x

Giegé, R., Sissler, M., & Florentz, C. (1998). Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Research*, *26*(22), 5017–5035. https://doi.org/10.1093/nar/26.22.5017

Giessen, T. W., Altegoer, F., Nebel, A. J., Steinbach, R. M., Bange, G., & Marahiel, M. A. (2015). A synthetic adenylation-domain-based tRNA-aminoacylation catalyst. *Angewandte Chemie - International Edition*, *54*(8), 2492–2496. https://doi.org/10.1002/anie.201410047

Gilson, P. R., Chisholm, S. A., Crabb, B. S., & de Koning-Ward, T. F. (2017). Host cell remodelling in malaria parasites: a new pool of potential drug targets. *International Journal for Parasitology*, *47*(2–3), 119–127. https://doi.org/10.1016/j.ijpara.2016.06.001

Gilson, P. R., & Crabb, B. S. (2009). Morphology and kinetics of the three distinct phases of red blood cell invasion by Plasmodium falciparum merozoites. *International Journal for Parasitology*, *39*(1), 91–96. https://doi.org/10.1016/j.ijpara.2008.09.007

Goldgur, Y., Mosyak, L., Reshetnikova, L., Ankilova, V., Lavrik, O., Khodyreva, S., & Safro, M. (1997). The crystal structure of phenylalanyl-tRNA synthetase from Thermus thermophilus complexed with cognate tRNA(Phe). *Structure*, *5*(1), 59–68. https://doi.org/10.1016/S0969-2126(97)00166-4

Goldsmith, C. (2011). *Battling malaria: on the front lines against a global killer*. Lerner Publishing Group.

Golinelli-Cohen, M. P., & Mirande, M. (2007). Arc1p is required for cytoplasmic confinement of synthetases and tRNA. *Molecular and Cellular Biochemistry*, *300*(1–2), 47–59. https://doi.org/10.1007/s11010-006-9367-4

Golinelli-Cohen, M. P., Zakrzewska, A., & Mirande, M. (2004). Complementation of yeast Arc1p by the p43 component of the human multisynthetase complex does not require its association with yeast MetRS and GluRS. *Journal of Molecular Biology*, *340*(1), 15–27. https://doi.org/10.1016/j.jmb.2004.04.040

Goodchild, S. C., Angstmann, C. N., Breit, S. N., Curmi, P. M. G., & Brown, L. J. (2011). Transmembrane extension and oligomerization of the CLIC1 chloride intracellular channel protein upon membrane interaction. *Biochemistry*, *50*(50), 10887–10897. https://doi.org/10.1021/bi2012564

Goodman, C. D., Buchanan, H. D., & McFadden, G. I. (2017). Is the Mitochondrion a Good Malaria Drug Target? *Trends in Parasitology*, *33*(3), 185–193. https://doi.org/10.1016/j.pt.2016.10.002

Goodman, C. D., Pasaje, C. F. A., Kennedy, K., McFadden, G. I., & Ralph, S. A. (2016). Targeting Protein Translation in Organelles of the Apicomplexa. *Trends in Parasitology*, *32*(12), 953–965. https://doi.org/10.1016/j.pt.2016.09.011

Goodsell, D. (2001). Aminoacyl-tRNA Synthetases. *RCSB PDB Molecule of the Month*. https://doi.org/10.2210/rcsb_pdb/mom_2001_4

Goss, D. J., & Kleiman, F. E. (2013). Poly(A) binding proteins: are they all created equal? *Wiley Interdiscip Rev RNA*, *4*(2), 167–179.

Gowri, V. S., Ghosh, I., Sharma, A., & Madhubala, R. (2012). Unusual domain architecture of aminoacyl tRNA synthetases and their paralogs from Leishmania major. *BMC Genomics*, *13*, 621. https://doi.org/10.1186/1471-2164-13-621

Graindorge, J. S., Senger, B., Tritch, D., Simos, G., & Fasiolo, F. (2005). Role of Arc1p in the modulation of yeast glutamyl-tRNA synthetase activity. *Biochemistry*, *44*(4), 1344–1352. https://doi.org/10.1021/bi049024z

Grant, T. D., Luft, J. R., Wolfley, J. R., Snell, M. E., Tsuruta, H., Corretore, S., Quartley, E., Phizicky, E. M., Grayhack, E. J., & Snell, E. H. (2013). The structure of yeast glutaminyl-tRNA synthetase and modeling of its interaction with tRNA. *Journal of Molecular Biology*, *425*(14), 2480–2493. https://doi.org/10.1016/j.jmb.2013.03.043

Grant, T. D., Snell, E. H., Luft, J. R., Quartley, E., Corretore, S., Wolfley, J. R., Elizabeth Snell, M., Hadd, A., Perona, J. J., Phizicky, E. M., & Grayhack, E. J. (2012). Structural conservation of an ancient tRNA sensor in eukaryotic glutaminyl-tRNA synthetase. *Nucleic Acids Research*, *40*(8), 3723–3731. https://doi.org/10.1093/nar/gkr1223

Gratzer, W. (1984). More red than dead. *Nature*, *310*, 368–369.

Greenwood, B., & Doumbo, O. K. (2016). Implementation of the malaria candidate vaccine RTS,S/AS01. *The Lancet*, *387*, 318–319. https://doi.org/10.1016/S0140-6736(15)00807-7

Gunderson, J. H., Sogin, M. L., Wollett, G., Hollingdale, M., de la Cruz, V. F., Waters, A. P., & McCutchan, T. F. (1987). Structurally Distinct, Stage-Specific Ribosomes Occur in Plasmodium. *Science*, *238*(4829), 933–937. https://doi.org/10.1126/science.3672135

Guo, M., & Schimmel, P. (2013). Essential nontranslational functions of tRNA synthetases. *Nature Chemical Biology*, *9*(3), 145–153. https://doi.org/10.1038/nchembio.1158

Guo, M., Schimmel, P., & Yang, X. L. (2010). Functional expansion of human tRNA synthetases achieved by structural inventions. *FEBS Letters*, *584*(2), 434–442. https://doi.org/10.1016/j.febslet.2009.11.064

Guo, M., & Yang, X. L. (2014). Architecture and metamorphosis. *Topics in Current Chemistry*, *344*, 89–118. https://doi.org/10.1007/128_2013_424

Gupta, A., Shah, P., Haider, A., Gupta, K., Siddiqi, M. I., Ralph, S. A., & Habib, S. (2014). Reduced ribosomes of the apicoplast and mitochondrion of Plasmodium spp. and predicted interactions with antibiotics. *Open Biology*, *4*(5), 140045. https://doi.org/10.1098/rsob.140045

Gupta, S., Chhibber-Goel, J., Sharma, M., Parvez, S., Harlos, K., Sharma, A., & Yogavel, M. (2020). Crystal structures of the two domains that constitute the Plasmodium vivax p43 protein. *Acta Crystallographica Section D: Structural Biology*, *76*(2), 135–146. https://doi.org/10.1107/S2059798319016413

H

Habib, S., Vaishya, S., & Gupta, K. (2016). Translation in Organelles of Apicomplexan Parasites. *Trends in Parasitology*, *32*(12), 939–952. https://doi.org/10.1016/j.pt.2016.07.005

Hadd, A., & Perona, J. J. (2014). Coevolution of specificity determinants in eukaryotic glutamyl- and glutaminyl-tRNA synthetases. *Journal of Molecular Biology*, *426*(21), 3619–3633. https://doi.org/10.1016/j.jmb.2014.08.006

Hahn, H., Park, S. H., Kim, H. J., Kim, S., & Han, B. W. (2019). The DRS-AIMP2-EPRS subcomplex acts as a pivot in the multi-tRNA synthetase complex. *IUCrJ*, *6*(5), 958–967. https://doi.org/10.1107/S2052252519010790

Haider, A., Allen, S. M., Jackson, K. E., Ralph, S. A., & Habib, S. (2015). Targeting and function of proteins mediating translation initiation in organelles of Plasmodium falciparum. *Molecular Microbiology*, *96*(4), 796–814. https://doi.org/10.1111/mmi.12972

Hajizadeh, N. R., Franke, D., Jeffries, C. M., & Svergun, D. I. (2018). Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data. *Scientific Reports*, *8*(1), 1–13. https://doi.org/10.1038/s41598-018-25355-2

Hakamada, K., Watanabe, H., Kawano, R., Noguchi, K., & Yohda, M. (2017). Expression and characterization of the Plasmodium translocon of the exported proteins component EXP2. *Biochemical and Biophysical Research Communications*, *482*(4), 700–705. https://doi.org/10.1016/j.bbrc.2016.11.097

Halfmann, R., Alberti, S., Krishnan, R., Lyle, N., O'Donnell, C. W., King, O. D., Berger, B., Pappu, R. V., & Lindquist, S. (2011). Opposing Effects of Glutamine and Asparagine Govern Prion Formation by Intrinsically Disordered Proteins. *Molecular Cell*, *43*(1), 72–84. https://doi.org/10.1016/j.molcel.2011.05.013

Hamilton, W. L., Claessens, A., Otto, T. D., Kekre, M., Fairhurst, R. M., Rayner, J. C., & Kwiatkowski, D. (2017). Extreme mutation bias and high AT content in Plasmodium falciparum. *Nucleic Acids Research*, *45*(4), 1889–1901. https://doi.org/10.1093/nar/gkw1259

Han, J. M., Jeong, S. J., Park, M. C., Kim, G., Kwon, N. H., Kim, H. K., Ha, S. H., Ryu, S. H., & Kim, S. (2012). Leucyl-tRNA synthetase is an intracellular leucine sensor for the mTORC1-signaling pathway. *Cell*, *149*(2), 410–424. https://doi.org/10.1016/j.cell.2012.02.044

Hancock, K., & Hajduk, S. L. (1990). The Mitochondrial tRNAs of Trypanosoma brucei Are Nuclear Encoded. *The Journal of Biological Chemistry*, *265*(31), 19208–19215.

Hare, J. E., Goodchild, S. C., Breit, S. N., Curmi, P. M. G., & Brown, L. J. (2016). Interaction of Human Chloride Intracellular Channel Protein 1 (CLIC1) with Lipid Bilayers: A Fluorescence Study. *Biochemistry*, *55*(27), 3825–3833. https://doi.org/10.1021/acs.biochem.6b00080

Havrylenko, S., Legouis, R., Negrutskii, B., & Mirande, M. (2010). Methionyl-tRNA synthetase from Caenorhabditis elegans: A specific multidomain organization for convergent functional evolution. *Protein Science*, *19*(12), 2475–2484. https://doi.org/10.1002/pro.529

Havrylenko, S., & Mirande, M. (2015). Aminoacyl-tRNA synthetase complexes in evolution. *International Journal of Molecular Sciences*, *16*(3), 6571–6594. https://doi.org/10.3390/ijms16036571

Hawass, Z., Gad, Y. Z., Ismail, S., Khairat, R., Fathalla, D., Hasan, N., Ahmed, A., Elleithy, H., Ball, M., Gaballah, F., Wasef, S., Fateen, M., Amer, H., Gostner, P., Selim, A., Zink, A., & Pusch, C. M. (2010). Ancestry and pathology in King Tutankhamun's family. *JAMA - Journal of the American Medical Association*, *303*(7), 638–647. https://doi.org/10.1001/jama.2010.121

Heiber, A., Kruse, F., Pick, C., Grüring, C., Flemming, S., Oberli, A., Schoeler, H., Retzlaff, S., Mesén-Ramírez, P., Hiss, J. A., Kadekoppala, M., Hecht, L., Holder, A. A., Gilberger, T. W., & Spielmann, T. (2013). Identification of New PNEPs Indicates a Substantial Non-PEXEL Exportome and Underpins Common Features in Plasmodium falciparum Protein Export. *PLoS Pathogens*, *9*(8), e1003546. https://doi.org/10.1371/journal.ppat.1003546

Hellen, C. U. T. (2018). Translation termination and ribosome recycling in eukaryotes. *Cold Spring Harbor Perspectives in Biology*, *10*(10), a032656. https://doi.org/10.1101/cshperspect.a032656

Herman, J. D., Pepper, L. R., Cortese, J. F., Estiu, G., Galinsky, K., Zuzarte-Luis, V., Derbyshire, E. R., Ribacke, U., Lukens, A. K., Santos, S. A., Patel, V., Clish, C. B., Sullivan, W. J., Zhou, H., Bopp, S. E., Schimmel, P., Lindquist, S., Clardy, J., Mota, M. M., … Mazitschek, R. (2015). The cytoplasmic prolyl-tRNA synthetase of the malaria parasite is a dual-stage target of febrifugine and its analogs. *Science Translational Medicine*, *7*(288), 288ra77. https://doi.org/10.1126/scitranslmed.aaa3575

Hiller, N. L., Bhattacharjee, S., Van Ooij, C., Liolios, K., Harrison, T., Lopez-Estraño, C., & Haldar, K. (2004). A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science*, *306*(5703), 1934–1937. https://doi.org/10.1126/science.1102737

Ho, C. M., Beck, J. R., Lai, M., Cui, Y., Goldberg, D. E., Egea, P. F., & Zhou, Z. H. (2018). Malaria parasite translocon structure and mechanism of effector export. *Nature*, *561*(7721), 70–75. https://doi.org/10.1038/s41586-018-0469-4

Hoepfner, D., McNamara, C. W., Lim, C. S., Studer, C., Riedl, R., Aust, T., McCormack, S. L., Plouffe, D. M., Meister, S., Schuierer, S., Plikat, U., Hartmann, N., Staedtler, F., Cotesta, S., Schmitt, E. K., Petersen, F., Supek, F., Glynne, R. J., Tallarico, J. A., … Winzeler, E. A. (2012). Selective and specific inhibition of the plasmodium falciparum lysyl-tRNA synthetase by the fungal secondary metabolite cladosporin. *Cell Host and Microbe*, *11*(6), 654–663. https://doi.org/10.1016/j.chom.2012.04.015

Hopkins, J. B., Gillilan, R. E., & Skou, S. (2017). BioXTAS RAW: Improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *Journal of Applied Crystallography*, *50*(5), 1545–1553. https://doi.org/10.1107/S1600576717011438

Hopper, A. K., & Nostramo, R. T. (2019). tRNA processing and subcellular trafficking proteins multitask in pathways for other RNAs. *Frontiers in Genetics*, *10*(96). https://doi.org/10.3389/fgene.2019.00096

Hughes, A. L. (2004). The evolution of amino acid repeat arrays in Plasmodium and other organisms. *Journal of Molecular Evolution*, *59*(4), 528–535. https://doi.org/10.1007/s00239-004-2645-4

Hyeon, D. Y., Kim, J. H., Ahn, T. J., Cho, Y., Hwang, D., & Kim, S. (2019). Evolution of the multi-tRNA synthetase complex and its role in cancer. *Journal of Biological Chemistry*, *294*(14), 5340–5351. https://doi.org/10.1074/jbc.REV118.002958

I

Ibba, M., & Soll, D. (2000). Aminoacyl-tRNA synthesis. *Annual Review of Biochemistry*, *69*, 617–650.

Idro, R., Marsh, K., John, C. C., & Newton, C. R. J. (2010). Cerebral malaria: Mechanisms of brain injury and strategies for improved neurocognitive outcome. *Pediatric Research*, *68*(4), 267–274. https://doi.org/10.1203/PDR.0b013e3181eee738

Ilyin, V. A., Temple, B., Hu, M., Li, G., Yin, Y., Vachette, P., & Carter, C. W. (2000). 2.9 Å Crystal structure of ligand-free tryptophanyl-tRNA synthetase: Domain movements fragment the adenine nucleotide binding site. *Protein Science*, *9*(2), 218–231.

Imprasittichail, W., Roytrakul, S., Krungkrai, S. R., & Krungkrail, J. (2014). A unique insertion of low complexity amino acid sequence underlies protein-protein interaction in human malaria parasite orotate phosphoribosyltransferase and orotidine 5'-monophosphate decarboxylase. *Asian Pacific Journal of Tropical Medicine*, *7*(3), 184–192. https://doi.org/10.1016/S1995-7645(14)60018-3

Istvan, E. S., Dharia, N. V., Bopp, S. E., Gluzman, I., Winzeler, E. A., & Goldberg, D. E. (2011). Validation of isoleucine utilization targets in Plasmodium falciparum. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1627–1632. https://doi.org/10.1073/pnas.1011560108

J

Jackson, K. E., Habib, S., Frugier, M., Hoen, R., Khan, S., Pham, J. S., Pouplana, L. R. de, Royo, M., Santos, M. A. S., Sharma, A., & Ralph, S. A. (2011). Protein translation in Plasmodium parasites. *Trends in Parasitology*, *27*(10), 467–476. https://doi.org/10.1016/j.pt.2011.05.005

Jackson, K. E., Pham, J. S., Kwek, M., De Silva, N. S., Allen, S. M., Goodman, C. D., McFadden, G. I., Ribas de Pouplana, L., & Ralph, S. A. (2012). Dual targeting of aminoacyl-tRNA synthetases to the apicoplast and cytosol in Plasmodium falciparum. *International Journal for Parasitology*, *42*(2), 177–186. https://doi.org/10.1016/j.ijpara.2011.11.008

Jain, V., Kikuchi, H., Oshima, Y., Sharma, A., & Yogavel, M. (2014). Structural and functional analysis of the anti-malarial drug target prolyl-tRNA synthetase. *Journal of Structural and Functional Genomics*, *15*(4), 181–190. https://doi.org/10.1007/s10969-014-9186-x

Jain, V., Yogavel, M., & Sharma, A. (2016). Dimerization of Arginyl-tRNA Synthetase by Free Heme Drives Its Inactivation in Plasmodium falciparum. *Structure*, *24*(9), 1476–1487. https://doi.org/10.1016/j.str.2016.06.018

Jeppesen, M. G., Ortiz, P., Shepard, W., Goss Kinzy, T., Nyborg, J., & Andersen, G. R. (2003). The Crystal Structure of the GST-like domain of Elongation Factor 1Bγ from Saccharomyces cerevisiae. *Journal of Biological Chemistry*, *278*(47), 47190–47198.

Ji, Y.-E., Mericlea, B. L., Rehkopp, D. H., Feagirpb, J. E., & Anderson, J. D. (1996). The Plasmodium falciparum 6 kb element is polycistronically transcribed. *Molecular and Biochemical Parasitology*, *81*(2), 211–223. https://doi.org/10.1016/0166-6851(96)02712-0.

Jia, J., Arif, A., Ray, P. S., & Fox, P. L. (2008). WHEP Domains Direct Noncanonical Function of Glutamyl-Prolyl tRNA Synthetase in Translational Control of Gene Expression. *Molecular Cell*, *29*(6), 679–690. https://doi.org/10.1016/j.molcel.2008.01.010

Johnson, R. A., McFadden, G. I., & Goodman, C. D. (2011). Characterization of two malaria parasite organelle translation elongation factor G proteins: The likely targets of the anti-malarial fusidic acid. *PLoS ONE*, *6*(6), e20633. https://doi.org/10.1371/journal.pone.0020633

K

Kabanova, S., Kleinbongard, P., Volkmer, J., Andrée, B., Kelm, M., & Jax, T. W. (2009). Gene expression analysis of human red blood cells. *International Journal of Medical Sciences*, *6*(4), 156–159.

Kaiser, A., Hammels, I., Gottwald, A., Nassar, M., Zaghloul, M. S., Motaal, B. A., Hauber, J., & Hoerauf, A. (2007). Modification of eukaryotic initiation factor 5A from Plasmodium vivax by a truncated deoxyhypusine synthase from Plasmodium falciparum: An enzyme with dual enzymatic properties. *Bioorganic and Medicinal Chemistry*, *15*(18), 6200–6207. https://doi.org/10.1016/j.bmc.2007.06.026

Kaminska, M., Deniziak, M., Kerjan, P., Barciszewski, J., & Mirande, M. (1999). A recurrent general RNA binding domain appended to plant methionyl-tRNA synthetase acts as a cis-acting cofactor for aminoacylation. *The EMBO Journal*, *19*(24), 6908–6917.

Kaminska, M., Francin, M., Shalak, V., & Mirande, M. (2007). Role of HIV-1 Vpr-induced apoptosis on the release of mitochondrial lysyl-tRNA synthetase. *FEBS Letters*, *581*(16), 3105–3110. https://doi.org/10.1016/j.febslet.2007.05.076

Kaminska, M., Havrylenko, S., Decottignies, P., Gillet, S., Le Maréchal, P., Negrutskii, B., & Mirande, M. (2009). Dissection of the structural organization of the aminoacyl-tRNA synthetase complex. *Journal of Biological Chemistry*, *284*(10), 6053–6060. https://doi.org/10.1074/jbc.M809636200

Kaminska, M., Shalak, V., & Mirande, M. (2001). The appended C-domain of human methionyl-tRNA synthetase has a tRNA-sequestering function. *Biochemistry*, *40*(47), 14309–14316. https://doi.org/10.1021/bi015670b

Kanaya, S., Yamada, Y., Kinouchi, M., Kudo, Y., & Ikemura, T. (2001). Codon usage and tRNA genes in eukaryotes: Correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *Journal of Molecular Evolution*, *53*(4–5), 290–298. https://doi.org/10.1007/s002390010219

Kao, J., Fan, Y. G., Haehnel, I., Brett, J., Greenberg, S., Clauss, M., Kayton, M., Houck, K., Kisiel, W., Seljelid, R., Burnier, J., & Stern, D. (1994). A peptide derived from the amino terminus of endothelial-monocyte-activating polypeptide II modulates mononuclear and polymorphonuclear leukocyte functions, defines an apparently novel cellular interaction site, and induces an acute inflammatory response. *Journal of Biological Chemistry*, *269*(13), 9774–9782.

Kao, J., Ryan, J., Brett, G., Chen, J., Shen, H., Fan, Y.-G., Godman, G., Famillettill, P. C., Wangll, F., Panli, Y.-C. E., Sternqii, D., & Clauss, M. (1992). Endothelial Monocyte-activating Polypeptide II. A Novel Tumor-derived polypeptide that activates host-response mechanisms. *The Journal of Biological Chemistry*, *267*(28), 20239–20247.

Kapishnikov, S., Grolimund, D., Schneider, G., Pereiro, E., McNally, J. G., Als-Nielsen, J., & Leiserowitz, L. (2017). Unraveling heme detoxification in the malaria parasite by in situ correlative X-ray fluorescence microscopy and soft X-ray tomography. *Scientific Reports*, *7*(1), 7610. https://doi.org/10.1038/s41598-017-06650-w

Kapps, D., Cela, M., Théobald-Dietrich, A., Hendrickson, T., & Frugier, M. (2016). OB or Not OB: Idiosyncratic utilization of the tRNA-binding OB-fold domain in unicellular, pathogenic eukaryotes. *FEBS Letters*, *590*(23), 4180–4191. https://doi.org/10.1002/1873-3468.12441

Kato, N., Comer, E., Sakata-Kato, T., Sharma, A., Sharma, M., Maetani, M., Bastien, J., Brancucci, N. M., Bittker, J. A., Corey, V., Clarke, D., Derbyshire, E. R., Dornan, G. L., Duffy, S., Eckley, S., Itoe, M. A., Koolen, K. M. J., Lewis, T. A., Lui, P. S., … Schreiber, S. L. (2016). Diversity-oriented synthesis yields novel multistage antimalarial inhibitors. *Nature*, *538*(7625), 344–349. https://doi.org/10.1038/nature19804

Ke, H., & Mather, M. W. (2017). +Targeting Mitochondrial Functions as Antimalarial Regime, What Is Next? *Current Clinical Microbiology Reports*, *4*, 175–191. https://doi.org/10.1007/s40588-017-0075-5

Keller, T. L., Zocco, D., Sundrud, M. S., Hendrick, M., Edenius, M., Yum, J., Kim, Y. J., Lee, H. K., Cortese, J. F., Wirth, D. F., Dignam, J. D., Rao, A., Yeo, C. Y., Mazitschek, R., & Whitman, M. (2012). Halofuginone and other febrifugine derivatives inhibit prolyl-tRNA synthetase. *Nature Chemical Biology*, *8*(3), 311–317. https://doi.org/10.1038/nchembio.790

Kemp, D. J., Coppel, R. L., & Anders, R. F. (1987). Repetitive proteins and genes of malaria. *Annual Review of Microbiology*, *41*, 181–208. https://doi.org/10.1146/annurev.micro.41.1.181

Khan, K., Baleanu-Gogonea, C., Willard, B., Gogonea, V., & Fox, P. L. (2020). 3-Dimensional architecture of the human multi-tRNA synthetase complex. *Nucleic Acids Research*, *48*(15), 8740–8754. https://doi.org/10.1093/nar/gkaa569

Khan, S. (2016). Recent advances in the biology and drug targeting of malaria parasite aminoacyl-tRNA synthetases. *Malaria Journal*, *15*, 203. https://doi.org/10.1186/s12936-016-1247-0

Khan, S., Garg, A., Camacho, N., Van Rooyen, J., Kumar Pole, A., Belrhali, H., Ribas De Pouplana, L., Sharma, V., & Sharma, A. (2013). Structural analysis of malaria-parasite lysyl-tRNA synthetase provides a platform for drug development. *Acta Crystallographica Section D: Biological Crystallography*, *69*(5), 785–795. https://doi.org/10.1107/S0907444913001923

Khan, S., Garg, A., Sharma, A., Camacho, N., Picchioni, D., Saint-Léger, A., de Pouplana, L. R., Yogavel, M., & Sharma, A. (2013). An Appended Domain Results in an Unusual Architecture for Malaria Parasite Tryptophanyl-tRNA Synthetase. *PLoS ONE*, *8*(6), e66224. https://doi.org/10.1371/journal.pone.0066224

Kim, D. G., Choi, J. W., Lee, J. Y., Kim, H., Oh, Y. S., Lee, J. W., Tak, Y. K., Song, J. M., Razin, E., Yun, S. H., & Kim, S. (2012). Interaction of two translational components, lysyl-tRNA synthetase and p40/37LRP, in plasma membrane promotes laminin-dependent cell migration. *FASEB Journal*, *26*(10), 4142–4159. https://doi.org/10.1096/fj.12-207639

Kim, J. E., Kim, K. H., Lee, S. W., Seol, W., Shiba, K., & Kima, S. (2000). An elongation factor-associating domain is inserted into human cysteinyl-tRNA synthetase by alternative splicing. *Nucleic Acids Research*, *28*(15), 2866–2872. https://doi.org/10.1093/nar/28.15.2866

Kim, J., Tan, Y. Z., Wicht, K. J., Erramilli, S. K., Dhingra, S. K., Okombo, J., Vendome, J., Hagenah, L. M., Giacometti, S. I., Warren, A. L., Nosol, K., Roepe, P. D., Potter, C. S., Carragher, B., Kossiakoff, A. A., Quick, M., Fidock, D. A., & Mancia, F. (2019). Structure and drug resistance of the Plasmodium falciparum transporter PfCRT. *Nature*, *576*(7786), 315–320. https://doi.org/10.1038/s41586-019-1795-x

Kim, J. Y., Kang, Y. S., Lee, J. W., Kim, H. J., Ahn, Y. H., Park, H., Ko, Y. G., & Kim, S. (2002). p38 is essential for the assembly and stability of macromolecular tRNA synthetase complex: Implications for its physiological significance. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 7912–7916. https://doi.org/10.1073/pnas.122110199

Kim, K. J., Min, C. P., So, J. C., Young, S. O., Choi, E. C., Hyo, J. C., Myung, H. K., Kim, S. H., Dong, W. K., Kim, S., & Beom, S. K. (2008). Determination of three-dimensional structure and residues of the novel tumor suppressor AIMP3/p18 required for the interaction with ATM. *Journal of Biological Chemistry*, *283*(20), 14032–14040. https://doi.org/10.1074/jbc.M800859200

Kim, T., Park, S. G., Kim, J. E., Seol, W., Ko, Y. G., & Kim, S. (2000). Catalytic peptide of human glutaminyl-tRNA synthetase is essential for its assembly to the aminoacyl-tRNA synthetase. *Journal of Biological Chemistry*, *275*(28), 21768–21772. https://doi.org/10.1074/jbc.M002404200

Kim, Y., Shin, J., Li, R., Cheong, C., Kim, K., & Kim, S. (2000). A novel anti-tumor cytokine contains an RNA binding motif present in aminoacyl-tRNA synthetases. *Journal of Biological Chemistry*, *275*(35), 27062–27068. https://doi.org/10.1074/jbc.C000216200

Koehler, C., Round, A., Simader, H., Suck, D., & Svergun, D. (2013). Quaternary structure of the yeast Arc1p-aminoacyl-tRNA synthetase complex in solution and its compaction upon binding of tRNAs. *Nucleic Acids Research*, *41*(1), 667–676. https://doi.org/10.1093/nar/gks1072

Koh, C. Y., Kim, J. E., Napoli, A. J., Verlinde, C. L. M. J., Fan, E., Buckner, F. S., Van Voorhis, W. C., & Hol, W. G. J. (2013). Crystal structures of Plasmodium falciparum cytosolic tryptophanyl-tRNA synthetase and its potential as a target for structure-guided drug design. *Molecular and Biochemical Parasitology*, *189*(1–2), 26–32. https://doi.org/10.1016/j.molbiopara.2013.04.007

Koh, C. Y., Kim, J. E., Wetzel, A. B., de van der Schueren, W. J., Shibata, S., Ranade, R. M., Liu, J., Zhang, Z., Gillespie, J. R., Buckner, F. S., Verlinde, C. L. M. J., Fan, E., & Hol, W. G. J. (2014). Structures of Trypanosoma brucei Methionyl-tRNA Synthetase with Urea-Based Inhibitors Provide Guidance for Drug Design against Sleeping Sickness. *PLoS Neglected Tropical Diseases*, *8*(4), e2775. https://doi.org/10.1371/journal.pntd.0002775

Komar, A. A. (2016). The Yin and Yang of Codon Usage. *Human Molecular Genetics*, *25*(R2), R77–R85.

Kooij, T. W. A., Janse, C. J., & Waters, A. P. (2006). Plasmodium post-genomics: Better the bug you know? *Nature Reviews Microbiology*, *4*(5), 344–357. https://doi.org/10.1038/nrmicro1392

Koonin, E. V., Mushegian, A. R., Tatusov, R. L., Altschul, S. F., Bryant, S. H., Bork, P., & Valencia, A. (1994). Eukaryotic translation elongation factor 1gamma contains a glutathion transferase domain - Study of a diverse, ancient protein superfamily using motif search and structural modeling. *Protein Science*, *3*(11), 2045–2054. https://doi.org/10.3180/react_1477.1

Koppel, D. E. (1972). Analysis of macromolecular polydispersity in intensity correlation spectroscopy: The method of cumulants. *The Journal of Chemical Physics*, *57*(11), 4814–4820. https://doi.org/10.1063/1.1678153

Koutmou, K. S., Schuller, A. P., Brunelle, J. L., Radhakrishnan, A., Djuranovic, S., & Green, R. (2015). Ribosomes slide on lysine-encoding homopolymeric A stretches. *ELife*, *4*, e05534. https://doi.org/10.7554/eLife.05534

L

Lalchhandama, K. (2014). The making of modern malariology: from miasma to mosquito-malaria theory. *Science Vision*, *14*(1), 2–17.

Lamarque, M., Tastet, C., Poncet, J., Demettre, E., Jouin, P., Vial, H., & Dubremetz, J. F. (2008). Food vacuole proteome of the malarial parasite Plasmodium falciparum. *Proteomics - Clinical Applications*, *2*(9), 1361–1374. https://doi.org/10.1002/prca.200700112

Lamonte, G., Philip, N., Reardon, J., Lacsina, J. R., Chapman, L., Thornburg, C. D., Telen, M. J., Ohler, U., Nicchitta, C. V, Haystead, T., & Chi, J. (2012). Translocation of sickle cell erythrocyte microRNAs into Plasmodium falciparum inhibits parasite translation and contributes to malaria resistance. *Cell Host Microbe*, *12*(2), 187–199. https://doi.org/10.1016/j.chom.2012.06.007.Translocation

Lamour, V., Quevillon, S., Diriong, S., N'Guyen, V. C., Lipinski, M., & Mirande, M. (1994). Evolution of the Glx-tRNA synthetase family: The glutaminyl enzyme as a case of horizontal gene transfer. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(18), 8670–8674. https://doi.org/10.1073/pnas.91.18.8670

Laporte, D., Huot, J. L., Bader, G., Enkler, L., Senger, B., & Becker, H. D. (2014). Exploring the evolutionary diversity and assembly modes of multi-aminoacyl-tRNA synthetase complexes: Lessons from unicellular organisms. *FEBS Letters*, *588*(23), 4268–4278. https://doi.org/10.1016/j.febslet.2014.10.007

Larson, E. T., Kim, J. E., Zucker, F. H., Kelley, A., Mueller, N., Napuli, A. J., Verlinde, C. L. M. J., Fan, E., Buckner, F. S., Van Voorhis, W. C., Merritt, E. A., & Hol, W. G. J. (2011). Structure of Leishmania major methionyl-tRNA synthetase in complex with intermediate products methionyladenylate and pyrophosphate. *Biochimie*, *93*(3), 570–582. https://doi.org/10.1016/j.biochi.2010.11.015

Laurens, M. B. (2018). The Promise of a Malaria Vaccine - Are We Closer? *Annual Review of Microbiology*, *72*, 273–292. https://doi.org/10.1146/annurev-micro-090817-062427

Lee, S. W., Cho, B. H., Park, S. G., & Kim, S. (2004). Aminoacyl-tRNA synthetase complexes: Beyond translation. *Journal of Cell Science*, *117*(17), 3725–3734. https://doi.org/10.1242/jcs.01342

Lee, W. C., Russell, B., & Rénia, L. (2019). Sticking for a cause: The falciparum malaria parasites cytoadherence paradigm. *Frontiers in Immunology*, *10*(1444). https://doi.org/10.3389/fimmu.2019.01444

Legrand, P. (2017). XDSME: XDS Made Easier. *GitHub Repository*, *https://github.com/legrandp/xdsme.* https://doi.org/DOI 10.5281/zenodo.837885

Lew, V. L., Tiffert, T., & Ginsburg, H. (2003). Excess hemoglobin digestion and the osmotic stability of Plasmodium falciparum - Infected red blood cells. *Blood*, *101*(10), 4189–4194. https://doi.org/10.1182/blood-2002-08-2654

Li, J., Gutell, R. R., Damberger, S. H., Wirtz, R. A., Kissinger, J. C., Rogers, M. J., Sattabongkot, J., & McCutchan, T. F. (1997). Regulation and trafficking of three distinct 18 S ribosomal RNAs during development of the malaria parasite. *Journal of Molecular Biology*, *269*(2), 203–213. https://doi.org/10.1006/jmbi.1997.1038

Liebau, E., Dawood, K. F., Fabrini, R., Fischer-Riepe, L., Perbandt, M., Stella, L., Pedersen, J. Z., Bocedi, A., Petrarca, P., Federici, G., & Ricci, G. (2009). Tetramerization and cooperativity in Plasmodium falciparum glutathione S-transferase are mediated by atypic loop 113-119. *Journal of Biological Chemistry*, *284*(33), 22133–22139. https://doi.org/10.1074/jbc.M109.015198

Lim, L., & McFadden, G. I. (2010). The evolution, metabolism and functions of the apicoplast. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*, 749–763. https://doi.org/10.1098/rstb.2009.0273

Lindner, S. E., Swearingen, K. E., Harupa, A., Vaughan, A. M., Sinnis, P., Moritz, R. L., & Kappe, S. H. I. (2013). Total and putative surface proteomics of malaria parasite salivary gland sporozoites. *Molecular & Cellular Proteomics : MCP*, *12*(5), 1127–1143. https://doi.org/10.1074/mcp.M112.024505

Lingelbach, K. R. (1993). Plasmodium falciparum: A molecular view of protein transport from the parasite into the host erythrocyte. *Experimental Parasitology*, *76*, 318–327. https://doi.org/10.1006/expr.1993.1039

Liu, J., Istvan, E. S., Gluzman, I. Y., Gross, J., & Goldberg, D. E. (2006). Plasmodium falciparum ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(23), 8840–8845. https://doi.org/10.1073/pnas.0601876103

Lobanov, A. V., Delgado, C., Rahlfs, S., Novoselov, S. V., Kryukov, G. V., Gromer, S., Hatfield, D. L., Becker, K., & Gladyshev, V. N. (2006). The Plasmodium selenoproteome. *Nucleic Acids Research*, *34*(2), 496–505. https://doi.org/10.1093/nar/gkj450

Lorber, B. (2020). Analytical light scattering methods in molecular and structural biology : Some experimental aspects and results. *HAL Archives-Ouvertes*, hal-02909121.

Lorber, B., Fischer, F., Bailly, M., Roy, H., & Kern, D. (2012). Protein analysis by dynamic light scattering: Methods and techniques for students. *Biochemistry and Molecular Biology Education*, *40*(6), 372–382. https://doi.org/10.1002/bmb.20644

Lund, E., & Dahlberg, J. E. (1998). Proofreading and aminoacylation of tRNAs before export from the nucleus. *Science*, *282*(5396), 2082–2085. https://doi.org/10.1126/science.282.5396.2082

Lustig, B., Arora, S., & Jernigan, R. L. (1997). RNA base-amino acid interaction strengths derived from structures and sequences. *Nucleic Acids Research*, *25*(13), 2562–2565. https://doi.org/10.1093/nar/25.13.2562

M

MacRaild, C. A., Richards, J. S., Anders, R. F., & Norton, R. S. (2016). Antibody Recognition of Disordered Antigens. *Structure*, *24*(1), 148–157. https://doi.org/10.1016/j.str.2015.10.028

Mailu, B. M., Li, L., Arthur, J., Nelson, T. M., Ramasamy, G., Fritz-Wolf, K., Becker, K., & Gardner, M. J. (2015). Plasmodium apicoplast Gln-tRNAGln biosynthesis utilizes a unique GatAB amidotransferase essential for erythrocytic stage parasites. *Journal of Biological Chemistry*, *290*(49), 29629–29641. https://doi.org/10.1074/jbc.M115.655100

Mailu, B. M., Ramasamay, G., Mudeppa, D. G., Li, L., Lindner, S. E., Peterson, M. J., DeRocher, A. E., Kappe, S. H. I., Rathod, P. K., & Gardner, M. J. (2013). A nondiscriminating glutamyl-tRNA synthetase in the plasmodium apicoplast: The first enzyme in an indirect aminoacylation pathway. *Journal of Biological Chemistry*, *288*(45), 32539–32552. https://doi.org/10.1074/jbc.M113.507467

Malaria Vaccine Funders Group. (2013). *Malaria Vaccine Technology Roadmap*.

Mantel, P. Y., Hoang, A. N., Goldowitz, I., Potashnikova, D., Hamza, B., Vorobjev, I., Ghiran, I., Toner, M., Irimia, D., Ivanov, A. R., Barteneva, N., & Marti, M. (2013). Malaria-infected erythrocyte-derived microvesicles mediate cellular communication within the parasite population and with the host immune system. *Cell Host and Microbe*, *13*(5), 521–534. https://doi.org/10.1016/j.chom.2013.04.009

Marapana, D. S., Dagley, L. F., Sandow, J. J., Nebl, T., Triglia, T., Pasternak, M., Dickerman, B. K., Crabb, B. S., Gilson, P. R., Webb, A. I., Boddey, J. A., & Cowman, A. F. (2018). Plasmepsin V cleaves malaria effector proteins in a distinct endoplasmic reticulum translocation interactome for export to the erythrocyte. *Nature Microbiology*, *3*(9), 1010–1022. https://doi.org/10.1038/s41564-018-0219-2

Marín-Navarro, J., Manuell, A. L., Wu, J., & Mayfield, S. P. (2007). Chloroplast translation regulation. *Photosynthesis Research*, *94*(2–3), 359–374. https://doi.org/10.1007/s11120-007-9183-z

Markus, M. B. (2020). Transition from Plasmodial Hypnozoite to Schizont Demonstrated. *Trends in Parasitology*, *36*(5), 407–408. https://doi.org/10.1016/j.pt.2020.01.011

Marti, M., Good, R. T., Rug, M., Knuepfer, E., & Cowman, A. F. (2004). Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science*, *306*(5703), 1930–1933. https://doi.org/10.1126/science.1102452

Martin, R. E. (2020). The transportome of the malaria parasite. *Biological Reviews*, *95*(2), 305–332. https://doi.org/10.1111/brv.12565

Martin, R. E., Henry, R. I., Abbey, J. L., Clements, J. D., & Kirk, K. (2005). The "permeome" of the malaria parasite: an overview of the membrane transport proteins of Plasmodium falciparum. *Genome Biology*, *6*(3), R26. https://doi.org/10.1186/gb-2005-6-3-r26

Martin, R. P., Schneller, J. M., Stahl, A. J. C., & Dirheimer, G. (1977). Study of yeast mitochondrial tRNAs by two-dimensional polyacrylamide gel electrophoresis: Characterization of isoaccepting species and search for imported cytoplasmic tRNAs. *Nucleic Acids Research*, *4*(10), 3497–3510. https://doi.org/10.1093/nar/4.10.3497

Matthews, K., Kalanon, M., Chisholm, S. A., Sturm, A., Goodman, C. D., Dixon, M. W. A., Sanders, P. R., Nebl, T., Fraser, F., Haase, S., Mcfadden, G. I., Gilson, P. R., Crabb, B. S., & De Koning-Ward, T. F. (2013). The Plasmodium translocon of exported proteins (PTEX) component thioredoxin-2 is important for maintaining normal blood-stage growth. *Molecular Microbiology*, *89*(6), 1167–1186. https://doi.org/10.1111/mmi.12334

Matthews, K. M., Pitman, E. L., & de Koning-Ward, T. F. (2019). Illuminating how malaria parasites export proteins into host erythrocytes. *Cellular Microbiology*, *21*(4), e13009. https://doi.org/10.1111/cmi.13009

Matz, J. M., Matuschewski, K., & Kooij, T. W. A. (2013). Two putative protein export regulators promote Plasmodium blood stage development in vivo. *Molecular and Biochemical Parasitology*, *191*(1), 44–52. https://doi.org/10.1016/j.molbiopara.2013.09.003

McCutchan, T. F., de la Cruz, V. F., Lal, A. A., Gunderson, J. H., Elwood, H. J., & Sogin, M. L. (1988). Primary sequences of two small subunit ribosomal RNA genes from Plasmodium falciparum. *Molecular and Biochemical Parasitology*, *28*(1), 63–68. https://doi.org/10.1016/0166-6851(88)90181-8

McFadden, G. I., & Yeh, E. (2017). The apicoplast: now you see it, now you don't. *International Journal for Parasitology*, *47*(2–3), 137–144. https://doi.org/10.1016/j.ijpara.2016.08.005

Mechulam, Y., Schmitt, E., Maveyraud, L., Zelwer, C., Nureki, O., Yokoyama, S., Konno, M., & Blanquet, S. (1999). Crystal structure of Escherichia coli methionyl-tRNA synthetase highlights species-specific features. *Journal of Molecular Biology*, *294*(5), 1287–1297. https://doi.org/10.1006/jmbi.1999.3339

Melnikov, S., Ben-Shem, A., Garreau De Loubresse, N., Jenner, L., Yusupova, G., & Yusupov, M. (2012). One core, two shells: Bacterial and eukaryotic ribosomes. *Nature Structural and Molecular Biology*, *19*(6), 560–567. https://doi.org/10.1038/nsmb.2313

Milton, M. E., & Nelson, S. W. (2016). Replication and maintenance of the Plasmodium falciparum apicoplast genome. *Molecular and Biochemical Parasitology*, *208*(2), 56–64. https://doi.org/10.1016/j.molbiopara.2016.06.006

Minns, A. M., Hart, K. J., Subramanian, S., Hafenstein, S., & Lindner, S. E. (2018). Nuclear, Cytosolic, and Surface-Localized Poly(A)-Binding Proteins of Plasmodium yoelii. *MSphere*, *3*(1), :e00435-17. https://doi.org/10.1128/msphere.00435-17

Mirande, M. (1991). Aminoacyl-tRNA Synthetase Family from Prokaryotes and Eukaryotes: Structural Domains and Their Implications. *Progress in Nucleic Acid Research and Molecular Biology*, *40*, 95–142. https://doi.org/10.1016/S0079-6603(08)60840-5

Mirande, M. (2017). The aminoacyl-tRNA synthetase complex. *Sub-Cellular Biochemistry*, *83*, 505–522. https://doi.org/10.1007/978-3-319-46503-6_18

Mirande, M., Le Corre, D., & Waller, J. -P. (1985). A complex from cultured Chinese hamster ovary cells containing nine aminoacyl-tRNA synthetases: Thermolabile leucyl-tRNA synthetase from the tsH1 mutant cell line is an integral component of this complex. *European Journal of Biochemistry*, *147*(2), 281–289. https://doi.org/10.1111/j.1432-1033.1985.tb08748.x

Morales, A. J., Swairjo, M. A., & Schimmel, P. (1999). Structure-specific tRNA-binding protein from the extreme thermophile Aquifex aeolicus. *The EMBO Journal*, *18*(12), 3475–3483. https://doi.org/10.1093/emboj/18.12.3475

Moras, D. (1992). Structural and functional relationships between aminoacyl-tRNA synthetases. *Trends in Biochemical Sciences*, *17*(4), 159–164. https://doi.org/10.1016/0968-0004(92)90326-5

Moriyama, E. N., & Powell, J. R. (1997). Codon usage bias and tRNA abundance in Drosophila. *Journal of Molecular Evolution*, *45*(5), 514–523. https://doi.org/10.1007/PL00006256

Mosyak, L., Reshetnikova, L., Goldgur, Y., Delarue, M., & Safro, M. (1995). Structure of phenylalanyl-tRNA synthetase from Thermus thermophilus. *Nature Structural Biology*, *2*(7), 537–547.

Mourier, T., Pain, A., Barrell, B., & Griffiths-Jones, S. (2005). A selenocysteine tRNA and SECIS element in Plasmodium falciparum. *RNA*, *11*(2), 119–122. https://doi.org/10.1261/rna.7185605

Mukhopadhyay, R., Jia, J., Arif, A., Ray, P. S., & Fox, P. L. (2009). The GAIT system: a gatekeeper of inflammatory gene expression. *Trends in Biochemical Sciences*, *34*(7), 324–331. https://doi.org/10.1016/j.tibs.2009.03.004

Muralidharan, V., & Goldberg, D. E. (2013). Asparagine Repeats in Plasmodium falciparum Proteins: Good for Nothing? *PLoS Pathogens*, *9*(8), e1003488. https://doi.org/10.1371/journal.ppat.1003488

Muralidharan, V., Oksman, A., Iwamoto, M., Wandless, T. J., & Goldberg, D. E. (2011). Asparagine repeat function in a Plasmodium falciparum protein assessed via a regulatable fluorescent affinity tag. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(11), 4411–4416. https://doi.org/10.1073/pnas.1018449108

Muralidharan, V., Oksman, A., Pal, P., Lindquist, S., & Goldberg, D. E. (2012). Plasmodium falciparumheat shock protein 110 stabilizes the asparagine repeat-rich parasite proteome during malarial fevers. *Nature Communications*, *3*, 1310. https://doi.org/10.1038/ncomms2306

Musto, H., Rodriguez-Maseda, H., & Bernardi, G. (1995). Compositional properties of nuclear genes from Plasmodium falciparum. *Gene*, *152*(1), 127–132.

N

Naganuma, M., Sekine, S. I., Fukunaga, R., & Yokoyama, S. (2009). Unique protein architecture of alanyl-tRNA synthetase for aminoacylation, editing, and dimerization. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(21), 8489–8494. https://doi.org/10.1073/pnas.0901572106

Nakanishi, K., Ogiso, Y., Nakama, T., Fukai, S., & Nureki, O. (2005). Structural basis for anticodon recognition by methionyl-tRNA synthetase. *Nature Structural and Molecular Biology*, *12*(10), 931–932. https://doi.org/10.1038/nsmb988

Negrutskii, B. S., Shalak, V. F., Kerjan, P., El'skaya, A. V., & Mirande, M. (1999). Functional interaction of mammalian valyl-tRNA synthetase with elongation factor EF-1α in the complex with EF-1H. *Journal of Biological Chemistry*, *274*(8), 4545–4550. https://doi.org/10.1074/jbc.274.8.4545

Ng, C. S., Sinha, A., Aniweh, Y., Nah, Q., Babu, I. R., Gu, C., Chionh, Y. H., Dedon, P. C., & Preiser, P. R. (2018). tRNA epitranscriptomics and biased codon are linked to proteome expression in Plasmodium falciparum. *Molecular Systems Biology*, *14*(10), e8009. https://doi.org/10.15252/msb.20178009

Nicholls, C., Li, H., & Liu, J. P. (2012). GAPDH: A common enzyme with uncommon functions. *Clinical and Experimental Pharmacology and Physiology*, *39*(8), 674–679. https://doi.org/10.1111/j.1440-1681.2011.05599.x

Nilsson Bark, S. K., Ahmad, R., Dantzler, K., Lukens, A. K., De Niz, M., Szucs, M. J., Jin, X., Cotton, J., Hoffmann, D., Bric-Furlong, E., Oomen, R., Parrington, M., Milner, D., Neafsey, D. E., Carr, S. A., Wirth, D. F., & Marti, M. (2018). Quantitative proteomic profiling reveals novel plasmodium falciparum surface antigens and possible vaccine candidates. *Molecular and Cellular Proteomics*, *17*(1), 43–60. https://doi.org/10.1074/mcp.RA117.000076

Norcum, M. T., & Boisset, N. (2002). Three-dimensional architecture of the eukaryotic multisynthetase complex determined from negatively stained and cryoelectron micrographs. *FEBS Letters*, *512*(1–3), 298–302. https://doi.org/10.1016/S0014-5793(02)02262-7

Novoa, E. M., Camacho, N., Tor, A., Wilkinson, B., Moss, S., Marín-García, P., Azcárate, I. G., Bautista, J. M., Mirando, A. C., Francklyn, C. S., Varon, S., Royo, M., Cortés, A., & De Pouplana, L. R. (2014). Analogs of natural aminoacyl-tRNA synthetase inhibitors clear malaria in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(51), E5508–E5517. https://doi.org/10.1073/pnas.1405994111

Nozawa, A., Ito, D., Ibrahim, M., Santos, H. J., Tsuboi, T., & Tozawa, Y. (2020). Characterization of mitochondrial carrier proteins of malaria parasite Plasmodium falciparum based on in vitro translation and reconstitution. *Parasitology International*, *79*(102160). https://doi.org/10.1016/j.parint.2020.102160

Nureki, O., Vassylyev, D. G., Katayanagi, K., Shimizu, T., Sekine, S. I., Kigawa, T., Miyazawa, T., Yokoyama, S., & Morikawa, K. (1995). Architectures of class-defining and specific domains of glutamyl-tRNA synthetase. *Science*, *267*(5206), 1958–1965. https://doi.org/10.1126/science.7701318

O

O'Donoghue, P., & Luthey-Schulten, Z. (2003). On the Evolution of Structure in Aminoacyl-tRNA Synthetases. *Microbiology and Molecular Biology Reviews*, *67*(4), 550–573. https://doi.org/10.1128/mmbr.67.4.550-573.2003

Ofir-Birin, Y., Fang, P., Bennett, S. P., Zhang, H. M., Wang, J., Rachmin, I., Shapiro, R., Song, J., Dagan, A., Pozo, J., Kim, S., Marshall, A. G., Schimmel, P., Yang, X. L., Nechushtan, H., Razin, E., & Guo, M. (2013). Structural Switch of Lysyl-tRNA Synthetase between Translation and Transcription. *Molecular Cell*, *49*(1), 30–42. https://doi.org/10.1016/j.molcel.2012.10.010

Oshikane, H., Sheppard, K., Fukai, S., Nakamura, Y., Ishitani, R., Numata, T., Sherrer, R. L., Feng, L., Schmitt, E., Panvert, M., Blanquet, S., Mechulam, Y., Söll, D., & Nureki, O. (2006). Structural basis of RNA-dependent recruitment of glutamine to the genetic code. *Science*, *312*(5782), 1950–1954. https://doi.org/10.1126/science.1128470

P

Pallarès, I., de Groot, N. S., Iglesias, V., Sant'Anna, R., Biosca, A., Fernàndez-Busquets, X., & Ventura, S. (2018). Discovering putative prion-like proteins in Plasmodium falciparum: A computational and experimental analysis. *Frontiers in Microbiology*, *9*(1737). https://doi.org/10.3389/fmicb.2018.01737

Panchal, M., Rawat, K., Kumar, G., Kibria, K. M., Singh, S., Kalamuddin, M., Mohmmed, A., Malhotra, P., & Tuteja, R. (2014). Plasmodium falciparum signal recognition particle components and anti-parasitic effect of ivermectin in blocking nucleo-cytoplasmic shuttling of SRP. *Cell Death and Disease*, *5*(1), e994. https://doi.org/10.1038/cddis.2013.521

Panda, S. K., & Mahapatra, R. K. (2017). In-silico screening, identification and validation of a novel vaccine candidate in the fight against Plasmodium falciparum. *Parasitology Research*, *116*(4), 1293–1305. https://doi.org/10.1007/s00436-017-5408-z

Pang, Y. L. J., Poruri, K., & Martinis, S. A. (2014). tRNA synthetase: TRNA aminoacylation and beyond. *Wiley Interdisciplinary Reviews: RNA*, *5*(4), 461–480. https://doi.org/10.1002/wrna.1224

Parhizgar, A. R., & Tahghighi, A. (2017). Introducing new antimalarial analogues of chloroquine and amodiaquine: A narrative review. *Iranian Journal of Medical Sciences*, *42*(2), 115–128.

Patankar, S., Narayan, A., Chaudhari, R., & Chakrabarti, R. (2013). Encyclopedia of Malaria. In M. Hommel & P. Kremsner (Eds.), *Encyclopedia of Malaria*. Springer New York. https://doi.org/10.1007/978-1-4614-8757-9

Penny, M. A., Camponovo, F., Chitnis, N., Smith, T. A., & Tanner, M. (2020). Future use-cases of vaccines in malaria control and elimination. *Parasite Epidemiology and Control*, *10*, e00145. https://doi.org/10.1016/j.parepi.2020.e00145

Perbandt, M., Eberle, R., Fischer-Riepe, L., Cang, H., Liebau, E., & Betzel, C. (2015). High resolution structures of Plasmodium falciparum GST complexes provide novel insights into the dimer-tetramer transition and a novel ligand-binding site. *Journal of Structural Biology*, *191*(3), 365–375. https://doi.org/10.1016/j.jsb.2015.06.008

Perona, J. J., & Hadd, A. (2012). Structural diversity and protein engineering of the aminoacyl-tRNA Synthetases. *Biochemistry*, *51*(44), 8705–8729. https://doi.org/10.1021/bi301180x

Petoukhov, M. V., & Svergun, D. I. (2015). Ambiguity assessment of small-angle scattering curves from monodisperse systems. *Acta Crystallographica Section D: Biological Crystallography*, *71*, 1051–1058. https://doi.org/10.1107/S1399004715002576

Pham, J. S., Sakaguchi, R., Yeoh, L. M., De Silva, N. S., Mcfadden, G. I., Hou, Y. M., & Ralph, S. A. (2014). A dual-targeted aminoacyl-tRNA synthetase in Plasmodium falciparum charges cytosolic and apicoplast tRNACys. *Biochemical Journal*, *458*(3), 513–523. https://doi.org/10.1042/BJ20131451

Piiadov, V., Ares de Araújo, E., Oliveira Neto, M., Craievich, A. F., & Polikarpov, I. (2019). SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. *Protein Science*, *28*(2), 454–463. https://doi.org/10.1002/pro.3528

Pizzi, E., & Frontali, C. (2001). Low-Complexity Regions in Plasmodium falciparum Proteins. *Genome Research*, *11*(2), 218–229. https://doi.org/10.1101/gr.gr-1522r

Provencher, S. W. (1982). Contin: A general purpose constrainded regularization program for inverting noisy linear algebraic and integral equations. *Computer Physics Communications*, *27*, 229–242. https://doi.org/10.1016/s0010-4655(84)82935-5

Prudêncio, M., Rodriguez, A., & Mota, M. M. (2006). The silent path to thousands of merozoites: The Plasmodium liver stage. *Nature Reviews Microbiology*, *4*(11), 849–856. https://doi.org/10.1038/nrmicro1529

Pütz, J., Giegé, R., & Florentz, C. (2010). Diversity and similarity in the tRNA world: Overall view and case study on malaria-related tRNAs. *FEBS Letters*, *584*(2), 350–358. https://doi.org/10.1016/j.febslet.2009.11.050

Q

Qin, X., Hao, Z., Tian, Q., Zhang, Z., Zhou, C., & Xie, W. (2014). Cocrystal structures of glycyl-tRNA synthetase in complex with tRNA suggest multiple conformational states in glycylation. *Journal of Biological Chemistry*, *289*(29), 20359–20369. https://doi.org/10.1074/jbc.M114.557249

Quevillon, S., Agou, F., Robinson, J. C., & Mirande, M. (1997). The p43 component of the mammalian multi-synthetase complex is likely to be the precursor of the endothelial monocyte-activating polypeptide II cytokine. *Journal of Biological Chemistry*, *272*(51), 32573–32579. https://doi.org/10.1074/jbc.272.51.32573

Quevillon, S., & Mirande, M. (1996). The p18 component of the multisynthetase complex shares a protein motif with the β and γ subunits of eukaryotic elongation factor 1. *FEBS Letters*, *395*(1), 63–67. https://doi.org/10.1016/0014-5793(96)01005-8

Quevillon, S., Robinson, J. C., Berthonneau, E., Siatecka, M., & Mirande, M. (1999). Macromolecular assemblage of aminoacyl-tRNA synthetases: Identification of protein-protein interactions and characterization of a core protein. *Journal of Molecular Biology*, *285*(1), 183–195. https://doi.org/10.1006/jmbi.1998.2316

R

Rajendran, V., Kalita, P., Shukla, H., Kumar, A., & Tripathi, T. (2018). Aminoacyl-tRNA synthetases: Structure, function, and drug discovery. *International Journal of Biological Macromolecules*, *111*, 400–414. https://doi.org/10.1016/j.ijbiomac.2017.12.157

Rambo, R. P., & Tainer, J. A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*, *496*(7446), 477–481. https://doi.org/10.1038/nature12070

Razakantoanina, V., Florent, I., & Jaureguiberry, G. (2008). Plasmodium falciparum: Functional mitochondrial ADP/ATP transporter in Escherichia coli plasmic membrane as a tool for selective drug screening. *Experimental Parasitology*, *118*(2), 181–187. https://doi.org/10.1016/j.exppara.2007.07.015

Regev-Rudzki, N., Wilson, D. W., Carvalho, T. G., Sisquella, X., Coleman, B. M., Rug, M., Bursac, D., Angrisano, F., Gee, M., Hill, A. F., Baum, J., & Cowman, A. F. (2013). Cell-cell communication between malaria-infected red blood cells via exosome-like vesicles. *Cell*, *153*(5), 1120–1133. https://doi.org/10.1016/j.cell.2013.04.029

Reiter, P. (2000). From Shakespeare to Defoe: Malaria in England in the Little Ice Age. *Emerging Infectious Diseases*, *6*(1), 1–11.

Renault, L., Kerjan, P., Pasqualato, S., Ménétrey, J., Robinson, J. C., Kawaguchi, S. I., Vassylyev, D. G., Yokoyama, S., Mirande, M., & Cherfils, J. (2001). Structure of the EMAPII domain of human aminoacyl-tRNA synthetase complex reveals evolutionary dimer mimicry. *EMBO Journal*, *20*(3), 570–578. https://doi.org/10.1093/emboj/20.3.570

Rénia, L., & Goh, Y. S. (2016). Malaria parasites: The great escape. *Frontiers in Immunology*, *7*(463). https://doi.org/10.3389/fimmu.2016.00463

Rho, S. B., Kim, M. J., Lee, J. S., Seol, W., Motegi, H., Kim, S., Shiba, K., Han, J. M., Ku, M. J., Lee, S. Y., & Kim, S. (1999). Genetic dissection of protein-protein interactions in multi-tRNA synthetase complex. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(8), 4488–4493. https://doi.org/10.1073/pnas.96.8.4488

Rho, S. B., Lee, J. S., Jeong, E. J., Kim, K. S., Kim, Y. G., & Kim, S. (1998). A multifunctional repeated motif is present in human bifunctional tRNA synthetase. *Journal of Biological Chemistry*, *273*(18), 11267–11273. https://doi.org/10.1074/jbc.273.18.11267

Rho, S. B., Lee, K. H., Kim, J. W., Shiba, K., Jo, Y. J., & Kim, S. (1996). Interaction between human tRNA synthetases involves repeated sequence elements. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(19), 10128–10133. https://doi.org/10.1073/pnas.93.19.10128

Ribas de Pouplana, L., & Schimmel, P. (2001). Two classes of tRNA synthetases suggested by sterically compatible dockings on tRNA acceptor stem. *Cell*, *104*(2), 191–193. https://doi.org/10.1016/S0092-8674(01)00204-5

Rich, S. M., Hudson, R. R., & Ayala, F. J. (1997). Plasmodium falciparum antigenic diversity: Evidence of clonal population structure. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(24), 13040–13045.

Ridley, R. G. (1991). Proteins of unusual sequence composition from the malarial parasite Plasmodium falciparum. *Biochemical Society Transactions*, *19*(2), 525–528. https://doi.org/10.1042/bst0190525

Robinson, J. C., Kerjan, P., & Mirande, M. (2000). Macromolecular assemblage of aminoacyl-tRNA synthetases: Quantitative analysis of protein-protein interactions and mechanism of complex assembly. *Journal of Molecular Biology*, *304*(5), 983–994. https://doi.org/10.1006/jmbi.2000.4242

Rocha, E. P. C. (2004). Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization. *Genome Research*, *14*(11), 2279–2286. https://doi.org/10.1101/gr.2896904

Rose, A., & Meier, I. (2004). Scaffolds, levers, rods and springs: Diverse cellular functions of long coiled-coil proteins. *Cellular and Molecular Life Sciences*, *61*(16), 1996–2009. https://doi.org/10.1007/s00018-004-4039-6

Rossmann, M. G., Moras, D., & Olsen, K. W. (1974). Chemical and biological evolution of a nucleotide-binding protein. *Nature*, *250*(463), 194–199. https://doi.org/10.1038/250194a0

Rould, M. A., Perona, J. J., Söll, D., & Steitz, T. A. (1989). Structure of E. coli glutaminyl-tRNA synthetase complexed with tRNA Gln and ATP at 2.8 Å resolution. *Science*, *246*(4934), 1135–1142. https://doi.org/10.1126/science.2479982

Rould, M. A., Perona, J. J., & Steitz, T. A. (1991). Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature*, *352*(6332), 213–218. https://doi.org/10.1038/352213a0

Roy, H., Ling, J., Irnov, M., & Ibba, M. (2004). Post-transfer editing in vitro and in vivo by the β subunit of phenylalanyl-tRNA synthetase. *EMBO Journal*, *23*(23), 4639–4648. https://doi.org/10.1038/sj.emboj.7600474

Rubinson, K. A., Ladner, J. E., Tordova, M., & Gilliland, G. L. (2000). Cryosalts: Suppression of ice formation in macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *56*(8), 996–1001. https://doi.org/10.1107/S0907444900007587

Rubio-Gomez, M. A., & Ibba, M. (2020). Aminoacyl-tRNA synthetases. *RNA*, *26*(8), 910–936. https://doi.org/10.1261/rna

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C., & Moras, D. (1991). Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science*, *252*(5013), 1682–1689. https://doi.org/10.1126/science.2047877

Ruíz López del Prado, G., Hernán García, C., Moreno Cea, L., Fernández Espinilla, V., Muñoz Moreno, M. F., Márquez, A., Delgado Polo, M. J. P., & Andrés García, I. (2014). Malaria in developing countries. *Journal of Infection in Developing Countries*, *8*(1), 1–4. https://doi.org/10.3855/jidc.4610

S

Sallares, R. (2002). *Malaria and Rome: A History of Malaria in Ancient Italy*. Oxford University Press.

Sang Lee, J., Gyu Park, S., Park, H., Seol, W., Lee, S., & Kim, S. (2002). Interaction network of human aminoacyl-tRNA synthetases and subunits of elongation factor 1 complex. *Biochemical and Biophysical Research Communications*, *291*(1), 158–164. https://doi.org/10.1006/bbrc.2002.6398

Sangolgi, P. B., Balaji, C., Dutta, S., Jindal, N., & Jarori, G. K. (2016). Cloning, expression, purification and characterization of Plasmodium spp. glyceraldehyde-3-phosphate dehydrogenase. *Protein Expression and Purification*, *117*, 17–25. https://doi.org/10.1016/j.pep.2015.08.028

Satchwell, T. J. (2016). Erythrocyte invasion receptors for Plasmodium falciparum: New and old. *Transfusion Medicine*, *26*, 77–88. https://doi.org/10.1111/tme.12280

Schmitt, E., Meinnel, T., Blanquet, S., & Mechulam, Y. (1994). Methionyl-tRNA Synthetase Needs an Intact and Mobile 332KMSKS336 Motif in Catalysis of Methionyl Adenylate Formation. *Journal of Molecular Biology*, *242*(4), 566–577.

Schmitt, E., Panvert, M., Blanquet, S., & Mechulam, Y. (1995). Transition state stabilization by the "high" motif of class I aminoacyl-tRNA synthetases: The case of Escherichia coli methionyl-tRNA synthetase. *Nucleic Acids Research*, *23*(23), 4793–4798. https://doi.org/10.1093/nar/23.23.4793

Schneider, A. (2011). Mitochondrial tRNA import and its consequences for mitochondrial translation. *Annual Review of Biochemistry*, *80*, 1033–1053. https://doi.org/10.1146/annurev-biochem-060109-092838

Sekine, S.-I., Nureki, O., Dubois, D. Y., Bernier, S., Chenevert, R., Lapointe, J., Vassylyev, D. G., & Yokoyama, S. (2003). ATP binding by glutamyl-tRNA synthetase is switched to the productive mode by tRNA binding. *EMBO Journal*, *22*(3), 676–688.

Senda, M., Hayashi, T., Hatakeyama, M., Takeuchi, K., Sasaki, A. T., & Senda, T. (2016). Use of Multiple Cryoprotectants to Improve Diffraction Quality from Protein Crystals. *Crystal Growth and Design*, *16*(3), 1565–1571. https://doi.org/10.1021/acs.cgd.5b01692

Sexton, A. E., Doerig, C., Creek, D. J., & Carvalho, T. G. (2019). Post-Genomic Approaches to Understanding Malaria Parasite Biology: Linking Genes to Biological Functions. *ACS Infectious Diseases*, *5*(8), 1269–1278. https://doi.org/10.1021/acsinfecdis.9b00093

Shah, S. (2010). *The Fever: How Malaria Has Ruled Humankind for 500,000 Years*. Sarah Crichton Books.

Shalak, V., Kaminska, M., Mitnacht-Kraus, R., Vandenabeele, P., Clauss, M., & Mirande, M. (2001). The EMAPII Cytokine Is Released from the Mammalian Multisynthetase Complex after Cleavage of Its p43/proEMAPII Component. *Journal of Biological Chemistry*, *276*(26), 23769–23776. https://doi.org/10.1074/jbc.M100489200

Sharma, A., & Sharma, A. (2015). Plasmodium falciparum mitochondria import tRNAs along with an active phenylalanyl-tRNA synthetase. *Biochemical Journal*, *465*(3), 459–469. https://doi.org/10.1042/BJ20140998

Shaw, P. J., Ponmee, N., Karoonuthaisiri, N., Kamchonwongpaisan, S., & Yuthavong, Y. (2007). Characterization of human malaria parasite Plasmodium falciparum eIF4E homologue and mRNA 5′ cap status. *Molecular and Biochemical Parasitology*, *155*(2), 146–155. https://doi.org/10.1016/j.molbiopara.2007.07.003

Shears, M. J., Botté, C. Y., & McFadden, G. I. (2015). Fatty acid metabolism in the Plasmodium apicoplast: Drugs, doubts and knockouts. *Molecular and Biochemical Parasitology*, *199*, 34–50. https://doi.org/10.1016/j.molbiopara.2015.03.004

Sheehan, D., Meade, G., Foley, V. M., & Dowd, C. A. (2001). Structure, function and evolution of glutathione transferases: Implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical Journal*, *360*(1), 1–16. https://doi.org/10.1042/0264-6021:3600001

Sherling, E. S., & van Ooij, C. (2016). Host cell remodeling by pathogens: The exomembrane system in Plasmodium-infected erythrocytes. *FEMS Microbiology Reviews*, *40*(5), 701–721. https://doi.org/10.1093/femsre/fuw016

Siatecka, M., Rozek, M., Barciszewski, J., & Mirande, M. (1998). Modular evolution of the Glx-tRNA synthetase family. Rooting of the evolutionary tree between the bacteria and archaea/eukarya branches. *European Journal of Biochemistry*, *256*(1), 80–87. https://doi.org/10.1046/j.1432-1327.1998.2560080.x

Siau, A., Huang, X., Weng, M., Sze, S. K., & Preiser, P. R. (2016). Proteome mapping of Plasmodium: Identification of the P. yoelii remodellome. *Scientific Reports*, *6*, 31055. https://doi.org/10.1038/srep31055

Simader, H., Hothorn, M., Köhler, C., Basquin, J., Simos, G., & Suck, D. (2006). Structural basis of yeast aminoacyl-tRNA synthetase complex formation revealed by crystal structures of two binary sub-complexes. *Nucleic Acids Research*, *34*(14), 3968–3979. https://doi.org/10.1093/nar/gkl560

Simader, H., Hothorn, M., & Suck, D. (2006). Structures of the interacting domains from yeast glutamyl-tRNA synthetase and tRNA-aminoacylation and nuclear-export cofactor Arc1p reveal a novel function for an old fold. *Acta Crystallographica Section D: Biological Crystallography*, *62*(12), 1510–1519. https://doi.org/10.1107/S0907444906039850

Simos, G., Sauer, A., Fasiolo, F., & Hurt, E. C. (1998). A conserved domain within Arc1p delivers tRNA to aminoacyl-tRNA synthetases. *Molecular Cell*, *1*(2), 235–242. https://doi.org/10.1016/S1097-2765(00)80024-6

Simos, G., Segref, A., Fasiolo, F., Hellmuth, K., Shevchenko, A., Mann, M., & Hurt, E. C. (1996). The yeast protein Arc1p binds to tRNA and functions as a cofactor for the methionyl-and glutamyl-tRNA synthetases. *The EMBO Journal*, *15*(19), 5437–5448.

Simpson, A. M., Suyama, Y., Dewcs, H., Campbell, D. A., & Simpson, L. (1989). Kinetoplastid mitochondria contain functional tRNAs which are encoded in nuclear DNA and also contain small minicircle and maxicircle transcripts of unknown function. *Nucleic Acids Research*, *17*(14), 5427–5446. https://doi.org/10.1093/nar/17.14.5427

Singer, G. A. C., & Hickey, D. A. (2000). Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. *Molecular Biology and Evolution*, *17*(11), 1581–1588. https://doi.org/10.1093/oxfordjournals.molbev.a026257

Singh, G. P., Chandra, B. R., Bhattacharya, A., Akhouri, R. R., Singh, S. K., & Sharma, A. (2004). Hyper-expansion of asparagines correlates with an abundance of proteins with prion-like domains in Plasmodium falciparum. *Molecular and Biochemical Parasitology*, *137*(2), 307–319. https://doi.org/10.1016/j.molbiopara.2004.05.016

Smith, D. W., & McNamara, A. L. (1972). The transfer RNA content of rabbit reticulocytes: enumeration of the individual species per cell. *Biochimica et Biophysica Acta*, *269*(1), 67–77.

Smith, J. D., Rowe, J. A., Higgins, M. K., & Lavstsen, T. (2013). Malaria's deadly grip: Cytoadhesion of Plasmodium falciparum-infected erythrocytes. *Cellular Microbiology*, *15*(12), 1976–1983. https://doi.org/10.1111/cmi.12183

Sonoiki, E., Palencia, A., Guo, D., Ahyong, V., Dong, C., Li, X., Hernandez, V. S., Zhang, Y. K., Choi, W., Gut, J., Legac, J., Cooper, R., Alley, M. R. K., Freund, Y. R., DeRisi, J., Cusack, S., & Rosenthal, P. J. (2016). Antimalarial benzoxaboroles target Plasmodium falciparum leucyl-tRNA synthetase. *Antimicrobial Agents and Chemotherapy*, *60*(8), 4886–4895. https://doi.org/10.1128/AAC.00820-16

Spielmann, T., Hawthorne, P. L., Dixon, M. W. A., Hannemann, M., Klotz, K., Kemp, D. J., Klonis, N., Tilley, L., Trenholme, K. R., & Gardiner, D. L. (2006). A Cluster of Ring Stage–specific Genes Linked to a Locus Implicated in Cytoadherence in Plasmodium falciparum Codes for PEXEL-negative and PEXEL-positive Proteins Exported into the Host Cell. *Molecular Biology of the Cell*, *17*(8), 3613–3624. https://doi.org/10.1091/mbc.E06

Spillman, N. J., Beck, J. R., Ganesan, S. M., Niles, J. C., & Goldberg, D. E. (2017). The chaperonin TRiC forms an oligomeric complex in the malaria parasite cytosol. *Cellular Microbiology*, *19*(6), e12719. https://doi.org/10.1111/cmi.12719

Sprinzl, M., & Cramer, F. (1975). Site of aminoacylation of tRNAs from Escherichia coli with respect to the 2' or 3' hydroxyl group of the terminal adenosine. *Proceedings of the National Academy of Sciences of the United States of America*, *72*(8), 3049–3053. https://doi.org/10.1073/pnas.72.8.3049

Staines, H. M., Moore, C. M., Slavic, K., & Krishna, S. (2017). Transmembrane solute transport in the apicomplexan parasite Plasmodium. *Emerging Topics in Life Sciences*, *1*(6), 553–561. https://doi.org/10.1042/etls20170097

Struhl, K. (1989). Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. *Trends in Biochemical Sciences*, *14*(4), 137–140. https://doi.org/10.1016/0968-0004(89)90145-X

Sugawara, A., Tanaka, T., Hirose, T., Ishiyama, A., Iwatsuki, M., Takahashi, Y., Otoguro, K., Omura, S., & Sunazuka, T. (2013). Borrelidin analogues with antimalarial activity: Design, synthesis and biological evaluation against Plasmodium falciparum parasites. *Bioorganic and Medicinal Chemistry Letters*, *23*(8), 2302–2305. https://doi.org/10.1016/j.bmcl.2013.02.075

Sugiura, I., Nureki, O., Ugaji-Yoshikawa, Y., Kuwabara, S., Shimada, A., Tateno, M., Lorber, B., Giegé, R., Moras, D., Yokoyama, S., & Konno, M. (2000). The 2.0 Å crystal structure of Thermus thermophilus methionyl-tRNA synthetase reveals two RNA-binding modules. *Structure*, *8*(2), 197–208.

Sun, M., Li, W., Blomqvist, K., Das, S., Hashem, Y., Dvorin, J. D., & Frank, J. (2015). Dynamical features of the Plasmodium falciparum ribosome during translation. *Nucleic Acids Research*, *43*(21), 10515–10524. https://doi.org/10.1093/nar/gkv991

Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography*, *25*, 495–503. https://doi.org/10.1107/S0021889892001663

Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal*, *76*(6), 2879–2886. https://doi.org/10.1016/S0006-3495(99)77443-6

Svergun, D. I. (2010). Small-angle X-ray and neutron scattering as a tool for structural systems biology. *Biological Chemistry*, *391*(7), 737–743. https://doi.org/10.1515/BC.2010.093

Swairjo, M. A., Morales, A. J., Wang, C. C., Ortiz, A. R., & Schimmel, P. (2000). Crystal structure of Trbp111: A structure-specific tRNA-binding protein. *EMBO Journal*, *19*(23), 6287–6298. https://doi.org/10.1093/emboj/19.23.6287

Swearingen, K. E., & Lindner, S. E. (2018). Plasmodium Parasites Viewed through Proteomics. *Trends in Parasitology*, *34*(11), 945–960. https://doi.org/10.1016/j.pt.2018.08.003

Swearingen, K. E., Lindner, S. E., Shi, L., Shears, M. J., Harupa, A., Hopp, C. S., Vaughan, A. M., Springer, T. A., Moritz, R. L., Kappe, S. H. I., & Sinnis, P. (2016). Interrogating the Plasmodium Sporozoite Surface: Identification of Surface-Exposed Proteins and Demonstration of Glycosylation on CSP and TRAP by Mass Spectrometry-Based Proteomics. *PLoS Pathogens*, *12*(4), e1005606. https://doi.org/10.1371/journal.ppat.1005606

T

Takebe, S., Witola, W. H., Schimanski, B., Günzl, A., & Ben Mamoun, C. (2007). Purification of components of the translation elongation factor complex of Plasmodium falciparum by tandem affinity purification. *Eukaryotic Cell*, *6*(4), 584–591. https://doi.org/10.1128/EC.00376-06

Terada, T., Nureki, O., Ishitani, R., Ambrogelly, A., Ibba, M., Söll, D., & Yokoyama, S. (2002). Functional convergence of two lysyl-tRNA synthetases with unrelated topologies. *Nature Structural Biology*, *9*(4), 257–262. https://doi.org/10.1038/nsb777

Tham, W. H., Beeson, J. G., & Rayner, J. C. (2017). Plasmodium vivax vaccine research – we've only just begun. *International Journal for Parasitology*, *47*(2–3), 111–118. https://doi.org/10.1016/j.ijpara.2016.09.006

Tournu, H., Butts, A., & Palmer, G. E. (2019). Titrating Gene Function in the Human Fungal Pathogen Candida albicans through Poly-Adenosine Tract Insertion. *MSphere*, *4*(3), e00192-19. https://doi.org/10.1128/msphere.00192-19

Tuteja, R., & Pradhan, A. (2009). Isolation and functional characterization of eIF4F components and poly(A)-binding protein from Plasmodium falciparum. *Parasitology International*, *58*(4), 481–485. https://doi.org/10.1016/j.parint.2009.09.001

Tuteja, R., & Pradhan, A. (2010). PfeIF4E and PfeIF4A colocalize and their double-stranded RNA inhibits Plasmodium falciparum proliferation . *Communicative & Integrative Biology*, *3*(6), 611–613. https://doi.org/10.4161/cib.3.6.13396

Tzima, E., Reader, J. S., Irani-Tehrani, M., Ewalt, K. L., Schwartz, M. A., & Schimmel, P. (2003). Biologically active fragment of a human tRNA synthetase inhibits fluid shear stress-activated responses of endothelial cells. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(25), 14903–14907. https://doi.org/10.1073/pnas.2436330100

V

Vaidya, A. B., Akella, R., & Suplick, K. (1989). Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite. *Molecular and Biochemical Parasitology*, *35*(2), 97–108.

Vaishya, S., Kumar, V., Gupta, A., Siddiqi, M. I., & Habib, S. (2016). Polypeptide release factors and stop codon recognition in the apicoplast and mitochondrion of Plasmodium falciparum. *Molecular Microbiology*, *100*(6), 1080–1095. https://doi.org/10.1111/mmi.13369

van Rooyen, J. M., Murat, J.-B., Hammoudi, P.-M., Kieffer-Jaquinod, S., Coute, Y., Sharma, A., Pelloux, H., Belrhali, H., & Hakimi, M.-A. (2014). Assembly of the novel five-component apicomplexan multi-aminoacyl-tRNA synthetase complex is driven by the hybrid scaffold protein Tg-p43. *PloS One*, *9*(2), e89487. https://doi.org/10.1371/journal.pone.0089487

Van Spaendonk, R. M. L., Ramesar, J., Van Wigcheren, A., Eling, W., Beetsma, A. L., Van Gemert, G. J., Hooghof, J., Janse, C. J., & Waters, A. P. (2001). Functional Equivalence of Structurally Distinct Ribosomes in the Malaria Parasite, Plasmodium berghei. *Journal of Biological Chemistry*, *276*(25), 22638–22647. https://doi.org/10.1074/jbc.M101234200

Vaughan, A. M., & Kappe, S. H. I. (2017). Malaria parasite liver infection and exoerythrocytic biology. *Cold Spring Harbor Perspectives in Medicine*, *7*, a025486. https://doi.org/10.1101/cshperspect.a025486

Vembar, S. S., Droll, D., & Scherf, A. (2016). Translational regulation in blood stages of the malaria parasite Plasmodium spp.: systems-wide studies pave the way. *Wiley Interdisciplinary Reviews: RNA*, *7*(6), 772–792. https://doi.org/10.1002/wrna.1365

Vera, L., & Stura, E. A. (2014). Strategies for protein cryocrystallography. *Crystal Growth and Design*, *14*(2), 427–435. https://doi.org/10.1021/cg301531f

Videvall, E. (2018). Plasmodium parasites of birds have the most AT-rich genes of eukaryotes. *Microbial Genomics*, *4*(2), e000150. https://doi.org/10.1099/mgen.0.000150

W

Wahlgren, M., Goel, S., & Akhouri, R. R. (2017). Variant surface antigens of Plasmodium falciparum and their roles in severe malaria. *Nature Reviews Microbiology*, *15*(8), 479–491. https://doi.org/10.1038/nrmicro.2017.47

Wakasugi, K., & Schimmel, P. (1999). Two Distinct Cytokines Released from Human Aminoacyl-tRNA Synthetase. *Science*, *284*(5411), 147–151.

Wakasugi, K., Slike, B. M., Hood, J., Otani, A., Ewalt, K. L., Friedlander, M., Cheresh, D. A., & Schimmel, P. (2002). A human aminoacyl-tRNA synthetase as a regulator of angiogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(1), 173–177. https://doi.org/10.1073/pnas.012602099

Wang, Z., Wei, C., Hao, X., Deng, W., Zhang, L., Wang, Z., & Wang, H. (2019). Genome-wide identification and characterization of transfer RNA-derived small RNAs in Plasmodium falciparum. *Parasites and Vectors*, *12*(1), 36. https://doi.org/10.1186/s13071-019-3301-6

Waters, A. P., Syin, C., & McCutchan, T. F. (1989). Developmental regulation of stage-specific ribosome populations in Plasmodium. *Nature*, *342*(6248), 438–440.

Weiss, G. E., Gilson, P. R., Taechalertpaisarn, T., Tham, W. H., de Jong, N. W. M., Harvey, K. L., Fowkes, F. J. I., Barlow, P. N., Rayner, J. C., Wright, G. J., Cowman, A. F., & Crabb, B. S. (2015). Revealing the Sequence and Resulting Cellular Morphology of Receptor-Ligand Interactions during Plasmodium falciparum Invasion of Erythrocytes. *PLoS Pathogens*, *11*(2), e1004670. https://doi.org/10.1371/journal.ppat.1004670

White, M. R., & Garcin, E. D. (2016). The sweet side of RNA regulation: glyceraldehyde-3-phosphate dehydrogenase as a noncanonical RNA-binding protein. *Interdiscip Rev RNA*, *7*(1), 53–70.

Wilson, I., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D. J., Moore, P. W., & Williamson, D. H. (1996). Complete Gene Map of the Plastid-like DNA of the Malaria Parasite Plasmodium falciparum. *J. Mol. Biol*, *261*(2), 155–172.

Wolf, Y. I., Aravind, L., Grishin, N. V, & Koonin, E. V. (1999). Evolution of Aminoacyl-tRNA synthetases-analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Research*, *9*(8), 689–710. https://doi.org/10.1101/gr.9.8.689

Wong, W., Bai, X. C., Brown, A., Fernandez, I. S., Hanssen, E., Condron, M., Tan, Y. H., Baum, J., & Scheres, S. H. W. (2014). Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *ELife*, *3*, e03080. https://doi.org/10.7554/eLife.03080

World Health Organization. (2018). World Malaria Report 2018. *Global Malaria Programme*, 22.

X

Xu, X., Shi, Y., Zhang, H. M., Swindell, E. C., Marshall, A. G., Guo, M., Kishi, S., & Yang, X. L. (2012). Unique domain appended to vertebrate tRNA synthetase is essential for vascular development. *Nature Communications*, *3*, 681. https://doi.org/10.1038/ncomms1686

Xue, H. Y., & Forsdyke, D. R. (2003). Low-complexity segments in Plasmodium falciparum proteins are primarily nucleic acid level adaptations. *Molecular and Biochemical Parasitology*, *128*(1), 21–32. https://doi.org/10.1016/S0166-6851(03)00039-2

Y

Yadavalli, S. S., & Ibba, M. (2012). Quality control in aminoacyl-tRNA synthesis: Its role in translational fidelity. *Advances in Protein Chemistry and Structural Biology*, *86*, 1–43. https://doi.org/10.1016/B978-0-12-386497-0.00001-3

Yaremchuk, A., Kriklivyi, I., Tukalo, M., & Cusack, S. (2002). Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. *The EMBO Journal*, *21*(14), 3829–3840.

Yeh, E., & DeRisi, J. L. (2011). Chemical rescue of malaria parasites lacking an apicoplast defines organelle function in blood-stage plasmodium falciparum. *PLoS Biology*, *9*(8), e1001138. https://doi.org/10.1371/journal.pbio.1001138

Yogavel, M., Chaturvedi, R., Babbar, P., Malhotra, N., Jain, V., & Sharma, A. (2018). Drug targeting of one or more aminoacyl-tRNA synthetase in the malaria parasite Plasmodium falciparum. *Drug Discovery Today*, *23*(6), 1233–1240. https://doi.org/10.1016/j.drudis.2018.01.050

Yu, C. H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M. S., & Liu, Y. (2015). Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Molecular Cell*, *59*(5), 744–754. https://doi.org/10.1016/j.molcel.2015.07.018

Z

Zhao, H., Brown, P. H., & Schuck, P. (2011). On the distribution of protein refractive index increments. *Biophysical Journal*, *100*(9), 2309–2317. https://doi.org/10.1016/j.bpj.2011.03.004

Zhu, J., Waters, A. P., Appiah, A., McCutchan, T. F., Lal, A. A., & Hollingdale, M. R. (1990). Stage-specific ribosomal RNA expression switches during sporozoite invasion of hepatocytes. *Journal of Biological Chemistry*, *265*(21), 12740–12744.

Zilversmit, M. M., Volkman, S. K., Depristo, M. A., Wirth, D. F., Awadalla, P., & Hartl, D. L. (2010). Low-complexity regions in plasmodium falciparum: Missing links in the evolution of an extreme genome. *Molecular Biology and Evolution*, *27*(9), 2198–2209. https://doi.org/10.1093/molbev/msq108

Zimm, B. H. (1948). The scattering of light and the radial distribution function of high polymer solutions. *J. Chem. Phys*, *16*(12), 1093–1099.

Etude Biochimique et Structurale des Complexes MARS de *Plasmodium*

Plasmodium est un parasite protozoaire qui provoque le paludisme chez l'Homme et d'autres animaux vertébrés. L'équipe a mis en évidence, *in vitro,* l'internalisation d'ARNt exogènes dans le parasite et a identifié un transporteur éventuel; il s'agit de la protéine tRip (tRNA import protein). *In vitro,* tRip lie tous les ARNt en reconnaissant leur structure tridimensionnelle*. In vivo,* tRip est une protéine intégrale de la membrane et son domaine de liaison aux ARNt (tRBD) est exposé à l'extérieur du parasite. Bien que l'absence de tRip ne soit pas létale, la multiplication du parasite tRip-KO et sa synthèse protéique sont considérablement ralenties. En plus de son implication potentielle dans l'import des ARNt exogènes, tRip interagit avec trois aminoacyl-ARNt synthétases (aaRS) du parasite : la glutamyl- (ERS), la glutaminyl- (QRS) et la methionyl-ARNt synthétase (MRS), suggérant que tRip permettrait l'organisation d'un complexe Multi-Aminoacyl-ARNt Synthétasique (MARS) localisé à la membrane du parasite.

Au cours de ma thèse, j'ai utilisé des approches biochimiques et structurales pour caractériser le complexe MARS de *Plasmodium berghei in vitro*. En étudiant l'architecture des différentes protéines qui le constituent par analyse bioinformatique, j'ai mis en évidence qu'elles contiennent toutes un domaine GST N-terminal. J'ai montré que ces domaines GST sont suffisants pour reconstituer les interactions entre protéines. J'ai ainsi pu identifier les réseaux d'interaction entre les différents domaines GST ce qui m'a permis d'identifier et reconstituer 2 complexes distincts et homogènes : tRip:ERS:QRS (complexe Q) et tRip:ERS:MRS (complexe M). Chaque complexe a été caractérisé par des approches biophysiques avec comme but ultime de déterminer leur structure tridimensionnelle. Grâce à la résolution de la structure cristallographique du domaine GST de la ERS, j'ai pu tester les interfaces d'interaction par mutagenèse dirigée et ainsi proposer un modèle de la structure quaternaire des complexes MARS de *Plasmodium.* De façon intéressante, ces complexes homodimérisent grâce à tRip et sont donc constitués de deux molécules de tRip, deux molécules de ERS, et soit deux molécules de QRS soit deux molécules de MRS. La géométrie des interfaces d'interaction est compatible avec la localisation membranaire de tRip et la présence de domaines de liaison à l'ARNt additionnels aux extrémités C-terminales de la QRS et de la MRS compenseraient la localisation externe du tRBD de tRip.

José Refugio JARAMILLO PONCE

Etude biochimique et structurale des complexes MARS de *Plasmodium*

Résumé

Plasmodium est le parasite qui cause le paludisme. L'équipe a mis en évidence, *in vitro*, l'import d'ARNt exogène et a identifié un transporteur potentiel : la protéine tRip (tRNA import protein). *In vitro*, tRip lie tous les ARNt en reconnaissant leur structure. *In vivo*, tRip est une protéine transmembranaire et son domaine de liaison aux ARNt est exposé à l'extérieur du parasite. tRip n'est pas essentielle, mais la multiplication du parasite KO est significativement ralentie. En plus de son rôle dans l'import d'ARNt, tRip interagit avec 3 aminoacyl-ARNt synthétases : la glutamyl- (ERS), la glutaminyl- (QRS) et la methionyl-ARNt synthétase (MRS), suggérant que tRip permet l'organisation d'un complexe multi synthétasique (MARS) localisé à la membrane. J'ai pu montrer que l'assemblage du complexe est effectué par des domaines GST N-terminales présents dans les 4 partenaires. J'ai identifié, reconstitué et caractérisé 2 complexes distincts : complexe Q (tRip:ERS:QRS) et complexe M (tRip:ERS:MRS). La résolution de la structure cristallographique du GST de l'ERS ainsi que des tests de mutagenèse dirigée m'a permis de proposer des modèles des complexes MARS.

Mot clés : *Plasmodium*, aminoacyl-ARNt synthétases, complexes MARS.

Résumé en anglais

Plasmodium is a genus of protozoan parasites causing malaria. The team demonstrated, *in vitro*, the import of exogenous tRNA and identified a potential transporter: tRip (tRNA import protein). *In vitro*, tRip binds all tRNAs by recognizing their structure. *In vivo*, tRip is transmembrane protein and its tRNA binding domain is exposed outside the parasite. tRip is not essential, but its deletion slows down parasite multiplication and protein synthesis. In addition to its role in tRNA import, tRip interacts with 3 aminoacyl-ARNt synthetases: glutamyl- (ERS), glutaminyl- (QRS) and methionyl-ARNt synthetase (MRS), suggesting that tRip allows the organization of a multi tRNA synthetase complex (MARS) localized at the membrane. During my thesis, I demonstrated that assembly of the complex is mediated by GST domains appended to the N-terminus of each partner. I identified, reconstituted and characterize in solution 2 distinct complexes: complex Q (tRip:ERS:QRS) and complex M (tRip:ERS:MRS). The resolution of the crystal structure of ERS GST domain as well as point mutation experiments allowed me to propose a model for these complexes MARS.

Keywords: *Plasmodium*, aminoacyl-tRNA synthetases, MARS complexes