Targeting trypanosomatid translation factor interactions

Fábio Henrique dos Santos Rodrigues, M.Sc.

A thesis submitted for the degree of
Doctor of Philosophy

School of Life Sciences
University of Warwick

Supervisor: Prof. John E. G. McCarthy

November 2018

Declaration

I declare that I personally carried out the work submitted in this thesis, to the University of Warwick, in support of my application for the degree of Doctor of Philosophy. No part of this work has previously been submitted in any previous application for a degree.

The research presented was carried out by myself except in the cases outlined below:

PABP1 and its domains production and purification were performed by Dr Maja Firczuk.

Crystallography experiments were performed by Dr Maja Firczuk, in collaboration with Prof Alex Cameron.

The PAM2 peptide used in the NMR experiments were synthesized by the group of Dr Andrew Wilson, from the University of Leeds.

Part of this thesis was submitted for publication; however, the outcome of this application was not yet been received.

Acknowledgements

First of all, I would like to thank God for helping me to achieve this goal.

I would like to express my gratitude to my supervisor, Prof. John McCarthy, for the opportunity to perform my research project in his group, and all the support.
I would like to thank Dr Maja Firczuk for co-supervising me during my whole PhD, teaching me, giving advices during the everyday work, helping with experiments (especially providing proteins and performing the crystallographic work), your help was essential for me to complete this work.
I would like also to thank my current and former lab colleagues, Tailise, Xiang, Estelle, Gurd, Paola, Ollie, Byron, Andrew, you all are part of it, thanks for your help and friendship.

Also, thanks to my advisory panel, Dr Orkun Soyer and Dr Robin Allaby, for your help and suggestions along the way.

I would also like to express my gratitude to Dr Alex Cameron, whose help was essential in this work, especially in the crystallographic experiments. Also, thanks to Dr Alex Breeze, Dr Arnoult Kalverda and Dr Simon Skinner, that provided the conditions for the NMR experiments, a fundamental part of this work. Thanks to Dr Andy Wilson and Dr Martin Walko, that provided peptides for our experiments.

I also would like to thank Dr Cleidiane Zampronio, for all the help with the MS analysis, and for the good chat between incubation times. Thanks too to Dr Liz Fullam and Dr Sarah Bennet, that were always happy to help when needed.

I would like to thank my wife Tailise. This thesis is also yours, thanks for all the support, in the lab and at home, all the love and for being always by my side. Without you, this thesis would never have happened. Love you! Te Amo!
I would like to thank my family, my parents, Gisberto and Márcia, my grandma Darcy, my mother-in-law Elisabeth, without your love and support, I would never make it. Love you all.

Finally, thanks to University of Warwick, School of Life Sciences, and my funding agency, CNPq- Brazil.

Abstract

Trypanosomatids are eukaryotic parasites that migrate between insect vectors and mammalian hosts. They cause a number of serious diseases with major impacts on human health on a global scale. One example is *Leishmania*, which causes Leishmaniasis, a disease that affects approximately 12 million people in Asia, Middle East, Africa, South America and Southern Europe. Trans-splicing of trypanosomatid polycistronic transcripts produces polyadenylated monocistronic mRNAs that are modified at their 5'ends to form the cap4 structure (m7Gpppm36,6,2'Apm2'Apm2'Cpm23,2'U). Two trypanosomatid translation initiation factor 4E isomers (eIF4E3 and eIF4E4) have extended N-terminal regions that include PAM2 motifs; these regions can interact with PABC domains in the poly(A)-binding proteins PABP1.

Using NMR, X-ray crystallography and biophysical quantitation of biomolecular interactions, we show that tight binding between *Leishmania* eIF4E4 and PABP1, mediated largely by hydrophobic interactions, underpins a novel architecture in the translation factor complex bridging the 5' and 3' ends of the monocistronic mRNAs. We demonstrate that this complex can also accommodate *Leishmania* eIF4G3 which, in contrast to the standard eukaryotic initiation complex paradigm, is found to bind to eIF4E4, but not to PABP1. Exceptionally, therefore, by binding to two protein ligands and to the mRNA cap4 structure, the trypanosomatid N-terminally extended form of eIF4E acts as a core molecular scaffold for the mRNA-cap-binding complex. Finally, our data indicate that the eIF4E4 N-terminal extension is an intrinsically disordered region that transitions to a partly folded form upon binding to PABP1.

Abbreviations

a.a.	Amino acid
40S	Small ribosome subunit 40S
Amp	Ampicillin
bp	Base Pairs
BP	Branching Point
BSA	Bovine Serum Albumin
$CaCl_2$	Calcium chloride
CL	Cutaneous Leishmaniasis
$CoCl_2$	Cobalt chloride
$CuSO_4$	Copper sulphate
CV	Column Volume
ddH_2O	Double-distilled water
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dTTP	Deoxythymidine triphosphate
D_2O	Deuterium oxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
eIF4E	Eukaryotic Translation Initiation Factor 4E
eIF4G	Eukaryotic Translation Initiation Factor 4G
ESI	Electrospray Ionization
$FeSO_4$	Ferrous sulphate
FL	Full length
GST	Glutathione S-transferase tag

H_3BO_3	Boric acid
HCD	Higher-energy Collisional Dissociation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSQC	Heteronuclear Single Quantum Coherence
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Infrared
ITC	Isothermal Titration Calorimetry
$_k$DNA	Kinetoplast mitochondrial DNA
k_a	Association rate constant
k_d	Dissociation rate constant
K_A	Association constant
K_D	Dissociation constant
Kan	Kanamycin
KCl	Potassium chloride
KH_2PO_4	Potassium dihydrogen phosphate
LB	Lysogeny Broth
LC	Liquid Chromatography
mM	mili molar
mAU	Milli Absorbance Units
mRNA	Messenger RNA
M	Molar, *i.e.,* mol/L
MBP	Maltose Binding Protein tag
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulphate
$MnCl_2$	Manganese chloride
MCL	Muco-cutaneous Leishmaniasis
MS	Mass Spectrometry
MST	Microscale Thermophoresis
MYA	Millions of Years Ago

nt	nucleotide
NaCl	Sodium chloride
Na_2HPO_4	Disodium hydrogen phosphate
NAD	Nicotinamide adenine dinucleotide
$^{15}NH_4Cl$	^{15}N ammonium chloride
$(NH_4)_6(Mo_7)$	Ammonium molybdate
NMR	Nuclear Magnetic Resonance
NTA	Nitrilotriacetic acid
OD_{600}	Optical Density at 600 nm
pH	$\log_{10}[H^+]$
PABC	C-terminal PABC domain
PABP	Poly(A) Binding Protein
PAM2	PABP interaction motif
PEG8000	Polyethylene glycol 8000
PIC	Pre-initiation Complex
PMSF	Phenylmethylsulfonyl fluoride
rpm	rotations per minute
R_{max}	Maximum response
RNA	Ribonucleic acid
RRM	RNA Recognition Motif
SA	Streptavidin
SBP	Streptavidin Binding Protein
SDS-PAGE	Sodium Dodecyl Sulphate - Polyacrylamide Gel Electrophoresis
SL	Splice Leader
SL RNA	Splice Leader RNA
SPR	Surface Plasmon Resonance
tRNA	transfer RNA
TB	Terrific Broth

Th	Thomson (unit of mass-to-charge ratio)
Tris.HCl	Tris(hydroxymethyl)aminomethane Hydrochloride
TROSY	Transverse Relaxation-Optimized Spectroscopy
UTR	RNA Untranslated Region
V	volts
VL	Visceral Leishmaniasis
WHO	World Health Organization
ZnSO4	Zinc sulphate
µM	micro molar
°C	Degree Celsius
µL	microlitre

Amino acid abbreviations:

Amino acid	**3 letter code**	**1 letter code**	**Amino acid**	**3 letter code**	**1 letter code**
Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartate	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamate	Glu	E	Serine	Ser	S
Glutamine	Gln	Q	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

List of Figures

List of Tables

Chapter 1 – Trypanosomatids, Leishmaniasis and the Translation Initiation Complex

1.1: Trypanosomatids and Leishmaniasis

Leishmaniasis

Leishmaniasis is a neglected tropical disease (NTD) that affects a large number of countries around the globe, the exceptions being Australia and Antarctica (PACE, 2014). The disease is concentrated in countries in Southeast Asia, East Africa, the Middle East and Latin America. Leishmaniasis is also found in several Mediterranean countries in Southern Europe, like France, Italy and Spain (READY, 2010).

According to the World Health Organization, there are over 1 billion people living at endemic areas at risk of infection, with 300,000 new cases of visceral leishmaniasis every year, and more than 1 million cases of cutaneous leishmaniasis reported in the last 5 years - the number of deaths exceed 20,000 every year (WHO Weekly epidemiological record, 2016). The large number of people affected by leishmaniasis makes its study of great importance, especially studies whose outcome may help in the relief of the burden caused by the disease.

There are around 53 known *Leishmania* species, from which 20 are pathogenic to humans. The disease is transmitted by the bite of the female sandfly from the genera *Phlebotomus*, in the Old World, and *Lutzomyia* in the New World. 93 of the 800 sandfly species transmit

leishmaniasis (PACE, 2014). Among the several *Leishmania* species, the most common are *L. major, L. donovani, L. infantum and L. braziliensis.* Fossil evidence suggests that the first *Leishmania* appeared during the Cretaceous (146-65 MYA), and had reptiles as its main targets. The first evidence of humans infected by *Leishmania* dates from 2000 BC, being found in remains of Egyptian mummies; in the new world, archaeologists found a 6-year-old girl Peruvian mummy, dated 800 BC, infected with leishmaniasis (AKHOUNDI, 2016).

There are 3 major types of leishmaniasis: Cutaneous, Muco-cutaneous and Visceral (Kala-Azar).

1.1.1: Cutaneous leishmaniasis - CL

Cutaneous leishmaniasis (CL) is the most common manifestation of the disease. The disease has a varied incubation period, ranging from weeks to months. It has as characteristic the presence of ulcers on exposed areas of arms, legs, torso, and often appear in large numbers (BARRETT, 2012). These start as papules, that after a developing time, ulcerate. In some cases, the disease can last for more than two years (DOWLATI, 1996), and even after initial cure, relapsing can occur (VALESKY, 2007). The ulcers caused by CL leave severe scars after treated, causing lasting effects in the individuals afflicted by the disease, even after treatment.

Figure 1 - Child affected by cutaneous leishmaniasis. Large lesions on the cheek and near the nose and eyes. (Photo from WHO website, http://www.who.int/leishmaniasis/en/, C. Black).

1.1.2: Muco-cutaneous leishmaniasis - MCL

Muco-cutaneous (or mucous) leishmaniasis (MCL) is a disfiguring manifestation, and its lesions can partially or totally destroy the mucous membranes of the nose, mouth and throat cavities - it has great impact on the lives of those affected by MCL (PURSE, 2017). The disease starts when the upper respiratory tract is infected, displaying cutaneous lesions. As the infection aggravates, it damages the mucosal tract from oral and nasal cavities, even leading to nasal septum perforation and profound disfiguration (ROYER, 2002).

Figure 2 - Child affected by muco-cutaneous leishmaniasis. Note the severe damage in the nose/ upper lip region. (Photo from WHO website, http://www.who.int/leishmaniasis/en/).

1.1.3: Visceral leishmaniasis - VL

Visceral or Kala-Azar leishmaniasis (VL), whose symptoms are high fever, weight loss, swelling of the spleen and liver, and anaemia. VL is the most dangerous and potentially fatal variation of leishmaniasis. If left untreated, the disease can have a fatality rate of 100% within 2 years (PACE, 2014; COLLIN, 2004). The most characteristic symptoms of VL are hepato and splenomegalia, accompanied by weight loss, fever and weakness. The disease occurs when the parasites invade the reticulo-endothelial system, infecting the internal organs cells. (HERWALDT, 1999; CHAPPUIS, 2007).

Figure 3 - Child affected by visceral leishmaniasis. Both spleen and liver are massively enlarged, as outlined by green markings on the picture. (Photo from WHO website, http://www.who.int/leishmaniasis/en/).

1.1.4: Epidemiology

According to the World Health Organization, in its 2015 report, Leishmaniasis is present in a large number of countries, especially in the inter-tropical area.

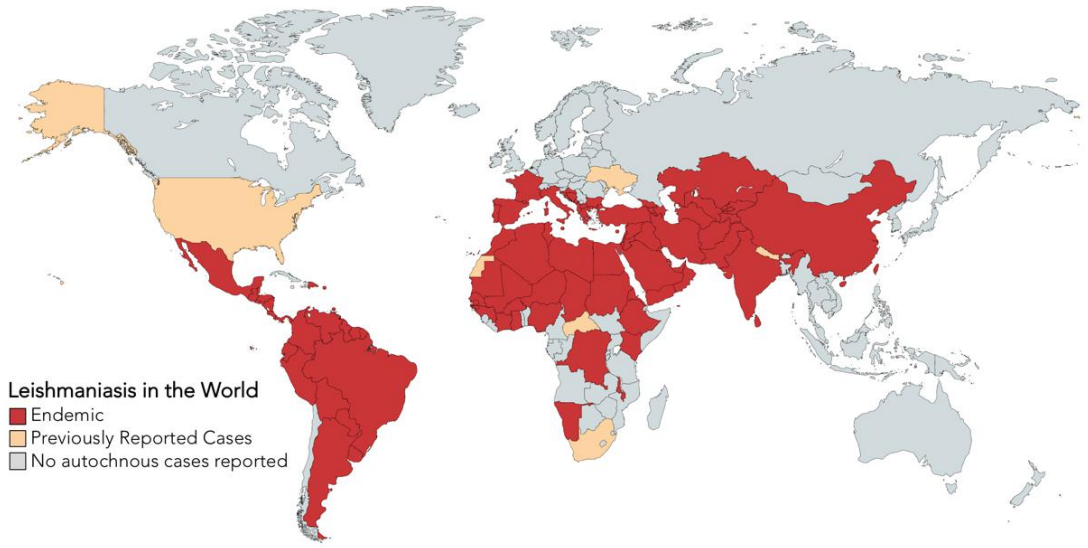


Figure 4: Leishmania epidemic map, endemic areas are highlighted in red. Data from WHO 2015 report.

Leishmaniasis has been identified as a growing concern for international travel. Studies have revealed that (MALTEZOU, 2010) there is an increasing number of people from non-endemic areas being diagnosed with Leishmaniasis. This group contains not only tourists and individuals travelling at work (MANSUETO, 2014), but also military

personnel deployed in endemic areas or that act in areas in combat operations. There are reports from military personnel from the UK armed forces and US army based in countries as Iraq, Belize, Panama and Afghanistan that were infected with Leishmaniasis. (PREEDY, 2014; ROYER, 2002). Leishmaniasis is one of the main parasitic diseases, and its mortality rate is only exceeded by malaria (WHO Technical Report, 2010).

1.1.5: Sandflies

All types of leishmaniasis are transmitted through the bite of the female sandfly. There are two sandfly genera that transmit the disease: *Phlebotomus* in the Old World, and *Lutzomyia* in the New World. According to (PACE, 2014), 93 of the around 800 known sandfly subspecies are able to spread leishmaniasis.

Sandfly is an insect that is mostly active at night, in open areas (although there are reports of some species that do bite during daytime, indoors). It has limited mobility, and its ability to travel vertically has a range of around 1.0 metres, and travel horizontally 1500 metres (AKHOUNDI, 2016). The first sandlies are believed to have originated during the Eocene (55-38 MYA), as *Phlebotomus*, while the separation into *Phlebotomus*/*Lutzomyia* happened during the Oligocene (38-25 MYA) (AKHOUNDI, 2016).

1.1.6: *Leishmania*

The *Leishmania* parasite was first described in 1900 by the Scottish pathologist William Boog Leishman, that was treating British soldiers in India. Not long after, the Irish doctor Charles Donovan also identified the disease independently. Their findings led to the discovery of the parasite that causes leishmaniasis. (STEVERDING, 2017).

(A) (B)

Figure 5: (A) Lieutenant General Sir William Boog Leishman; (B) Major Charles Donovan. (STEVERDING, 2017).

The parasite *Leishmania* belongs to the Class Kinetoplastida, Order Trypanossomatida, Genus *Leishmania*. It has two main morphologies during its lifecycle; a flagellated promastigote and a non-flagellated amastigote form. The amastigote cell morphology is smaller, spherical, with limited motility. The other cell morphology, promastigote, is larger, more elongated, and with a long flagellum, that confers mobility to the parasite (SUNTER, 2017).

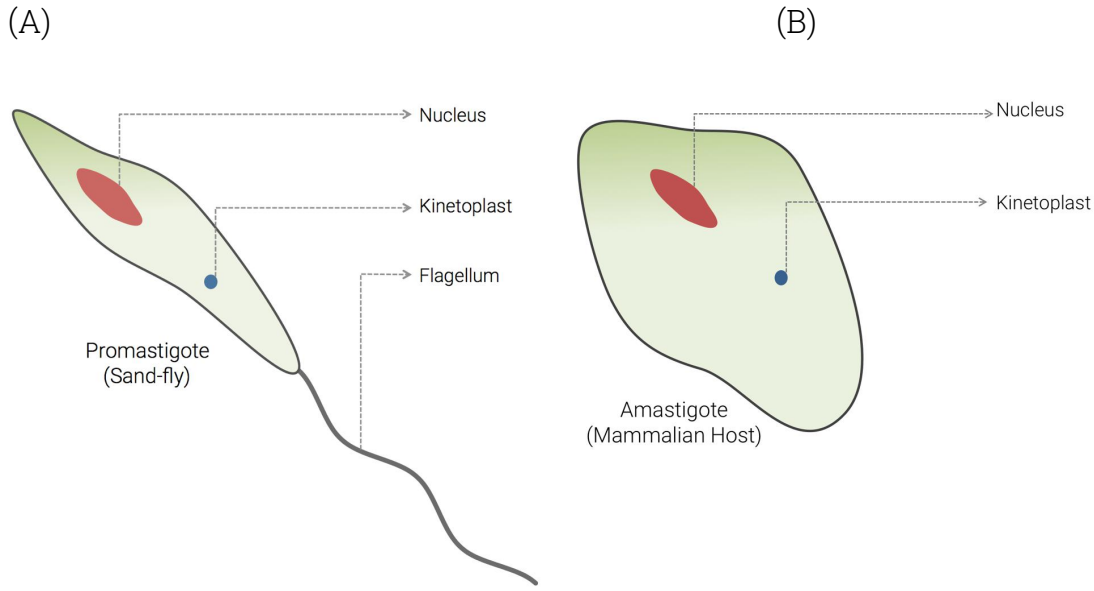


Figure 6: Leishmania main morphologies (A) Promastigote (insect); (B) amastigote (mammalian host).

The parasite promastigote, flagellated form is found in the sandfly stage of the disease. During this period, the parasite develops and multiply in the insect gut, and after the sandfly bites the mammal host, and infects it. After the host gets infected, the parasite transforms into its amastigote form, living mainly in phagocytes of the reticulo-endothelial system. The amastigotes then multiple and infect other cells, spreading the infection. If the host is bitten by another sandfly, then after the blood meal the insect ingests the amastigotes, that then develop into promastigotes in the sandfly gut, re-starting the infection cycle (PODINOVSKAIA, 2015; CARDOSO, 2015).

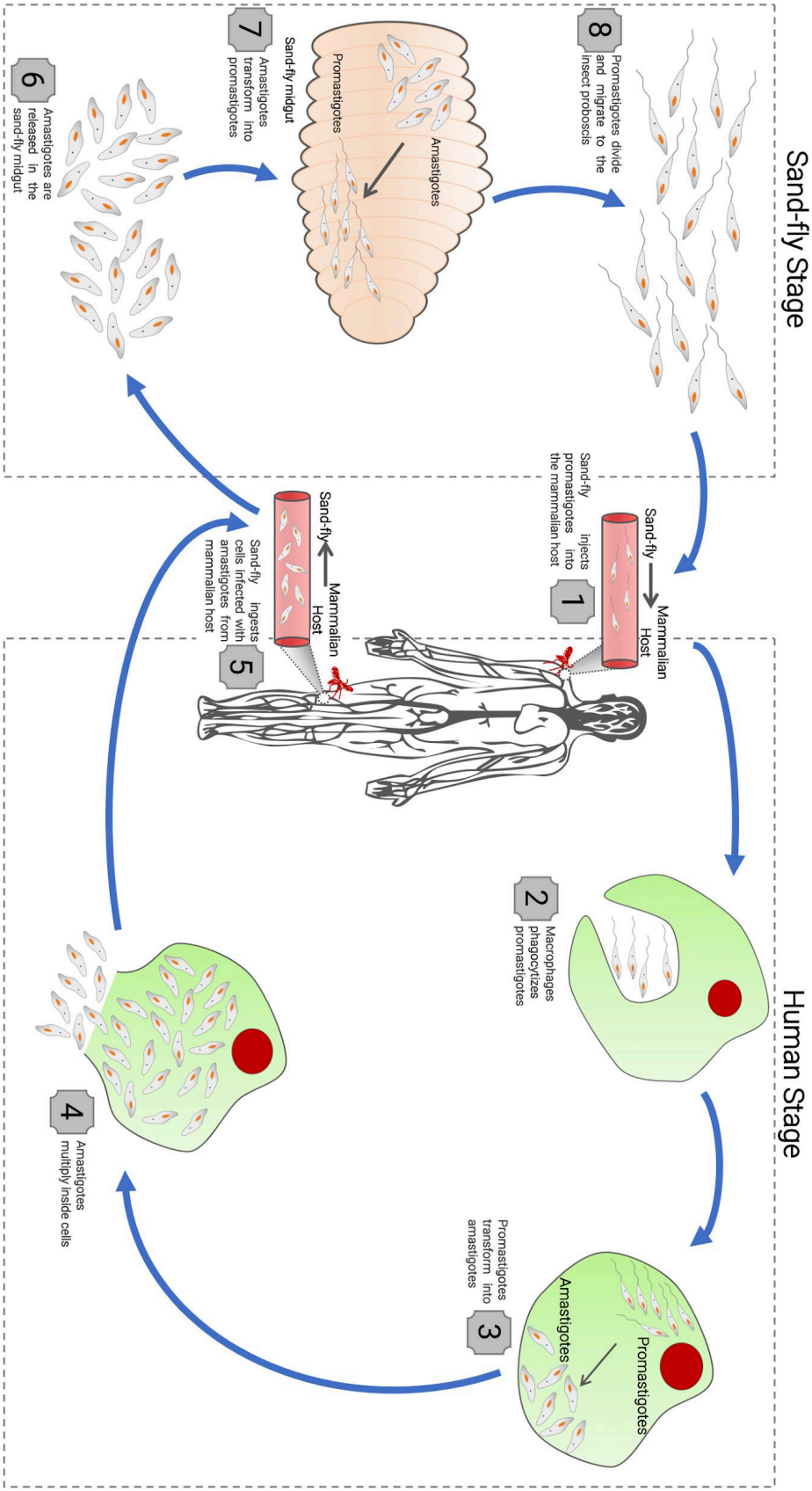


Figure 7: Leishmania life cycle. (Adapted from the American Center for Disease and Prevention website).

The Kinetoplastida class has a unique organelle called kinetoplast. Kinetoplasts are condensed disks of mitochondrial DNA (kDNA), organized in large circles (maxicircles - approximately 20 kb) and small circles (minicircles - approximately 1 kb) located in the mitochondria, near the flagellum base (POVELONES, 2014). There are around 25-50 maxicircles and 5000-10000 minicircles, interlocked together in a unique disk-shaped structure forming a single kDNA network (VARGAS-PARADA, 2010). The maxicircles encode genes that correspond to other eukaryotic mitochondrial DNA, like subunits of respiratory complex for example. However, the maxicircle transcripts are cryptic and require further RNA editing, and the minicircles are responsible for controlling RNAs that will act in the RNA editing, "correcting" the maxicircle transcripts (LIU, 2005).

1.1.7: Nutrition, available treatments, side effects and drug resistance

Several studies suggest that individual's nutrition plays an important role in the occurrence or resistance to leishmaniasis. Malnutrition can lead to weakened immune system, putting people in endemic areas in higher risk to display the disease (PEREZ, 1979; DIRO, 2015). For instance, studies showed that malnutritioned mice showed higher rates of severe VL, lower leukocyte count, lower immunological response, suggesting the importance of nutrition on resistance to leishmaniasis (CARRILLO, 2014). After infection occurs and the clinical symptoms takes place, a number of drug treatments are used. The most common drugs used in leishmaniasis treatment are Amphotericin B, Miltefosine, Paromomycin and Sodium Stibogluconate (LINDOSO, 2012; COPELAND, 2015), although other options are also available (CARDOSO, 2014) (Figure 8):

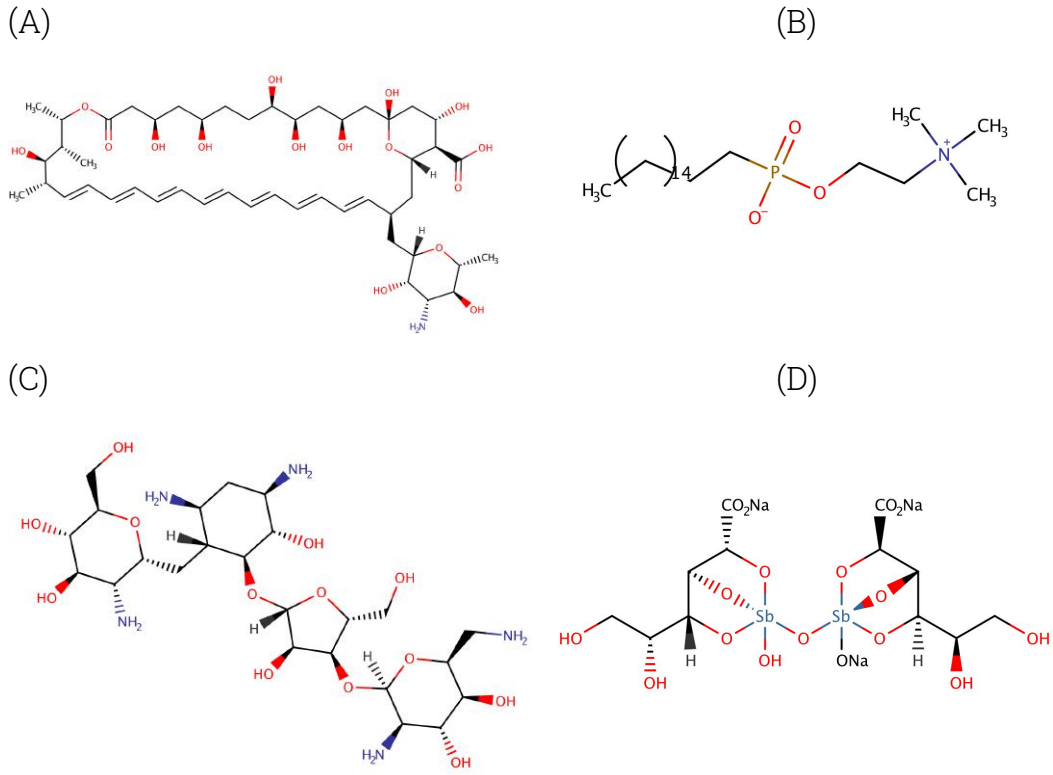


Figure 8: Most common drugs used in leishmaniasis treatment, (A) Amphotericin B; (B) Miltefosine; (C) Paromomycin and (D) Sodium Stibogluconate.

Amphotericin B is an antibiotic that is capable of killing *Leishmania* parasites, acting on the parasite's surface membrane. However, the drug is known to cause kidney damage and allergic reactions (WHO Frequently asked questions on visceral leishmaniasis kala-azar, 2013). Miltefosine was initially developed as an anti-cancer drug, and posteriorly tested as anti-*Leishmania* drug (TIUMAN, 2011). It needs a long treatment, and cannot be used during pregnancy, since it can cause fetus deformity. It can also lead to gastrointestinal problems, and its long half-life may lead to increased parasite resistance to the drug (WHO Frequently asked questions on visceral leishmaniasis kala-azar, 2013).

Paromomycin is an aminoglycoside antibiotic with anti-Leishmanial activity. It has poor oral absorption, and for this reason, requires deep intramuscular injections, daily for at least 21 days (TIUMAN, 2011). Sodium

stibugluconate is the most traditional drug for leishmaniasis treatment. Its mechanism of action is not completely understood, but there is evidence that it acts in the parasite's oxidative pathways (LINDOSO, 2012). It is a significantly toxic drug (toxic to liver, kidneys), being heart problems and acute pancreatitis particularly dangerous side-effects (TIUMAN, 2011; LINDOSO, 2012).

Among the several side effects related to these drugs, the most serious ones are decreased kidney function, liver problems, permanent hearing loss, pancreatitis, veins toxicity and cardiac failure. Also, the administration of these drugs is often unpleasant, like for example through the application very painful intramuscular injections, or through intravenous administration that requires hospitalization. For instance, sodium stibugluconate and miltefosine requires daily injection for at least 28 days, and hospitalization (WHO Frequently asked questions on visceral leishmaniasis kala-azar, 2013).

Table 1: Pros and cons of the main drugs used in leishmaniasis treatment.

Drug	Pros	Cons
Amphotericin B	Capable of destroying the parasites surface membrane, killing it.	Kidney damage, allergic reactions.
Miltefosine	It can be used to treat different types of leishmaniasis.	Long treatment; fetus deformity. Requires daily injections. Cannot be administrated to children.
Paromomycin	Alternative treatment, can be used in combination with other anti-leishmanial drugs.	Poor oral absorption; requires painful deep intramuscular injections, daily for at least 21 days. It can cause hearing loss.
Sodium Stibogluconate	Classical treatment, acts in the parasite's oxidative pathways.	Toxic drug, affects liver, kidneys, heart and pancreas.

Another important aspect related to leishmaniasis is the occurrence of cases in which the parasite is resistant to the current drugs available (CROFT, 2006). This, associated to a number of unpleasant side effects associated to these drugs, make the treatment of leishmaniasis more challenging and needs new drugs and strategies to fight the disease.

1.2: DNA transcription and trans-splicing

1.2.1: DNA transcription

Eukaryotic transcription is the process in which information contained in DNA is converted into RNA, being the first stage in gene expression. It has 3 major steps: transcription initiation, elongation and termination.

Transcription initiation begins with DNA sequence, called promoter, binding to RNA polymerase, starting the transcription process.

Any required transcription initiation factors also bind at this stage. The promoter-RNA polymerase complex goes through structural changes, in such way that the DNA starts to unwind, forming a transcription bubble of single stranded DNA. The process occurs in the 5' to 3' direction and requires only one DNA strand as template.

During elongation, multiple tasks take place simultaneously; RNA synthesis is catalyzed, DNA is initially unwinded, then reannealed and the RNA product dissociated as the RNA polymerase moves along, and finally, the RNA product is proofread, in a process similar to "quality control".

The final stage, termination, takes place after the RNA polymerase has transcribed the whole gene, stopping transcription. At this step, the RNA product is released, and RNA polymerase dissociates from DNA. Often, well defined DNA sequences trigger transcription termination, however, exceptions are common. (WATSON, 2014).

1.2.2: *trans*-splicing

Splicing is the process of joining exons in order to form mature mRNA. The majority of organisms use a *cis*-splicing method, in which the introns of the same pre-mRNA transcriptome are removed, and its exons joined together (LIANG, 2003).

One remarkable characteristic of trypanosomatids is their use of *trans*-splicing. In *trans*-splicing, exons of distinct pre-mRNA transcripts are joined together, in such a way that the mature mRNA is composed of exons derived from different origins. Although "normal" *cis*-splicing is observed in a few genes in trypanosomatids, *trans*-splicing is responsible for almost the totality of mRNA formation.

Splicing requires a short RNA sequence called the Splice Leader RNA (SL RNA), that provides the RNA cap, needed for translation initiation. In trypanosomatids, the SL RNA has around 141 nt, of which 39 nt comprise the SL exon (LASDA, 2011).

Mechanistically, *trans*-splicing needs SL RNA and pre-mRNA. Typically, the SL exon is followed by a GU dinucleotide sequence (5' splice site), while in the pre-mRNA, there is an A nucleotide that acts as the Branching Point (BP), followed by a polypyrimidine tract, responsible for spliceossome assembly, and before the exon sequence an AG dinucleotide sequence (3' splice site).

The SL RNA initially transfers part of the pre-mRNA BP via its 5' splice site, having its 3' end hydroxylated. Subsequently, the SL RNA then fuses to the pre-mRNA exon, via the 3' splice site, in such way that at the end of this process two products are obtained: a mature mRNA (SL-exon) and

a Y-structured intermediary, that will undergo debranching and degradation. Figure 9 illustrates this mechanism.

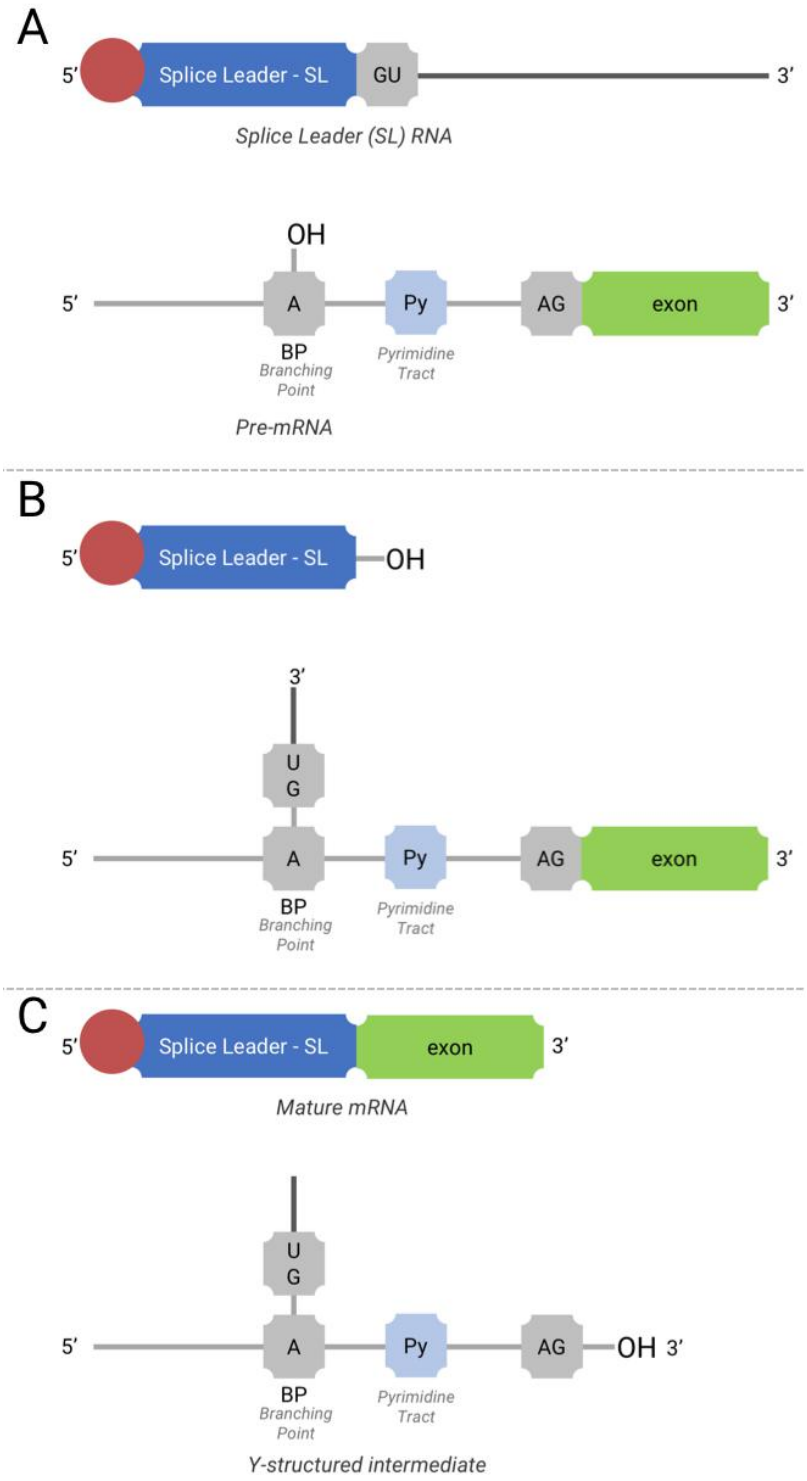


Figure 9: Mechanism of trans-Splicing (adapted from LIANG et al, 2003). (A)Initial components for trans-splicing; (B) first step of trans-splicing; (C) second step of trans-splicing, with Y-structures intermediate and mature RNA formation.

1.3: Trypanosomatid Translation Initiation Complex

Gene expression leading to protein synthesis is a vital process in all organisms, and requires a transcription stage, followed by translation. The translation stage is composed of four essential processes: initiation, elongation, termination and recycling.

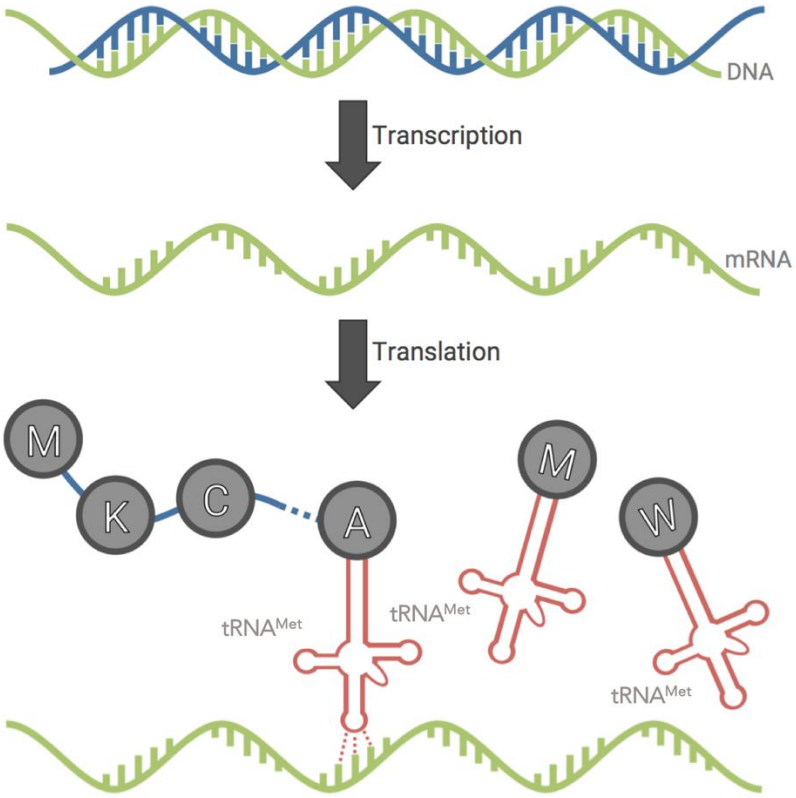


Figure 10: Overview of gene expression leading to protein synthesis.

Protein synthesis has a very important role in gene expression. Its first stage, translation initiation, is believed to be the most important point in mRNA translation regulation, and therefore is commonly the main focus of such studies.

In translation initiation, the ribosome is recruited to the mRNA and, with the aid of multiple initiation factors, a start codon is located. In elongation, a protein chain is synthesized from amino acids on the basis of a code in the RNA and in termination the polypeptide is released and

in recycling the ribosomal complex on the mRNA is disassembled and prepared for a new round of translation.

It is believed that translation initiation is the main control stage in the process of protein synthesis, since it consumes a very large amount of energy, with works describing the production of more than 10000 protein molecules per cell, every second (FIRCZUK, 2013; VON DER HAAR, 2008). For this reason, studies that focused on translation initiation might provide very important biological data, that can be used, for example in the identification of new therapeutic targets.

The trypanosomatid translation initiation complex has unique characteristics, and has its parts assembled in a different way than observed in other eukaryotes.

Of particular relevance to the present study is the observation that a strong interaction between the translation factors eIF4G and PABP1 plays a significant role in the translation machineries of eukaryotes other than the trypanosomatids. Figures 11 and 13 display the proposed models of translation initiation complex, for higher eukaryotes and trypanosomatids (adapted from ZINOVIEV, 2012).

1.3.1: The Eukaryotic Translation Initiation Complex

mRNA translation initiation is a complex, multi-step process that involves many different initiation factors. A brief summary of the most relevant factors is presented below:

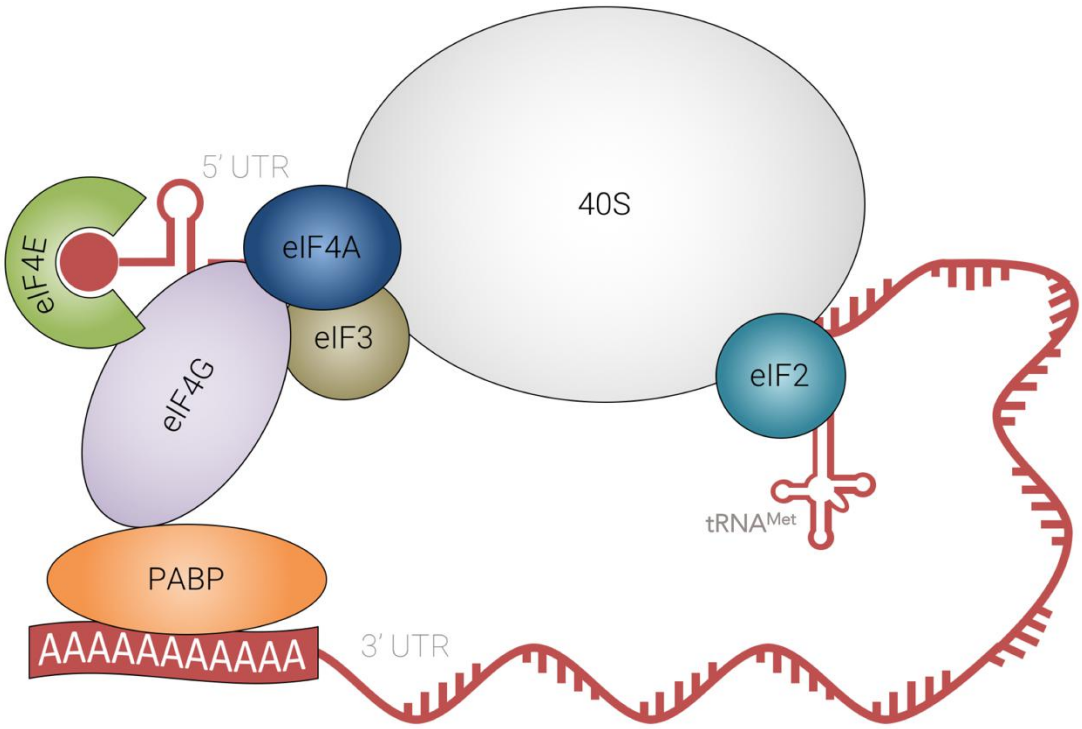


Figure 11: Model of higher eukaryotes translation initiation complex. Adapted from SHAPIRA, 2012.

eIF4E

Initiation factor eIF4E is responsible for m^7GDP binding, being therefore known as the "cap-binding" protein. The eIF4E structure is described as a "cupped hand" that holds the cap structure, and this binding enables the subsequent scanning process that leads to identification of a start codon (HINNESBUSCH, 2012).

PABP

PABP - Poly(A) Binding Protein binds to the mRNA poly(A) tail and plays a role in stabilizing it. PABP has two major sectors: a RNA recognition sector, comprising four RRM (RNA Recognition Motif), followed by a poly-glycine linker, that leads to the PABC domain. PABP also binds to eIF4G and eRF3, and in so doing is thought to act as a type of 'scaffolding' protein, bringing multiple factors together (compare eIF4G below). PABP1

is involved (together with eIF4F) in circularizing mRNA by binding to its 3' poly(A) tail, and simultaneously to the eIF4F complex (that in its turn binds to the 5' end of mRNA) (HINNEBUSCH, 2012).

eIF4G

eIF4G plays a recognized role as "scaffolding" protein. The reason for this is that it binds to several translation initiation factors, including eIF4E, PABP eIF4A, eIF1, eIF5, among others. This major role as a recruitment hub for other proteins assemble onto RNA, in such way that the mRNA circularizes by virtue of an interaction chain: 5'cap-eIF4E-eIF4G-PABP-poly(A) tail (KAPP, 2004).

eIF4A

eIF4A is an ATP-dependent DEAD RNA helicase, that is thought to act by removing secondary structure from the mRNA 5'-UTR region.

eIF2

eIF2 is involved in ternary complex formation, which comprises eIF2, GTP and Met-$_t$RNA. The ternary complex binds to the 40S subunit, thus helping to prepare the ribosomal subunit for start codon recognition.

eIF3

eIF3 is the largest translation initiation complex, comprising to up to thirteen subunits, depending on the organism (in *L. major*, 8 subunits, around 480 kDa; in humans, 13 subunits, around 800 kDa). Its activity is related to ribosomal subunit recruitment, and it also participates in a multiple number of different steps in translation initiation. It binds to the 40S subunit and promotes the subsequent mRNA scanning.

eIF4B

eIF4B is a RNA-binding factor related to eIF4A, enhancing its helicase activity. It is also involved in mRNA recruitment to the preinitiation complex (PIC), that is composed of Met_tRNA, eIF1, eIF1A, eIF2 and eIF3.

eIF1, eIF1A and eIF5

eIF1 and eIF1A are a translation initiation factors that are part of the 43S PIC. They are described as being involved in 43S PIC "open conformation" stabilization. Their activity is required for the correct ribosomal scanning and codon selection. eIF5 is a GTPase-activating protein, involved in initiation codon recognition, interacting with the 40S ribosomal subunit. eIF5 binds to eIF2, inducing GTP hydrolysis during initiation codon recognition. (HINNEBUSCH, 2012).

eIF4F complex

The eIF4F complex is composed of eIF4E, eIF4G and eIF4A, and is responsible for 40S recruitment and is involved in cap-dependent translation initiation.

As described above, translation initiation requires a significant number of translation factors, that are needed for mRNA assembly and further codon screening. This is a very complex process, and Figure 12 summarizes some of the main protein-protein interactions (interactome) involved in translation initiation.

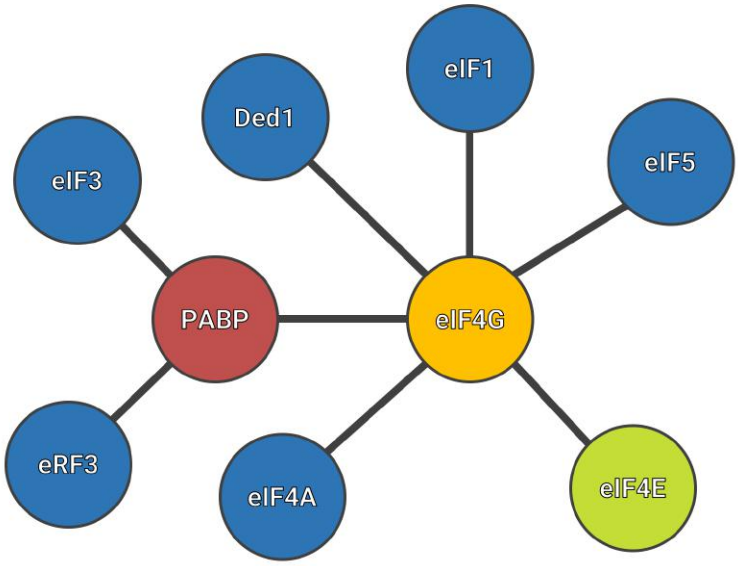


Figure 12: Simplified translation initiation complex interactome.

Among all the translation initiation factors, there are three that are particularly important in this work: eIF4E, PABP and eIF4G. In almost all eukaryotes, eIF4G acts as a "scaffold" factor, assembling a large number of different factors. However, in trypanosomatids, it is proposed that these interactions might be assembled differently, and as will be demonstrated in this work, an interaction between PABP and eIF4E.

Translation initiation has as one of its main steps the assembly of the PIC and screening for the start codon; in cap-dependent translation initiation, eIF4E binds to the mRNA's cap-structure, at the 5' end, while PABP binds to the mRNA's poly(A) tail, aiding in mRNA circularization. As mentioned above, in trypanosomatids, the interactions of the translation factors suggest that the translation initiation complex is assembled in a different way, with eIF4E interacting with PABP.

Trypanosomatids have multiple translation initiation factors isomers, and in the specific case of *Leishmania*, there are six eIF4E isoforms, two PABP isoforms and five eIF4G isoforms.

1.3.2: *Leishmania* eIF4E, PABP and eIF4G isoforms

Several *Leishmania* translation initiation proteins have multiple isoforms. For instance, *L. major* has 6 eIF4Es, 3 PABPs and 5 eIF4Gs. A summary of these proteins is presented below.

eIF4E

Leishmania has six eIF4E isoforms, eIF4E1 to eIF4E6, all located in the cytoplasm (FREIRE, 2017). The literature contains data regarding the first four isoforms, while little is known about the last two, more recently identified isomers. The eIF4E factors are expected to bind to both RNA's m^7GTP and methylated cap-4.

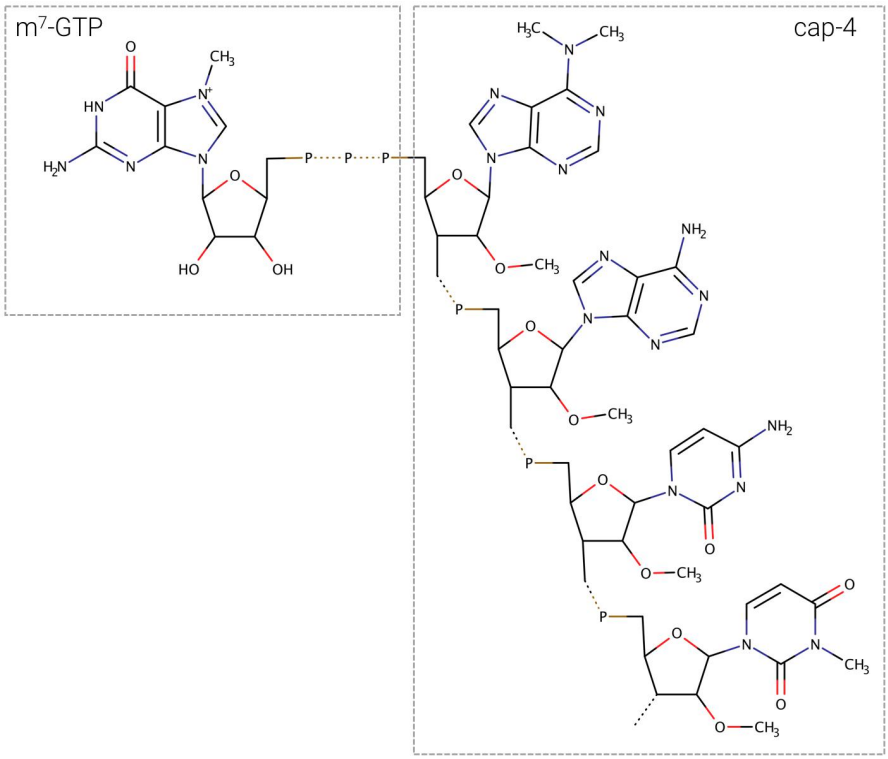


Figure 13: m^7GTP-cap-4 structure.

The literature shows that (ZINOVIEV, 2012; FREIRE, 2017) eIF4E1-6 have different affinities toward m^7GTP and cap4.

Table 2 Trypanosomatid eIF4Es affinities and L. major eIF4Es sizes.

	L. major eIF4Es	Binds To	
Isoform	Size/kDa	m^7GTP	cap-4
eIF4E1	23.8	✓	✓
eIF4E2	31.4	X	✓
eIF4E3	37.7	✓	X
eIF4E4	48.1	✓	✓
eIF4E5	23.9	✓	✓
eIF4E6	23.2	✓	✓

A particular feature in the isoforms eIF4E3 and eIF4E4 is the presence of a long, unique N-terminal segment, found only in trypanosomatids. The function of this longer N-terminal is not completely clear, and therefore, has been the object of considerable speculation. One major characteristic of this N-terminal extension is the lack of structure, possibly due to the presence of a large number of prolines in its sequence, since these are expected to destabilize secondary structure formation. eIF4E5 and eIF4E6 are the most recently described and least studied of the trypanosomatid eIF4Es. These are the small isoforms, with eIF4E6 being the smallest, and have limited homology with the other isoforms.

The existence of a number of eIF4E isoforms suggest that each version might have a different role during trypanosomatids life span. It has been identified in *Leishmania* (MCNICOLL, 2005) differences in gene

expression depending on the parasite developmental stage, in such way that genes responsible for some of the eIF4E isomers (but not all) expression had their expression increased in the amastigote stage (but not in the promastigote).

There were several attempts in the literature to identify the different functions of each of eIF4E isoforms, and a summary is presented below (FREIRE, 2017).

eIF4E-1

Previous studies have revealed (YOFFE, 2006) that eIF4E-1 levels are significantly higher in amastigotes than in promastigotes. These variations might be associated with differences in environmental changes in temperature, since amastigotes are found in the mammalian host (higher temperatures), and axenic cultures of promastigotes cultivated at higher temperatures undergo a morphology change, losing the flagellum and becoming more rounded, amastigote-like. The expression of the other isoforms did not show significant differences with temperature variation.

eIF4E-2

Little is known regarding eIF4E2 function, but in trypanosomatids studies suggest that it may not be involved in translation (FREIRE, 2017). It binds to cap-4 but not to m^7GTP, and it is suggested that it might be associated with histone mRNA stem-loop binding protein interactions (FREIRE, 2017).

eIF4E-3

eIF4E-3 binds to m^7GTP but not to cap4, and for this reason it is unlikely that it is the essential basal isoform (*i.e.* the one essential for cell survival) for trypanosomatid translation initiation. If eIF4E-3 is knocked out, cell death takes place both in promastigotes and amastigotes, and there are indications that it participates in translation, at least in promastigotes (FREIRE, 2011).

eIF4E-4

eIF4E-4 is the largest eIF4E isoform in trypanosomatids, and its long N-terminal sequence largely contributes to its increased size. This N-terminal sequence is not observed in any other eukaryotes, being unique to trypanosomatids. It has a high proline content, and it has been suggested that it is intrinsically disordered. These properties make the N-terminal region of eIF4E4 an important subject of research seeking to understand its function and behaviour. eIF4E-4 binds both to m^7GTP and cap-4, and its knockout leads to cell death in amastigotes and arrests translation and impairs cell viability in promastigotes.

eIF4E-5 and eIF4E-6

There is little information available about the roles of eIF4E5 and eIF4E6. Initial studies indicate that both proteins are probably involved in trypanosomatid cell motility and in procyclic flagellar attachment. This hypothesis has its origin in knockout experiments, since after eIF4E-5 knockout cells displayed impaired motility (swimming), while an eIF4E-6 knockout leads to flagellar detachment. However, more studies are required in order to obtain a clearer image of their respective functions.

PABP

Trypanosomatid PABPs have a highly conserved structure, composed of 4 RRMs in the N-terminal, a non-conserved linking region, and a conserved PABC domain, in the C-terminal region (DA COSTA LIMA, 2010). In normal conditions, all three isoforms are found in the cytoplasm, with only marginal amounts being detected in the nucleus.

PABP1

The PABP1 is the isoform with the major translation initiation activity, being found in the cytoplasm. PABP1 is responsible for mRNA binding, through its RRM domains. It interacts with eIF4G3, and thus participates in eIF4F complex function (although it is not part of the eIF4F complex).

PABP2 and PABP3

PABP2 and PABP3 are believed to not to participate in translation initiation, being probably involved in other regulatory functions (DA COSTA LIMA, 2010). Experiments in which transcription was inhibited, both PABP2 and PABP3 accumulated in the nucleus, although not directly associated to the chromatin (BATES, 2000). It has been also suggested that PABP2 may play an important role in mRNA metabolism and regulation (DA COSTA LIMA, 2010; BATES, 2000). PABP3 was not studied in detail, so many aspects of its properties and functions are still to be determined.

eIF4G

Trypanosomatid eIF4Gs have in common the presence of a conserved MIF4G domain, characteristic to all eIF4Gs (DHALIA, 2005). They can also

be divided into two groups with specific characteristics, as described below:

eIF4G1, eIF4G2 and eIF4G5

eIF4G1, eIF4G2 and eIF4G5 have so far not been linked functionally to translation initiation. They bind to eIF4E5 and eIF4E6 (FREIRE, 2014-1), and their functions are not yet clear, although it is suspected that they might have regulatory roles (FREIRE, 2014-2).

eIF4G3 and eIF4G4

eIF4G3 and eIF4G4 have been described as participating in translation initiation (DHALIA, 2004). Besides the characteristic MIF4G domain, the two proteins share several conserved segments, and its N-terminal is the most relevant. Both eIF4G3 and eIF4G4 have very short N-terminal regions, of only around 50 residues and 75 residues, respectively. This contrasts with the long eIF4Gs N-terminals from yeast (600 residues) and humans (700 residues) (MOURA, 2015).

1.3.3: Translation Elongation

Translation elongation is the most conserved step in protein translation (KAPP, 2004; DEVER, 2012), and its mechanism is very similar in different organisms, from bacteria to eukaryotes.

Elongation requires that tRNA binds to the ribosomal A site, in such way that the ribosome enables in peptide bond formation (given that the appropriate tRNA is present). The next step involves the movement of the next codon into the A site, and then having the next peptide bond formation. This step is repeated until a stop codon is found, and termination happens.

1.3.4: Translation Termination

When a stop codon is found during the elongation process, the nascent peptide is released from the ribosome. The ribosomal subunits can then be recycled, and then used in the formation of a new initiation complex, for the production of another peptide (KAPP, 2004).

1.3.5: Recycling

The last step in translation initiation is recycling. In this step, the ribosomal subunits are disassembled, and then used for a new round of translation initiation. During this process, there is the release of mRNA and tRNA (KAPP, 2004).

1.4: MLLE Domains and PAM2 motifs

1.4.1: MLLE Domain

The MLLE domain is a well-known peptide binding motif, composed of around 70 aminoacid residues, organized in five α-helices. The MLLE, also called "mademoiselle" domain is present in several protein, including PABP proteins (XIE, 2014). Although the MLLE domain comprises usually around 70 amino acids in its sequence, the highly conserved KITGMLLE sequence is particularly relevant (KOZLOV, 2010), since this has an important role in binding to the PAM2 peptide sequence, through its interactions with the correspondent amino acid present in the ligand.

1.4.2: PAM2 Motif

The PAM2 motif, or PABP-interacting Motif 2, recognizes proteins containing the MLLE domain, and has the following conserved amino acid sequence:

$$(L/P/F)x(P/V)xAxx(F/W)xP \quad (1)$$

Or commonly:

$$LNxxAxEFxP \quad (2)$$

The PAM2 motif binds to MLLE domains by interacting with hydrophobic pockets that are formed along the MLLE domain α-helices. There are three amino acid residues in particular in the PAM2 motif that are important for the binding to MLLE, the first conserved leucine, and the conserved alanine and phenylalanine (KOZLOV, 2010; XIE, 2014).

1.4.3: PAM2/MLLE Interaction

The PAM2 residues interact with a number of amino acid residues in MLLE domains (KOZLOV, 2010). However, as described previously, the first conserved leucine, the conserved alanine and the conserved phenylalanine are particularly relevant. The leucine and phenylalanine bind to hydrophobic pockets located in the α-helices around the central α-helix in the MLLE domain, while the PAM2 conserved alanine exploits the space left by the central glycine in the KITGMLLE sequence to link the two hydrophobic pockets. One example of interacting partners in which the PAM2 motif plays a key role is the PABP1/eRF3 complex, involved in translation termination (ALBRECHT, 2004; OSAWA, 2012).

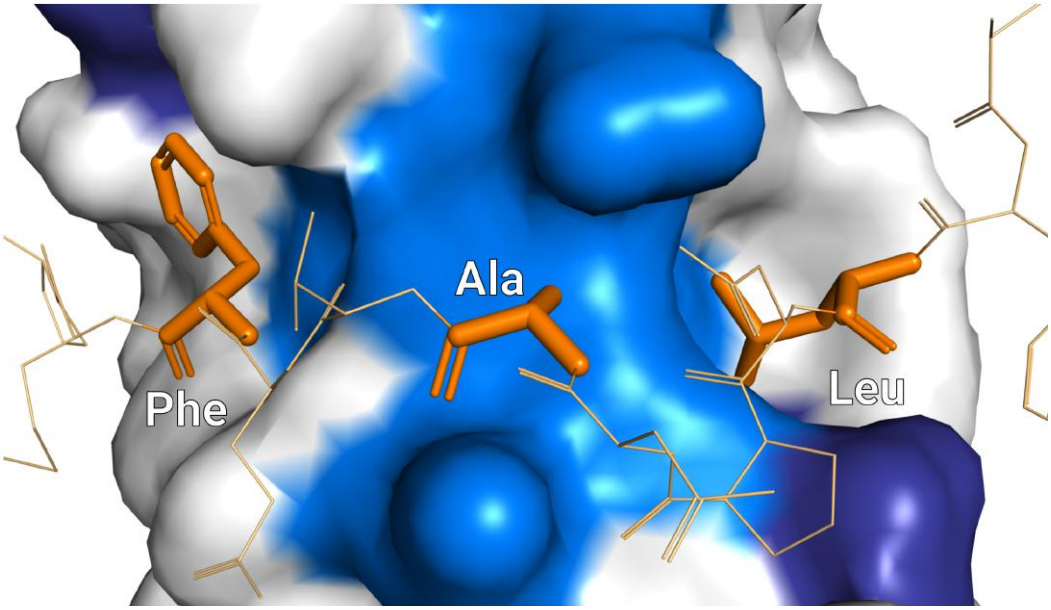


Figure 14: PAM2 key residues interacting with MLLE domain; hydrophobic pockets highlighted in dark blue.

The MLLE domain is well conserved, and studies indicate that it does not undergo significant conformational changes for PAM2 binding, being completely pre-formed for PAM recognition.

1.5: Objectives

1.5.1: General Objectives

1. To investigate the architecture of the trypanosomatid translation complex comprising eIF4E4, PABP1 and eIF4G3. Trypanosomatid eIF4E4 has a unique, long N-terminal, making it an interesting object of research, that could potentially be used as a drug target, since these interactions are not observed in humans.

1.5.2: Specific Objectives

1. To characterize the interactions between *Leishmania major* eIF4E4, eIF4G3, and PABP1.

2. To characterize structural details of the eIF4E4-PABP1 interaction.

3. To quantitate the relative affinities of the eIF4E4-PABP1 and the PABP1-eIF4G3 interactions.

Chapter 2 – Techniques Overview - Materials and Methods

2.1: Buffers List

A number of buffers were used in the various experiments, their names and composition are listed below:

Buffer A:
150 mM KCl
10 mM Tris-HCl pH 8
10% glycerol
10 mM $MgCl_2$
10 mM β-Mercaptoethanol

Buffer B:
150 mM KCl
10 mM HEPES pH 7.5
5% glycerol
1 mM DTT

Buffer C
150 mM KCl
20 mM Tris-HCl pH 7.5
5% Glycerol
10mM $MgCl_2$
5 mM β-Mercaptoethanol
0.01 % Triton X-100

Buffer D
100 mM KCl
20 mM Tris-HCl pH 7.5
10 mM $MgCl_2$
5% Glycerol

Buffer E
300 mM KCl
40 mM Tris-HCl pH 7.4
2 mM EDTA
0.1% Triton X-100
5% Glycerol
5 mM β-Mercaptoethanol
1 mM PMSF

Buffer F
50 mM HEPES pH 7.5
100 mM NaCl
1 mM DTT
0.01% sodium azide
0.1 mM EDTA

2.2: Constructs Preparation

Our studies aimed the investigation of interactions between *Leishmania*'s eIF4E4 and PABP1 proteins. eIF4E4 is a 48kDa protein from *Leishmania major*, that binds to mRNA's cap structure, and presents the following primary structure:

```
         10         20         30         40         50
MSTPLDVRAAEYSPSFAVTMKKTVAAAPPKSPAPAKSKISVTRTGVNTTY
         60         70         80         90        100
PMPPPMPEKNYAPFFAEGCQTIAASKASMPPVQPASPLPPMHSAPPTASV
        110        120        130        140        150
VSNSIPPSSPATAPGERSPAVAARSVPTRFSPATVPRHHMNPNATEFMPG
        160        170        180        190        200
RRNGPDGGLEALPTSTADMELAKTPAGAAAAAVHAPSLPGAVRRSLQNSP
        210        220        230        240        250
IIQPSRLSVKSASEIEAISKNSALNAAAAAYVPQRTLARVVLTQPSPLAL
        260        270        280        290        300
APSEDPAKNNIEMMLDDLWCLFYLPTTLGENIKEEDYNPTLVFRVDSILT
        310        320        330        340        350
FWRVVNNIAAPSELQLSTLYLFRDGIDPKWEDPANRDGGIVKVKATAAQV
        360        370        380        390        400
DEAWELLLCRTIGDSWSPSVRETVNGVVLKVRERAYWLELWVTKNSSALQ
        410        420        430        440    447
KDLAELWHPILGASFATTYLTHAMMQERSHAAAALAAEKQKKNRRRY
```

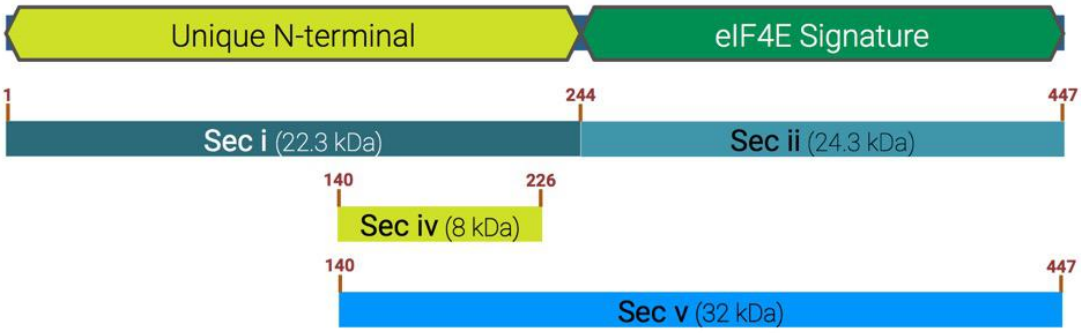


Figure 15: *Leishmania eIF4E-4 segmentation in several domains.*

PABP1 is a 62kDa protein from *Leishmania major*, that binds to mRNA's poly(A) tail, and presents the following primary structure:

```
        10         20         30         40         50
MAAAVQEAAAPVAHQPQMDKPMQIASIYVGDLDATINEPQLVELFKPFGT
        60         70         80         90        100
ILNVRVCRDIITQRSLGYGYVNFDNHDSAEKAIESMNFKRVGDKCVRLMW
       110        120        130        140        150
QQRDPALRYSGNGNVFVKNLEKDVDSKSLHDIFTKFGSILSCKVMQDEEG
       160        170        180        190        200
KSRGYGFVHFKDETSAKDAIVKMNGAADHASEDKKALYVANFIRRNARLA
       210        220        230        240        250
ALVANFTNVYIKQVLPTVNKDVIEKFFAKFGGITSAAACKDKSGRVFAFC
       260        270        280        290        300
NFEKHDDAVKAVEAMHDHHIDGITAPGEKLYVQRAQPRSERLIALRQKYM
       310        320        330        340        350
QHQALGNNLYVRNFDPEFTGADLLELFKEYGEVKSCRVMVSESGVSRGFG
       360        370        380        390        400
FVSFSNADEANAALREMNGRMLNGKPLIVNIAQRRDQRYTMLRLQFQQRL
       410        420        430        440        450
QMMMRQMHQPMPFVGSQGRPMRGRGGRQQLGGRAQGHPMPMPSPQQPQAP
       460        470        480        490        500
AQPQGFATPSAVGFVQATPKHSPGDVPETPPLPPITPQELESMSPQEQRA
       510        520        530        540        550
ALGDRLFLKVYEIAPELAPKITGMFLEMKPKEAYELLNDQKRLEERVTEA
       560
LCVLKAHQTA
```

The PABP1 was divided into subdomains, using its natural architecture as guide. The segmentation into smaller subdomains would be used as a toolkit, allowing the investigation and screen of the different PABP1 regions, in such way that it would be possible to analyze how they would behave when subjected to binding experiments. For instance, subdomain PABP1(A) explored the first two PABP1 RRM domains, while the subdomain PABP1(J) corresponds to the PABC domain.

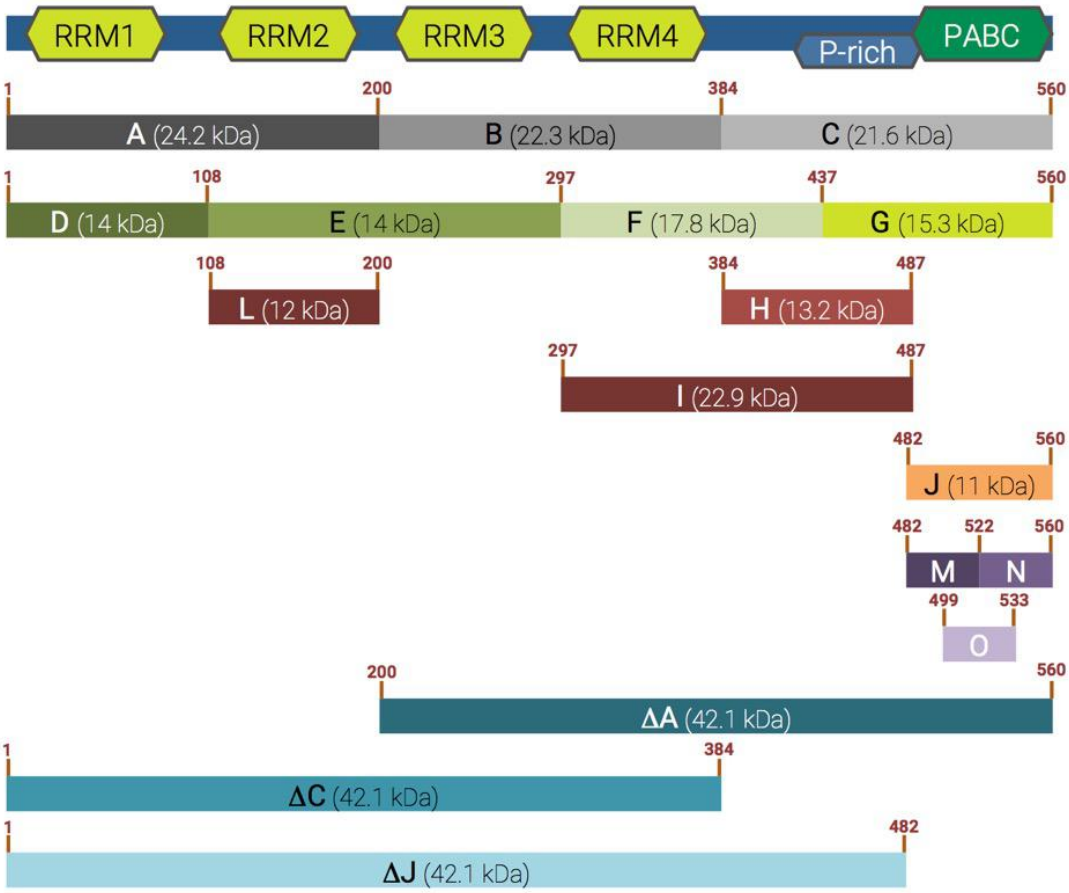


Figure 16: Leishmania PABP1 segmentation in several domains.

2.2.1: Restriction based cloning and Gibson assembly

Subcloning:

Cloning technique was used in order to obtain some of the vectors for protein expression.

The cloning procedure was performed as follows:

The gene of interest was amplified using primers with 5' overhangs containing the desired restriction sites.

The amplified insert and the target vector were then digested with the appropriate restriction enzyme. Unless otherwise stated, the restriction enzymes BamHI and HindIII were used in all cloning.

Insert and vector purification was performed using Thermo Scientific GeneJET Plasmid Purification kit, and if required, gel extraction was done using Thermo Scientific GeneJET Gel Extraction Kit.

The insert was digested using 100µL reactions, composed by 10µL of digestion buffer (specific to the restriction enzymes used), and 2µL of each restriction enzyme, and 88µL of nuclease-free water. The vector was digest also in 100µL reactions, composed by 10µL of digestion buffer (specific to the restriction enzymes used), and 2µL of each restriction enzyme, 10µL of vector and 78µL of nuclease-free water. The reactions were then incubated at 37˚C for 1 hour.

Vector dephosphorylation was made by adding 1µL of FastAP Thermosensitive Phosphatase (Fermentas), incubated at 37˚C for 15 minutes, and inactivated by incubation at 75˚C for 5 minutes. The products were then purified using Thermo Scientific GeneJET Plasmid Purification kit and used for ligation.

The insert/vector ligation was performed in 10µL reactions, using 1µL of 10*NEB Ligase buffer, and the appropriate volume of vector/insert in order to achieve a 1:3 vector/insert molar ratio; then 1µL of T4 ligase, and enough nuclease-free water to total 10µL.

The mixture was incubated at room temperature for 30 minutes, ant then transformed into *E. coli*.

Gibson Assembly:

The Gibson Assembly technique was used in order to obtain some of the vectors for protein expression. The Gibson Mix (5x Isothermal Reaction Mix) has the following components:

5x Isothermal Reaction Mix

-Tris-HCl pH 7.5: 500mM
-$MgCl_2$: 100mM
-dATP: 1mM
-dCTP: 1mM
-dGTP: 1mM
-dTTP: 1mM
-DTT: 50mM
-PEG8000: 30mM
-NAD: 5mM
-ddH_2O to the desired volume

1.33x Assembly Master Mix

Protocol for 1.2 mL of Master Mix:

-320µL 5X Isothermal Master Mix
-0.64µL 10 U/µL T5 exonuclease
-20µL 2 U/µL Phusion DNA Pol
-0.16µL 40000 U/µL Taq DNA Ligase
-860µL ddH_2O

For the reaction, 15µL of Gibson Master Mix was added to 2.5µL of digested vector, and then split in two 7.0µL aliquots, A and B. To aliquot A, 1.0 µL of sterile, nuclease-free water was added, to be used as negative control. To aliquot B, 1.0 µL of 10µM GeneBLOCK was added, and then incubated in a thermocycler at 50˚C for 30 minutes. After the reaction ended, 2.0 µL of the reaction was transformed into competent cells and plated on selective media. The Figure 17 summarizes the procedure.

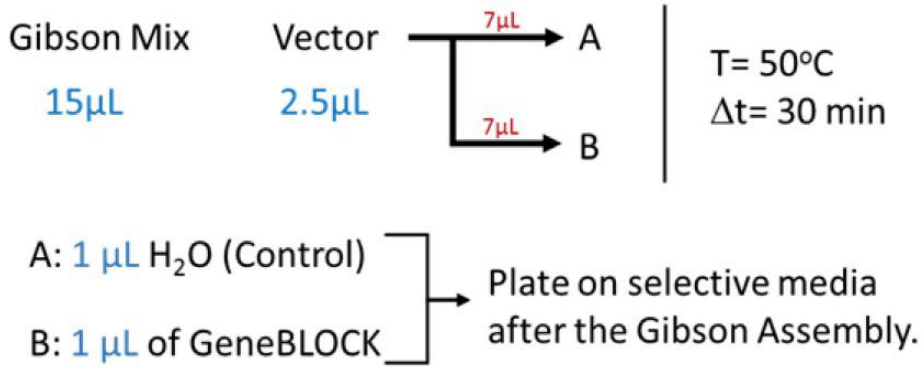


Figure 17: Gibson Assembly procedure summary.

2.2.2: *E. coli* transformation

E. coli transformation for protein expression was performed as follows:
The appropriate amount of competent cells were thawed and added into a single sterile 1.5mL Eppendorf tube. In another tube, 1.0 µL of vector was added, and then 40-70µL of competent cells were then added to the vector.
The tubes were then incubated on ice for 10 minutes, and then heat-shocked for 30 seconds at 42˚C. The tubes were then cooled down on ice for 1 minute, and 0.5-1.0 mL of sterile LB/SOC media added to the tubes, which were then incubated for 30-60 minutes at 37˚C. Finally, the cells were plated on selective media (LB-agar + antibiotic) and incubated overnight at 37˚C.

2.2.3: Protein overproduction

All proteins were prepared as follows, unless stated otherwise:

The media chosen for protein expression was Terrific Broth with 25 mM Ampicillin. When the OD_{600} reached 0.6-0.8, enough IPTG was added to reach 1.0mM concentration. The overexpression was conducted at 16˚C, for 16h.
The cells were then pelleted by centrifugation at 10000 rpm for 45 minutes at 4˚C in a Beckman-Coulter Avanti J-26 XP centrifuge using JLA 8.1 rotor, and then lysed through 10 cycles of 30 seconds sonication/ 1 minute rest, on ice.

After sonication, the cells were pelleted by centrifugation at 16000 rpm for 30 minutes at 4˚C in a Beckman-Coulter Allegra 64R centrifuge using F0685 rotor. The supernatant was collected and then used for protein extraction and purification.

2.3: Protein purification

Protein purification was performed using a combination of methods, essentially affinity and size exclusion.

2.3.1: Affinity purification

The supernatant obtained in the protein overproduction stage was incubated overnight with cobalt resin (HisPur, Thermo Fisher Scientific). After the incubation, the resin was washed thoroughly and then eluted with at least 10 1.0mL elutions of Buffer A with 150 mM imidazole. The elution aliquots were then joined into a single sample, and then concentrated using Sartorius Vivaspin 3000 MWCO PES 20mL concentrators, filter at 3000 rpm in an Eppendorf 5810R centrifuge using S-4-104 rotor until the volume was reduced to 500-1000 µL. The concentrated fractions were then used in size exclusion chromatography, in order to isolate monomeric protein from any aggregates.

2.3.2: Size Exclusion chromatography purification

The concentrated samples from the affinity purification were subjected to exclusion chromatography. For this, 500µL of sample was injected in an AKTA FPLC Purifier UPC100 with size exclusion columns Superdex

Increase 75/300 GL and Superdex Increase 200/300 GL (depending on the protein size), equilibrated with 2 CV until constant mAU, in order to separate protein monomers from aggregates (1.0mL loop, emptied with 0.7mL; flow of 0.3mL/min, 600µL fractions). After gel filtration, the fractions containing the protein monomer were, if needed, concentrated (but avoiding aggregation), aliquoted and stored at -80˚C for use in the various experiments.

2.4: Protein identification

In order to verify each step of protein production, it was necessary to confirm protein identities. A combination of electrophoresis and mass spectrometry was used, according to the stage of protein production/purification.

2.4.1: SDS-PAGE

After protein production and purification, these were subject to analysis by SDS-PAGE and Mass Spectrometry in order to confirm the success in its attainment.

The protein samples were prepared using Morris Formula Sample Buffer prior to electrophoretic run. SDS-PAGE was performed using a BioRad Mini PROTEAN TetraCell System, powered by a BioRad PowerPAC200 power supply. We used precast gradient gels, from Expedeon SDS Precast Gels 4-20% and 4-12%, that were run at constant voltage of 100V, for 80 minutes, and used Expedeon RunBlue Running Buffer. After the electrophoretic run, the gels were stained with Expedeon Instant Blue dye and then scanned using a GelDoc XR+ Imaging System.

2.4.2: Mass spectrometry

The identification of protein identities was performed by mass spectrometry, using a nanoLC-ESI-MS/MS approach. For this purpose, SDS-PAGE sections containing protein bands of interest were digested and analyzed according to the following protocol:

Initially, the bands containing the proteins of interest were cut from the SDS-PAGE gels and digested with trypsin, and then the peptides separated by reversed phase chromatography. For the separation, two columns were utilized, an Acclaim PepMap µ-precolumn cartridge 300 µm i.d. x 5 mm 5 µm 100 Å and an Acclaim PepMap RSLC 75 µm x 25 cm 2 µm 100 Å (Thermo Scientific), installed on an Ultimate 3000 RSLCnano system (Dionex). Two mobile phases were also used; the mobile phase A was composed by 0.1% formic acid in water and mobile phase B 0.1 % formic acid in acetonitrile. In the next step, the samples were loaded onto the µ-precolumn equilibrated in 2% aqueous acetonitrile containing 0.1% trifluoroacetic acid for 8 min at 10 µL min-1 after which peptides were eluted onto the analytical column at 300 nL min^{-1} by increasing the mobile phase B concentration from 4% B to 25% over 22 min then to 90% B over 3 min, followed by a 10 min re-equilibration at 4% B.

The isolated peptides were then subjected to electrospray ionization and analyzed on a Thermo Orbitrap Fusion system (Q-OT-qIT, Thermo Scientific). Precursor peptides scanning from 375 to 1500 m/z were performed at 120K resolution (at 200 m/z) with a 2×10^5 ion count target. Tandem MS was performed by isolation at 1.2 Th using the quadrupole, HCD fragmentation with normalized collision energy of 30, and rapid scan MS analysis in the ion trap. The MS2 ion count target was set to 3 x

10^3 and the max injection time was 200 ms. Precursors with charge state 2–6 were selected and sampled for MS2. The dynamic exclusion duration was set to 30 s with a 10 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on. The instrument was run in top speed mode with 1 s cycle.

Data Analysis

The obtained MS raw data was processed using MSConvert in ProteoWizard Toolkit (version 3.0.5759)1. All the MS2 spectra were searched with Mascot engine (Matrix Science, version 2.4.1) against provided *Leishmania* database and the common Repository of Adventitious Proteins Database (http://www.thegpm.org/cRAP/index.html). Peptides were generated from tryptic digestion with up to two missed cleavages, carbamidomethylation of cysteines as fixed modifications, and oxidation of methionines as variable modifications. Precursor mass tolerance was 10 ppm and product ions were searched at 0.8 Da tolerances.

Scaffold (version Scaffold_4.3.2, Proteome Software Inc.) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Scaffold algorithm. Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (NESVIZHSKII, 2003). Proteins sharing significant peptide evidence were grouped into clusters

2.5: Surface Plasmon Resonance

Surface Plasmon Resonance is a technique that detects changes of the refractive index on a surface upon ligand binding. In this technique, a gold-coated chip is covered with a chemical matrix, often dextran, that can then be used for a range of different methods of target immobilization. Protein-ligand studied can explore a number of different immobilization methods, such as amine coupling, thiol coupling, His-tag affinity, to mention a few.

The usual strategy involves the immobilization of a target protein on the surface chip, using one of the methods aforementioned. After immobilization, the ligand binding can then be investigated, by its addition at varied concentrations. If the ligand binds to the target protein, the local refractive index changes, and this change can be measured by changes in the angle of a laser that is irradiating the chip. The refractive index change is proportional to ligand binding, and through this measure, a number of data can be obtained, as affinity, k_{on} k_{off} rates, among others.

The SPR experiments were performed using a GE Biacore T200 instrument, and surface-binding chip CM5.
The immobilization strategy used was amine coupling, using eIF4E4(iv) (0.2mM) protein as the ligand, Buffer B, flow-path number 2, until an immobilization level of around 1000 units was obtained.

2.5.1: Yes/No binding experiment

Each Yes/No Binding experiment comprises a sequential run of different analytes, in this case PABP1 domains, at the same concentration. This

experiment is used to screen a number of ligands and verify if they bind to an immobilized partner on the chip's surface.

Experimental Conditions:

-Buffer C
-Regeneration Solution: $MgCl_2$ 1M
-Injection Time: 300 s
-Regeneration Time: 120 s
-Dissociation Time: 300 s
-Cell Flow: 10µL/min
-Cell Flow Mode: 2-1

2.5.2: Binding affinity experiment

The Affinity experiment comprises of the sequential run of different analytes, in this case PABP1 domains, at varied concentrations. These experiments are useful to assess the binding behaviour of a ligand, and in ideal conditions, determine its affinity towards a binding partner.
Experimental Conditions:

-Buffer C
-Regeneration Solution: $MgCl_2$ 1M
-Injection Time: 300 s
-Regeneration Time: 120 s
-Dissociation Time: 300 s
-Cell Flow: 10µL/min
-Cell Flow Mode: 2-1

2.6: Microscale Thermophoresis

The MST experiments were performed in a Nanotemper Monolith NT.115 instrument, using standard Monolith NT.115 Capillaries. Protein labelling with fluorescent tag was made using Monolith Protein Labelling Kit Red-Maleimide.

MST experiments detects binding on the basis of alterations in thermal diffusion when a complex is formed, when compared with the thermal diffusion of the individual complex components.

2.6.1: Binding affinity experiment

The binding affinity experiments were performed by labelling eIF4E4(iv) with a fluorescent tag (Monolith Protein Labelling Kit Red-Maleimide), as instructed in the kit manufacturer user manual. In this experiment, the fluorescent tagged protein (eIF4F4(iv)) is held at constant concentration, while the ligand protein (PABP1 domains) concentration is varied. The experiment is conducted by preparing 16 tubes with various combinations of eIF4E4(iv)/PABP1 domain concentrations. This is achieved by serial dilution, and then glass capillaries are added to each tube, until they are filled with the respective solutions. The capillaries are then loaded on a rack and inserted in the MST equipment for measurement. The experiment starts with a quality check, whose main parameters are fluorescence homogeneity between the capillaries, fluorescence intensity, adsorption, aggregation and ligand-induced photobleaching. After the checks, the measurements start, and a graph containing the data is plotted, and then a final signal-to-noise evaluation

is performed. The final step involves the calculation of affinity constants and overall data quality evaluation.

2.7: Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) is a very powerful technique that can provide valuable information in protein/protein and protein/ligand studies. NMR uses the magnetic properties of active atom isotopes, and through a large range of techniques, information regarding structure, dynamics and affinity can be obtained. Since the technique requires samples that are magnetically active, (*i.e.* contain magnetically active isotopes) the proteins used in this study were prepared as isotopically labelled samples.

The experiments in this thesis were performed in collaboration with the group of Prof. Alex Breeze, from the University of Leeds, using a 950 MHz Bruker Ascend Aeon™ NMR magnet with TXO-cryoprobe (5mm) and equipped with a Bruker Avance III HD console and 600 MHz Oxford NMR magnet with QCI-P-cryoprobe equipped with Bruker Avance III HD console.

2.7.1: {^{1}H-^{15}N}-HSQC and TROSY

In the investigation of protein/protein interactions, one of the most useful techniques is the HSQC - Heteronuclear Single Quantum Coherence experiment. In these 2D experiments, two different types of atom are monitored (most commonly the pairs ^{1}H and ^{13}C, or ^{1}H and ^{15}N).

In protein studies, {^{1}H-^{15}N} HSQC spectra provide peaks in a way that each peak corresponds to one aminoacid residue in the primary structure of the protein being studied.

When a ligand is added to a protein solution, the aminoacid residues that are involved in ligand binding can have their chemical shifts changed, *i.e.* the peak can change position (shift) or undergo changes in its intensity, often disappearing.

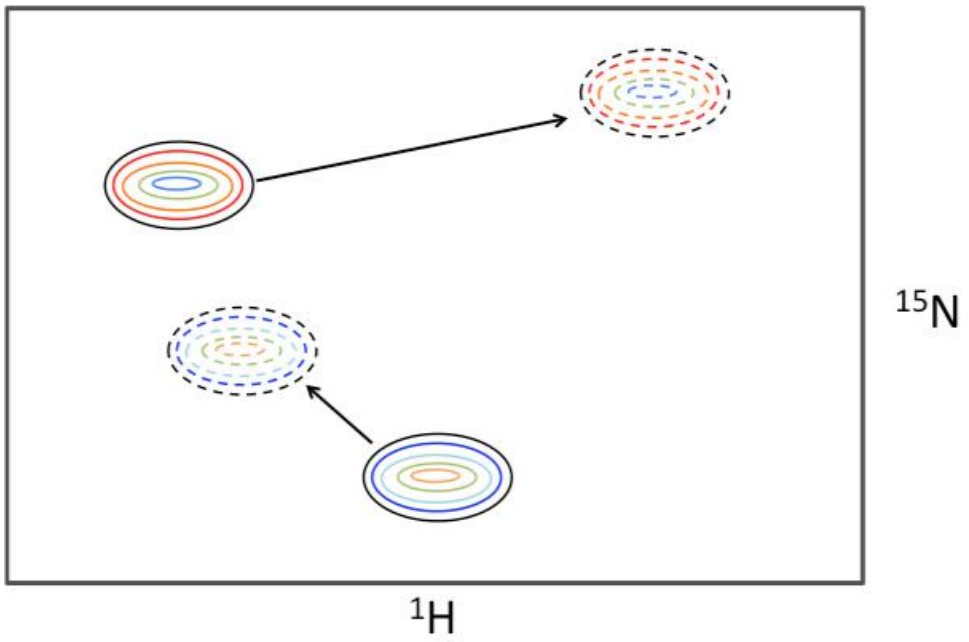


Figure 18: Chemical shifts in interacting peaks in a HSQC spectrum simulation.

These changes can give very good indications that binding is being observed. In order to map the binding interactions in proteins primary structure, it is necessary to assign each peak to an aminoacid, which is done through a number of triple resonance NMR experiments, which allow their assignment.

TROSY (Transverse relaxation-optimized spectroscopy) is an NMR experiment that is often used in the study of larger proteins. When large proteins are investigated by NMR, it is very common to observe signal

overlapping, making the data analysis difficult. TROSY can help in this situation, since it increases resolution and sensitivity (but it requires the use of higher magnetic fields).

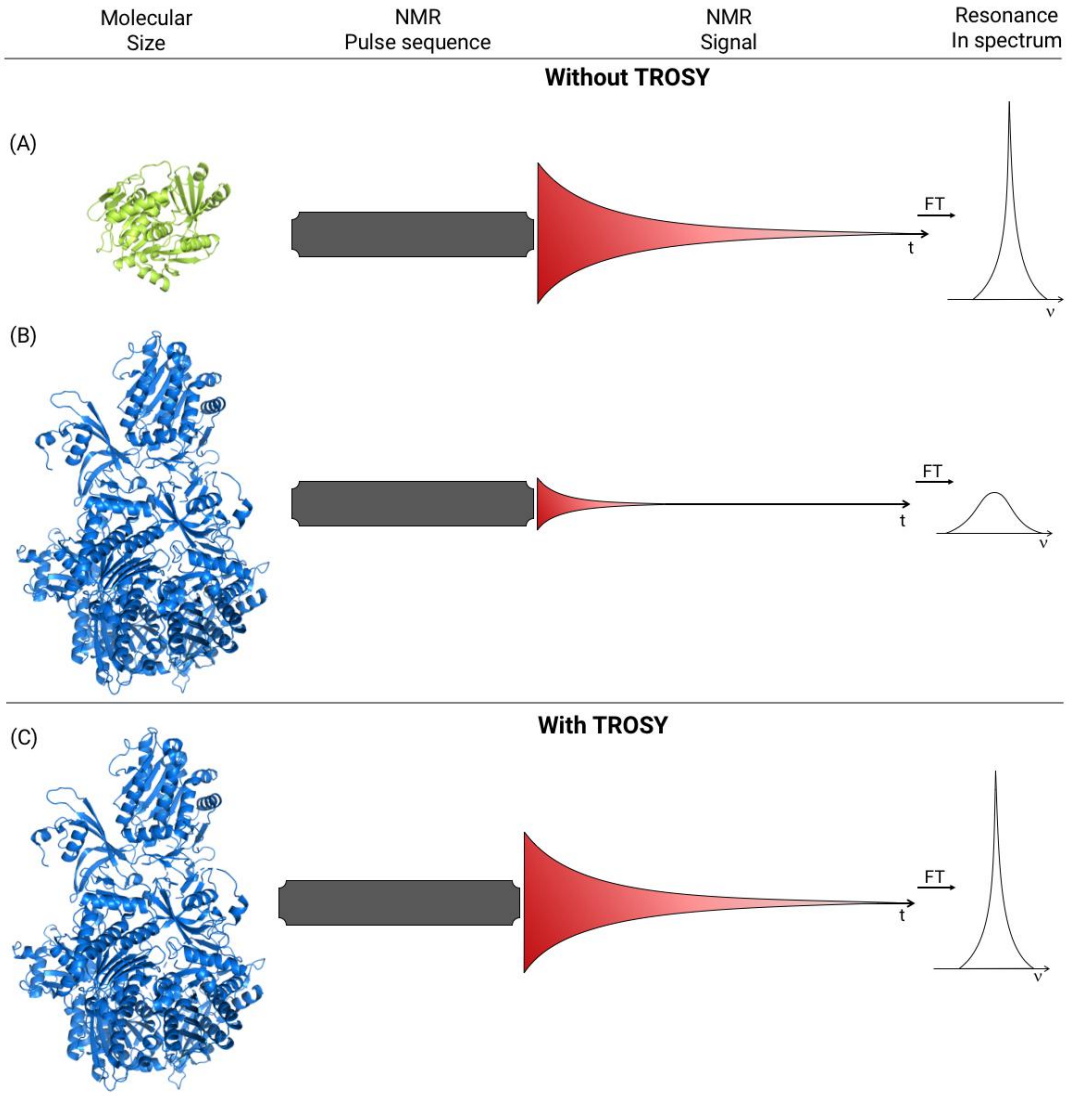


Figure 19: Diagram comparing the difference between TROSY and non-TROSY experiments (Adapted from FERNANDEZ, 2003).

2.7.2: Triple resonance and residue assignment

In order to assign amino acid residues, it is usually necessary, along with a HSQC, to perform experiments that will identify the Cα, Cβ and CO

moieties in each amino acid. Since each peak couples with its corresponding neighbour, it is possible to use these couplings for the assignments. For this, we need to do HNCA, HNCOCA, HNCACB, HNCOCACB, HNCO and HNCACO experiments. Each set of experiments will provide information regarding the identity of each amino acid residue, observing the amino acid in the position i and in the position i-1, for instance. The information associated with the coupling and chemical shifts can be used to perform the assignment, with the assistance of the appropriate software; for example, CCPN Analysis V2.4.2.

The triple resonance analysis was performed observing the data from a pair of experiments: HNCA/HNCOCA, HNCACB/HNCOCACB and HNCO/HNCACO. Below, Figures 20, 21 and 22 summarize information regarding each of these experiments (spectra strips and magnetization transfer for NMR detection).

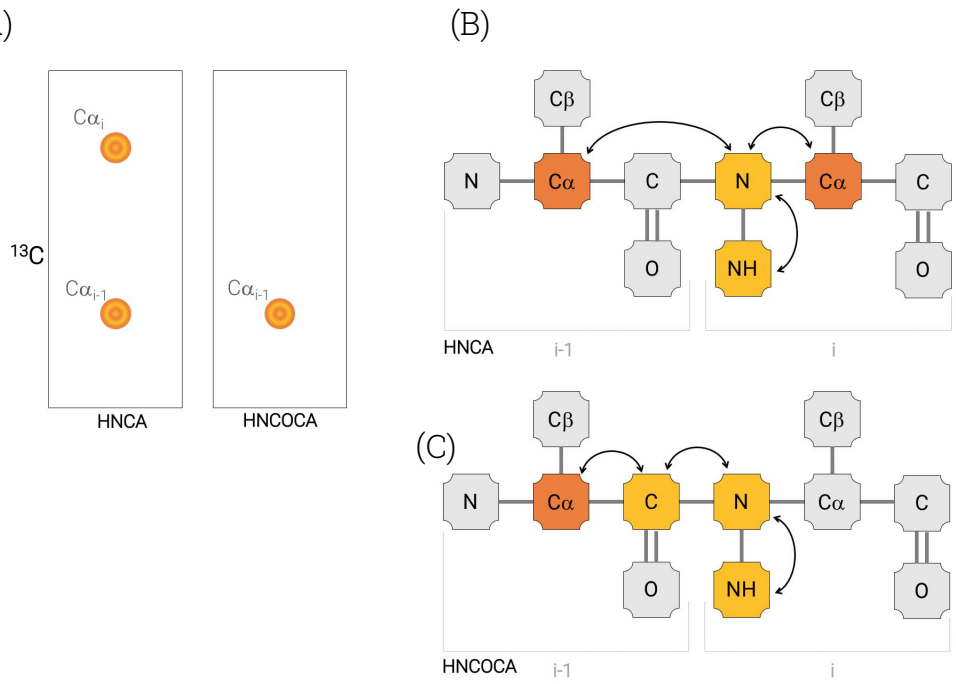


Figure 20: HNCA and HNCOCA diagrams, (A) HNCA and HNCOCA spectra strips, demonstrating the typical profile of each experiment; (B) HNCA magnetization transfer diagram; (C) HNCOCA magnetization transfer diagram.

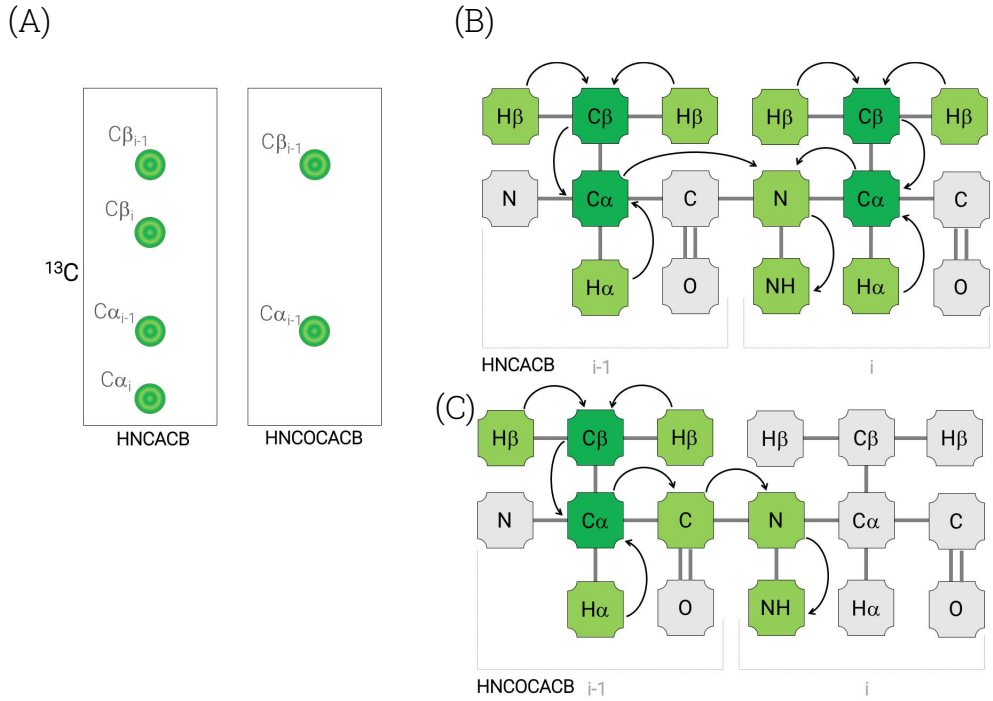


Figure 21: HNCACB and HNCOCACB diagrams, (A) HNCACB and HNCOCACB spectra strips, demonstrating the typical profile of each experiment; (B) HNCACB magnetization transfer diagram; (C) HNCOCACB magnetization transfer diagram.

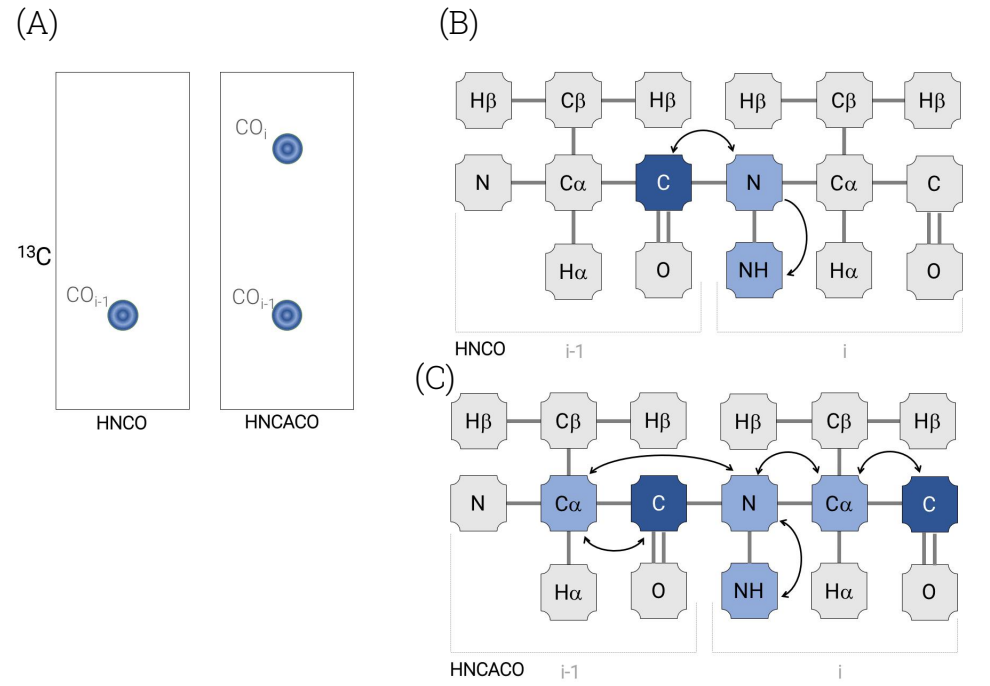


Figure 22: HNCO and HNCACO diagrams, (A) HNCO and HNCACO spectra strips, demonstrating the typical profile of each experiment; (B) HNCO magnetization transfer diagram; (C) HNCACO magnetization transfer diagram.

HNCA/HNCOCA experiments provide information regarding a sample's Cα. HNCA experiments display two peaks, one relative to the Cα of the amino acid residue at the i position, the second relative to the Cα at the i-1 position. The HNCOCA displays only the Cα at the i-1 position, allowing the identification of each residue. Since each HNCA spectrum strip has information regarding an amino acid and its predecessor, the combination of multiple strips, allied to the characteristic chemical shifts of certain residues, allows the linking of the strips to match the protein's primary structure. As an example, alanine commonly has its Cα at around 18 ppm, whereby and no other amino acid displays similar values at this range, allowing unequivocal the identification of alanine.

A similar approach is used with the HNCACB/HNCOCACB and HNCO/HNCACB pairs, in a such way that the HNCACO will provide information regarding Cα and Cβ at the position i, and also Cα and Cβ at position i-1. HNCOCB, however, displays only Cα and Cβ at the position i-1. In the last pair of experiments, HNCO/HNCACO, the HNCO peaks reveal only the carbonyls (CO) at the position i-1, while the HNCACO peaks correspond CO from both position i and position i-1.

These pairs of experiments complement each other, and their combination allows the assignment of an HSQC spectrum through the connection of multiple strips in a sequential way.

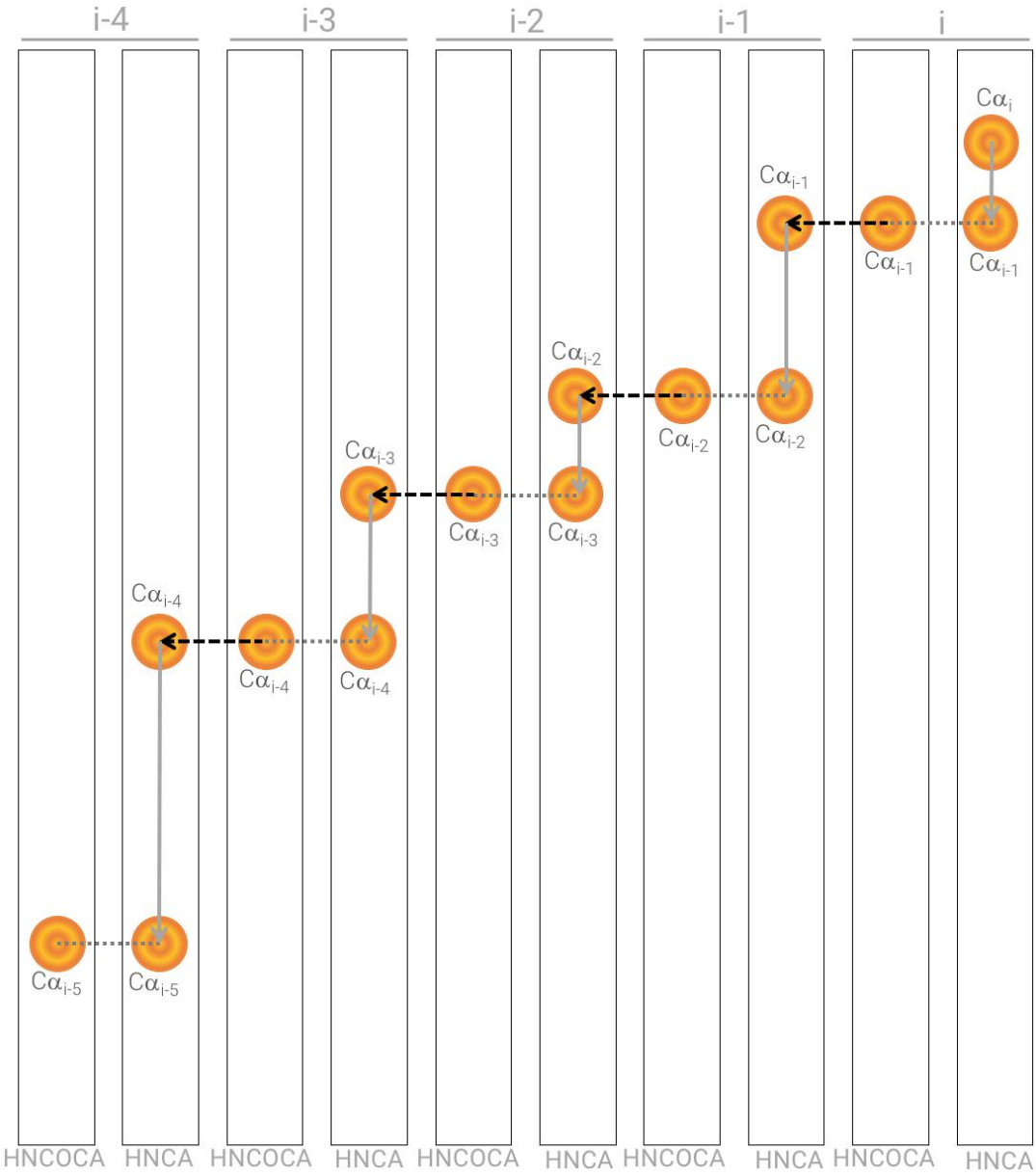


Figure 23: Diagram illustrating the process of amino acid residue assignment via triple resonance experiments.

Isotopically labelled proteins

The production of isotopically labelled proteins was performed using an Applikon ez-Control Bioreactor, in minimal media containing isotopically labelled reagents.

The minimal medium used has the following composition:

Isotopically labelled reagents:

-D-glucose (U13C6, 99%, Cambridge Isotope Labs)

-Ammonium Chloride (^{15}N, 99%).

Additional Components:

42.20mM Na_2HPO_4	5.55µM Myo-Inositol
22.0mM KH_2PO_4	2.45µM Pyridoxal.HCl
8.50mM NaCl	1.48µM Thiamine.HCl
18.70mM $^{15}NH_4Cl$	0.13µM Riboflavin
2.00mM $MgSO_4$	4.09µM Biotin
0.01mM $CaCl_2$	3.00µM $(NH_4)_6(Mo_7)$
0.01mM $FeSO_4$	0.40mM H_3BO_3
11.10mM ^{13}C-glucose	30.00µM $CoCl_2$
2.86µM Choline Chloride	0.10µM $CuSO_4$
0.90µM Folic Acid	0.80µM $MnCl_2$
2.28µM Pantothenic Acid	0.10µM $ZnSO_4$
4.09µM Nicotinamide	

The NMR experiments were performed, unless stated, in the following conditions:

-Buffer F	-Sample Volume: 300uL
-Temperature: 27˚C	-NMR tube: 10mm Shigemi tube D_2O matched
-Sample concentration: 50-200µM	-Deuterium Oxide concentration: 5%

All data analysis was performed using the software Bruker TopSpin 3.5pl7 and CCPN Analysis V2.4.2

Chapter 3 – eIF4E4 and PABP1 protein domains: production and purification

Protein production

The initial attempts for protein production were made using commercially produced synthetic DNA constructs, manufactured by the company DNA2.0. These constructs were designed to encode for *Leishmania* eIF4E4 and PABP1, as well varied domains of each (please see details in the Annexes). There were several attempts for the production of these proteins, in several conditions, but without success. Other strategies were then used in order to obtain the target proteins, for instance the preparation of fusion proteins with MBP-12His tag. The MBP tag can be used to increase protein solubility (ALEXANDROV, 2001), and it contains a C-protease cleavage site. It was possible to obtain a number of encoded proteins in these conditions; however, it was not found possible to cleave and remove the MBP tag.

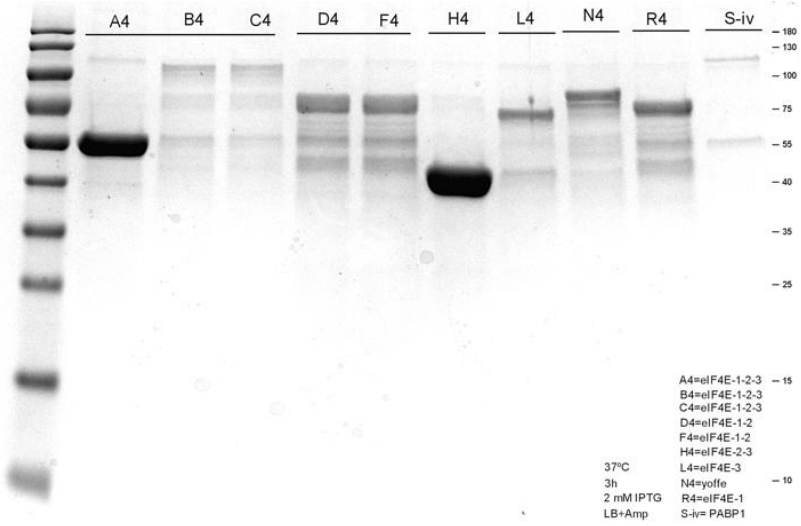


Figure 24: SDS-PAGE gel with the different MBP-fusion constructs of eIF4E4 and PABP1.

The next strategy involved substitution of the MBP tag by the GST tag. The GST-tagged constructs were less successful than those made with its MPB counterpart, and also displayed the same C-protease cleavage issues, rendering it unattractive for use in binding experiments.

Subsequently, the constructs were cloned into the pET22b vector, and it was found possible to obtain two His-tagged constructs, eIF4E4(i) and eIF4E4(iv). The eIF4E4(i) corresponds to the eIF4E4 N-terminal region, and it was used for antibodies production, but since the main objective was to identify the binding motif, eIF4E4(iv) was a better protein domain, due to its shorter sequence, especially since it had previously been described in the literature as a potential binding region in eIF4E4 (ZINOVIEV, 2012). After eIF4E4(iv) was obtained and optimized, protein production was initially performed by growing *E. coli* cultures in 2L flasks, in the conditions described in the Material and Methods. eIF4E4(iv) displayed a marked tendency to aggregate, protein yield was generally low. In order to help address these issues, the incubation temperature was lowered to 16˚C, giving rise to a limited improvement in the aggregation behaviour. Detergents, reducing agents and different salt concentrations were also tested, but little further improvement was observed.

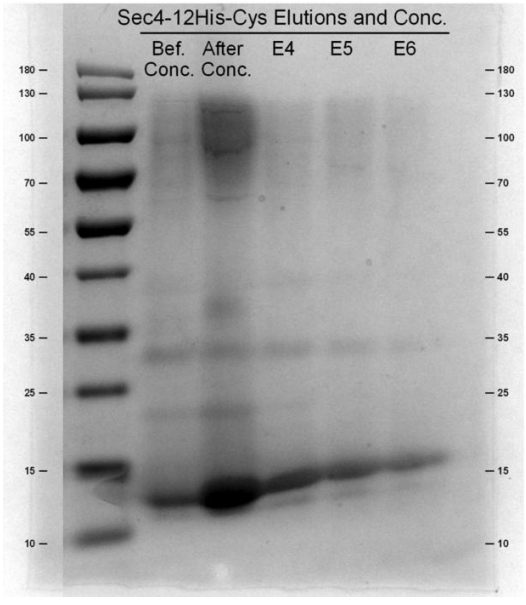


Figure 25: SDS-PAGE gel of eIF4E4(iv), - note the increased protein aggregation after concentration.

The utilization of a bioreactor increased substantially the protein yield. Cultures grown in flasks in incubators had a pellet yield of around 4g/L of medium, while cultures grown in bioreactors had this yield increased 7-10-fold. The resulting increase in the size of the pellet generated a greater amount of monomeric protein, but there was no increase in aggregation. Figure 26 shows the chromatogram of eIF4E4(iv) after size exclusion chromatography.

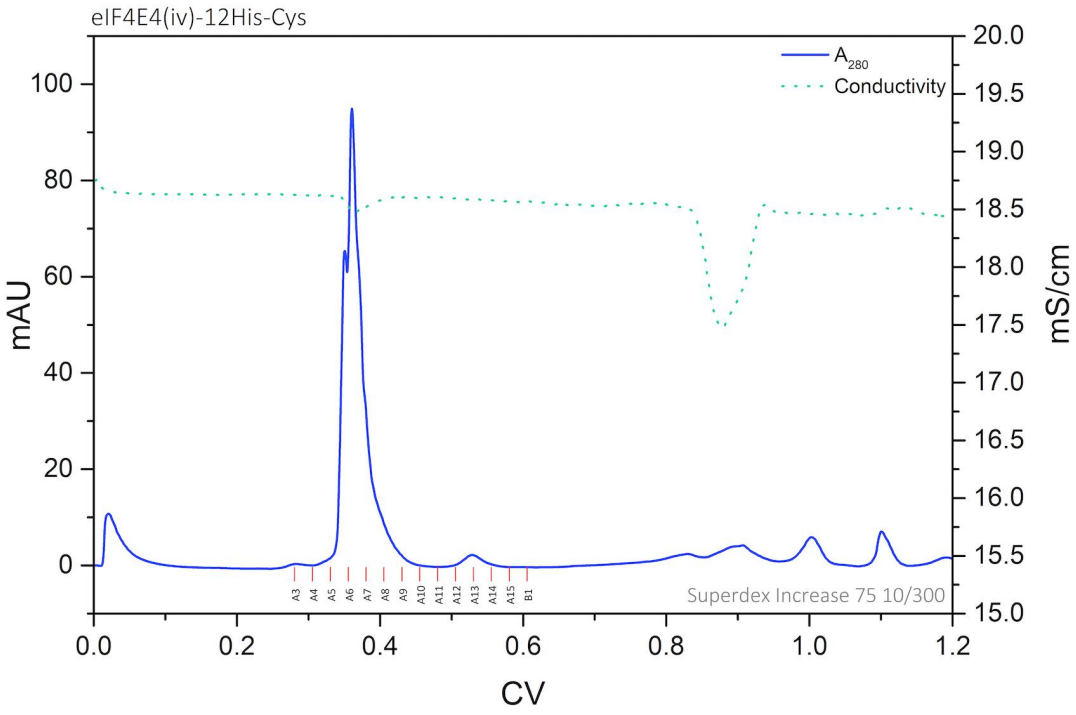


Figure 26: Size exclusion chromatogram of eIF4E4(iv), in Superdex Increase 75 10/300 GL column. Aggregated protein peak located between the fraction A5-A9; monomeric protein peak between fraction A12-A15. Void volume = 7.2 mL, 1 CV= 24 mL; molecular weight range= 3-70 kDa.

The large peak corresponding to fractions A5-A10 contains aggregated protein, while in a small peak in the fractions A11-A15 corresponds to monomeric protein. The column Superdex Increase 75 10/300 has a void volume of 7.2 mL, 1 CV is 24mL and the separatable molecular weight range lies between 3-70kDa; therefore, anything larger than 70 kDa is eluted before 7 mL, while anything smaller that 3 kDa leaves the column

after 24mL. As illustrated in Figure 26, the majority of protein is aggregated, requiring a larger amount of cell culture or more efficient production (as achieved using bioreactors) in order to obtain appropriate amount of monomeric protein. Figure 27 displays the SDS-PAGE gel for this size exclusion purification.

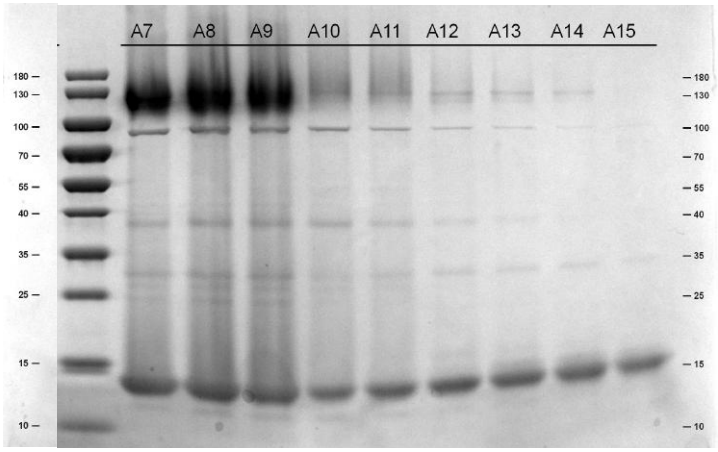


Figure 27: SDS-PAGE gel of eIF4E4(iv), in Superdex Increase 75 10/300 GL column. Aggregated protein peak located between the fraction A7-A9; monomeric protein peak between fraction A10-A15.

The protein fractions containing monomeric protein were collected, joined in a single sample, and concentrated. The protein concentration was determined by the use of SDS-PAGE BSA standards, as shown in Figure 28.

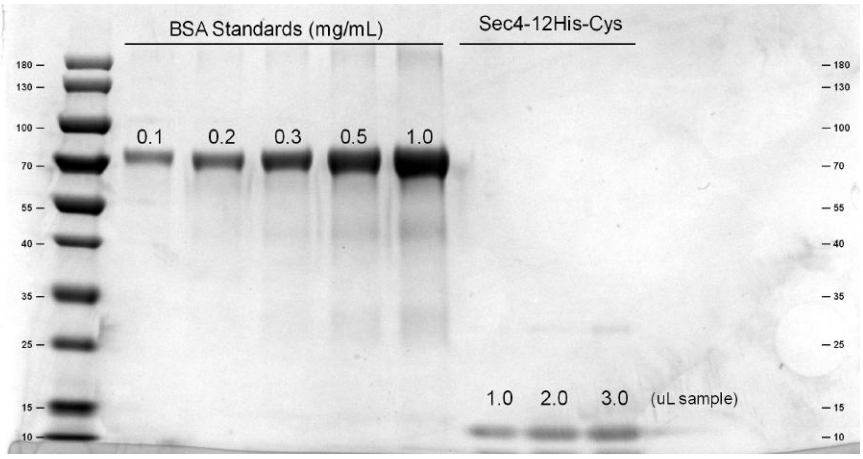


Figure 28: SDS-PAGE gel of eIF4E4(iv) concentration determination, using BSA standards. Protein concentration is determined by comparing the sample band intensity with a calibration curve obtained by the BSA standards.

Isotopically labelled protein production

The production of isotopically labelled protein utilized minimal medium, as described in the Material and Methods. The protein yield was superior to that obtained in rich medium and was suited to requirements.

The large peak between fractions A3-A10 corresponds to aggregated protein, while the small peak between the fractions A11-A14 corresponds to monomeric protein. Figure 30 displays the SDS-PAGE gel for this size exclusion chromatography.

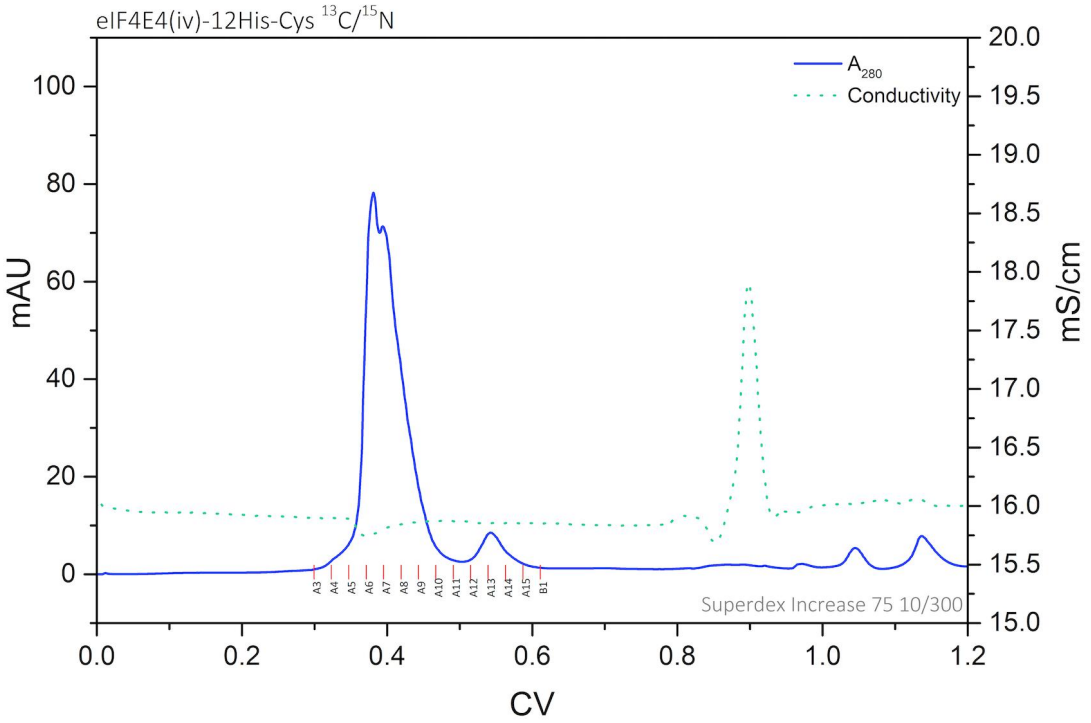


Figure 29: Size exclusion chromatogram of isotopically labelled eIF4E4(iv), in Superdex Increase 75 10/300 GL column. Aggregated protein peak located between the fraction A3-A10; monomeric protein peak between fraction A11-A14.

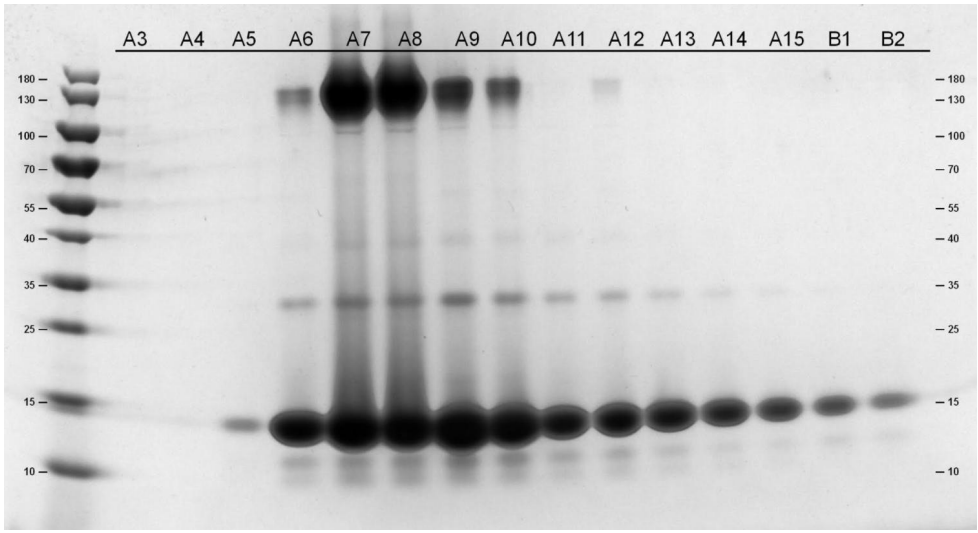


Figure 30: SDS-PAGE gel of isotopically labelled eIF4E4(iv), in Superdex Increase 200/300 GL column. Aggregated protein peak located between the fraction A10-A14; monomeric protein peak between fraction A15-B3.

The isotopically labelled protein fraction containing monomeric protein were also collected, joined in a single sample and concentrated. The protein concentration was determined by the use of BSA standards, as shown in Figure 31, side-by-side to non-labelled eIF4E4(iv).

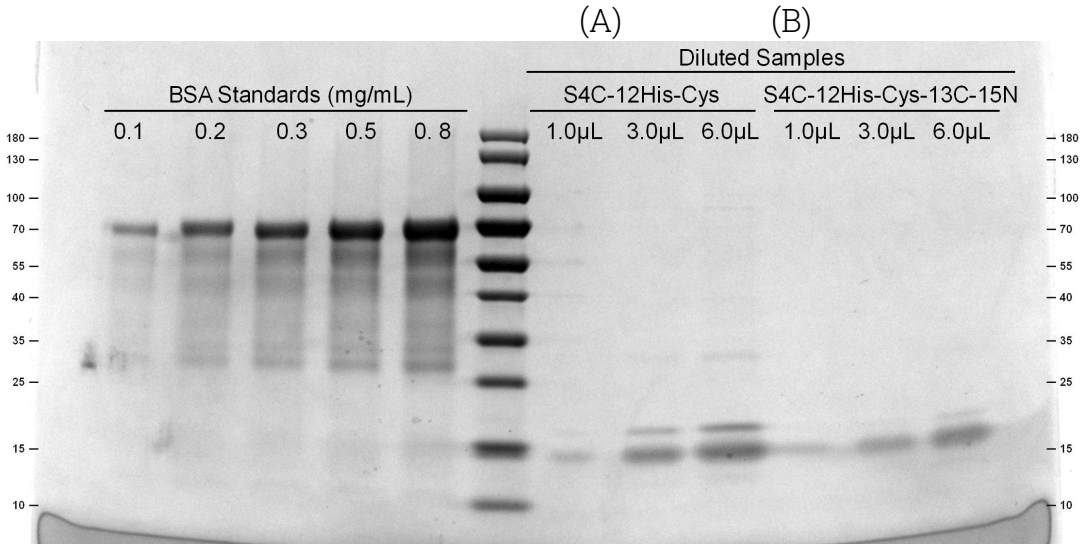


Figure 31: SDS-PAGE gel of (A) non-labelled eIF4E4(iv); (B) isotopically labelled concentration determination, using BSA standards.

PABP1 domains were prepared by another group member, Dr Maja Firczuk, being overproduced in *E. coli* with an N-terminal 12His-tag, and in most cases purified using cobalt affinity, followed by size exclusion chromatography, as described in the methods. For a few proteins, like full length PABP1, PABP1(A) and PABP1(B), the yield was too low to proceed with size exclusion chromatography.

Protein identification was also performed by mass spectrometry, through gel digestion protocol, as described in the methods. The mass spectrometry data confirmed the protein identity, allowing their confident use in the interaction and structural experiments.

Chapter 4 – Surface Plasmon Resonance

4.1: "Yes/No" Binding Experiment

Surface plasmon resonance was the first method used in the investigation of eIF4E4/PABP1 protein-protein interactions. The first strategy used was the initial screening of PABP1 domains against eIF4E4(iv). This initial screening used eIF4E4(iv) immobilized on a CM5 chip, through amine coupling (as described in the Materials and Methods). This experiment was conducted in a 2-1, 4-3 mode (Figure 32) *i.e.* cells 1 and 3 were used as reference, while cells 2 and 4 were used in target immobilization, in such way that the reference data was subtracted from the target/ligand data. A range of different ligands are then screened, verifying which among them bind to the immobilized target protein (Figure 32).

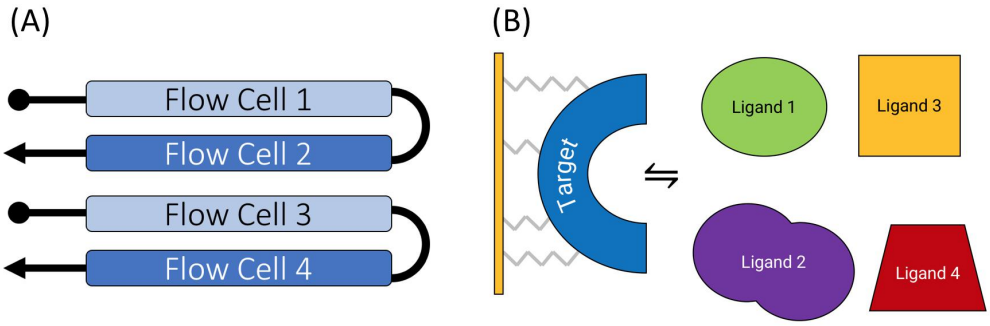


Figure 32: (A) Flow cell modes used in the Yes/No binding experiments; (B) schematic of the Yes/No binding screening experiment.

This kind of experiment is very useful in assessing which members of a varied group of ligands (in this case, PABP1 domains) binds to a target (eIF4E4(iv). The immobilization strategy, amine coupling, is very simple, allowing the immobilization of protein to a chip surface; however, there

is a drawback in this approach, since it binds the target protein in an isotropic manner, which can restrict the availability of the potential binding domain. Therefore, this strategy is very useful for qualitative screening, and provides a good means of detecting binding *per se*, although not with the highest precision.

At this first stage, we were interested in identifying the PABP1 regions that would bind to eIF4E4(iv), so a Yes/No binding experiment with an amine coupling immobilization was an adequate technique. We chose a number of PABP1 domains, as follows: full length PABP1, A, B, C, D, F, G, I, J (Figure 33) and used BSA and running buffer as negative controls; while immobilizing eIF4E4(iv) on the chip surface. Some of the domains gave low production yields, preventing additional purification by size exclusion chromatography.

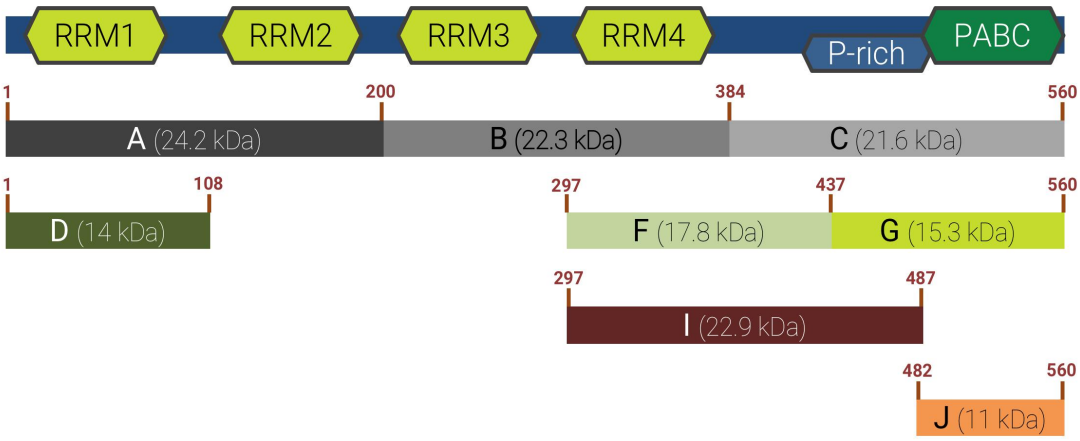


Figure 33: PABP1 domains used for Yes/No binding experiments.

The results obtained in this experiment are displayed in Figure 34:

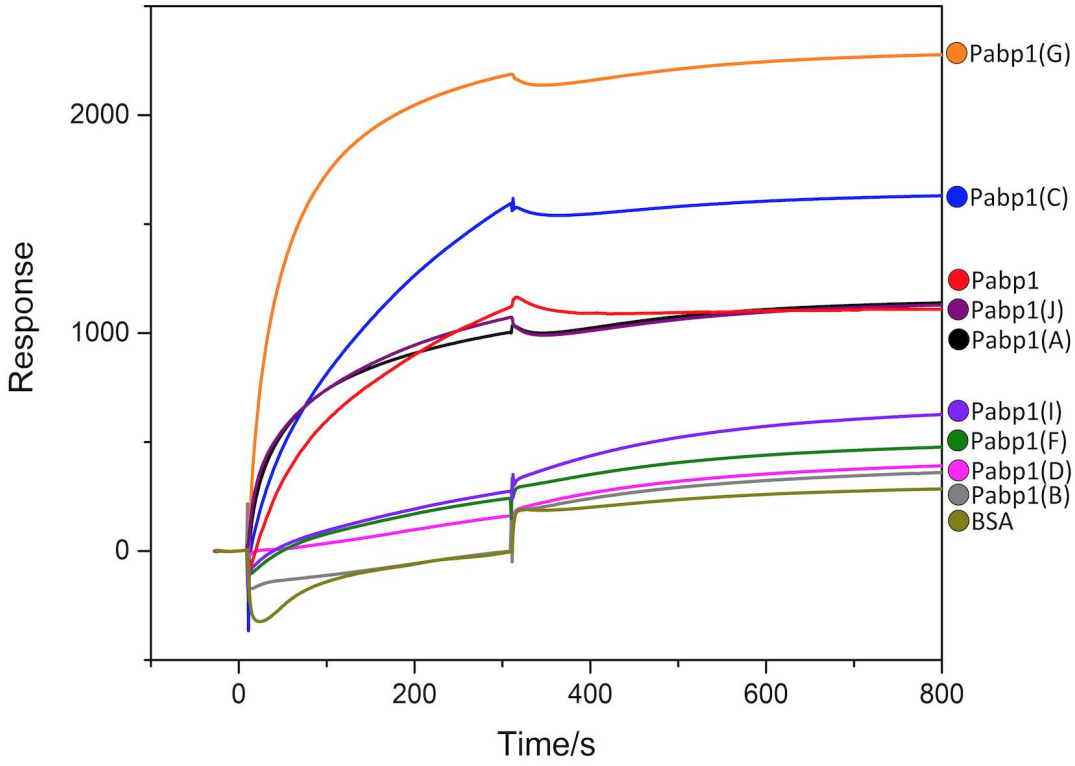


Figure 34: Yes/No binding sensorgram of PABP1 domains against eIF4E4(iv).

Due to the qualitative nature of this experiment, it was performed using the highest concentration available for each of the proteins (please see table in Annexes) in order to maximize the observation of binding to eIF4E4(iv). eIF4E4(iv) was immobilized by amine coupling to a concentration of 1000 units, and then screened against the PABP1 proteins.

The data displayed in the sensorgram indicates binding for full length PABP1 and for its C-terminal domains, C, G and J. Surprisingly, it also indicated that one N-terminal domain, A, also binds to eIF4E4(iv). The binding of full length PABP1, and especially its C-terminal domains supports the hypothesis of PABP1/eIF4E4 interaction. The negative controls, BSA and running buffer did not show any binding.

It was noted that the curves showed limited dissociation, which could have indicated the occurrence of very strong binding (tight binding with very low dissociation), or a contribution of unspecific binding. In order to further investigate these interactions, binding affinity experiments were then performed.

4.2: Binding Affinity Experiments

Binding affinity experiments, differently from the Yes/No binding detection approach, use a single ligand, at varied concentrations, to evaluate its affinity towards a target. In these experiments, a target (in this case eIF4E4(iv)) is also immobilized on a chip surface, and different ligand concentrations are tested in duplicates. We tested several PABP1 domains, and the most relevant results are shown in this chapter (further data are available in the Annexes).

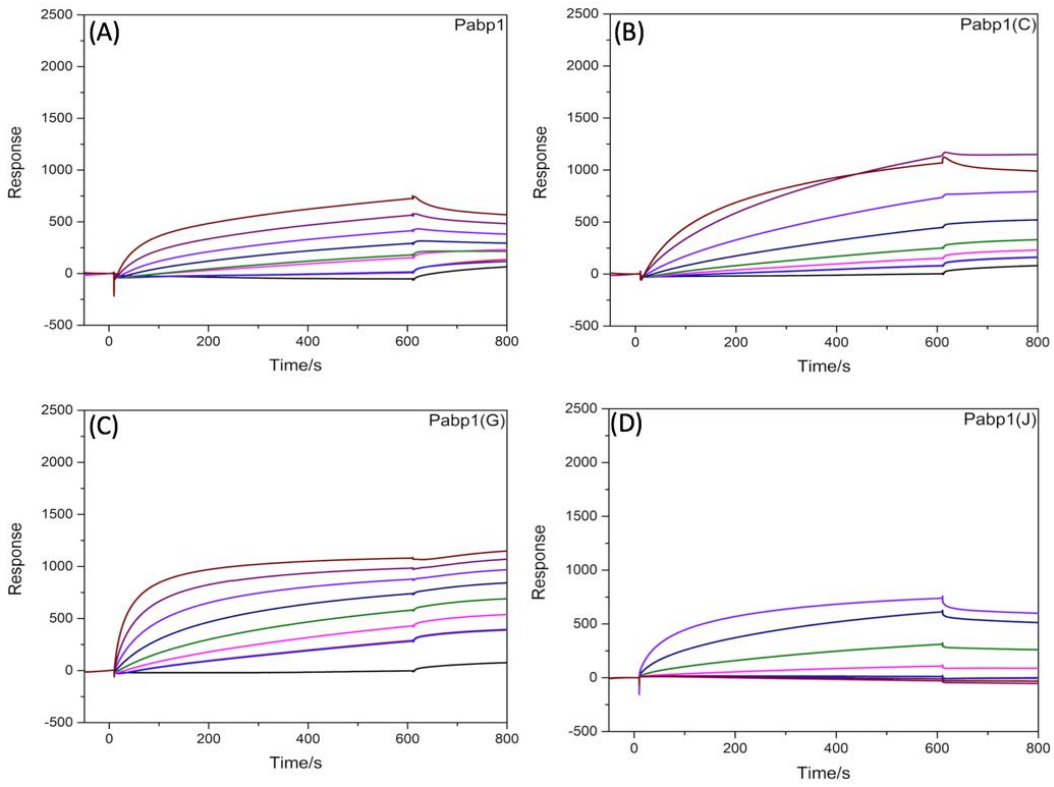


Figure 35: Binding affinity sensorgrams of eIF4E4(iv) against: ***(A)*** *PABP1 full length;* ***(B)*** *PABP1(C);* ***(C)*** *PABP1(G);* ***(D)*** *PABP1(J).*

The binding affinity experiments showed results that agreed with the Yes/No binding experiments. The sensorgrams suggested the occurrence of binding, however, there was a lack of significant dissociation in the sensorgrams, feature that must be considered for K_D calculations.

There are two main methods to calculate affinity using SPR, the kinetic method and the thermodynamic method. Considering a two components system, protein/ligand, in equilibrium the following equations applies:

$$[P] + [L] \rightleftharpoons [PL] \text{ and } \frac{[PL]}{[P][L]} = \frac{k_a}{k_d} = K_A$$

And therefore:

$$K_D = \frac{1}{K_A} \therefore K_D = \frac{k_d}{k_a}$$

Where K_A is the association constant, K_D is the dissociation constant, k_d dissociation rate constant and k_a the association rate constant.

Considering the equations above, it is possible to observe that by evaluating the association and dissociation rates in a sensorgram, it is possible to calculate affinity, by fitting data into the curves.

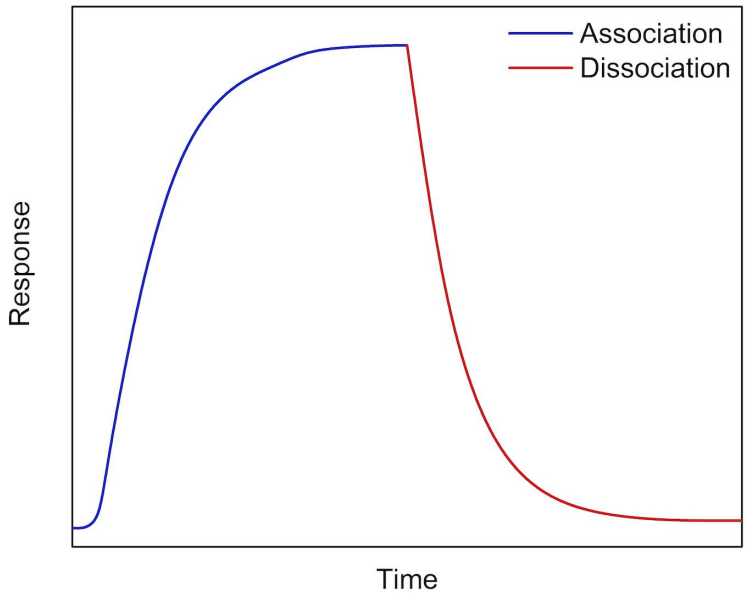


Figure 36: Example of a SPR curve, with ligand association phase highlighted by the blue line, and the dissociation phase highlighted in red.

The other method, thermodynamic, considers the equipment response to a determined ligand concentration, and for instance, K_D is equivalent to the ligand concentration that has a response of half of the maximum response (R_{max}) obtained in a measurement (K_D is [L] whose response is $R_{max}/2$).

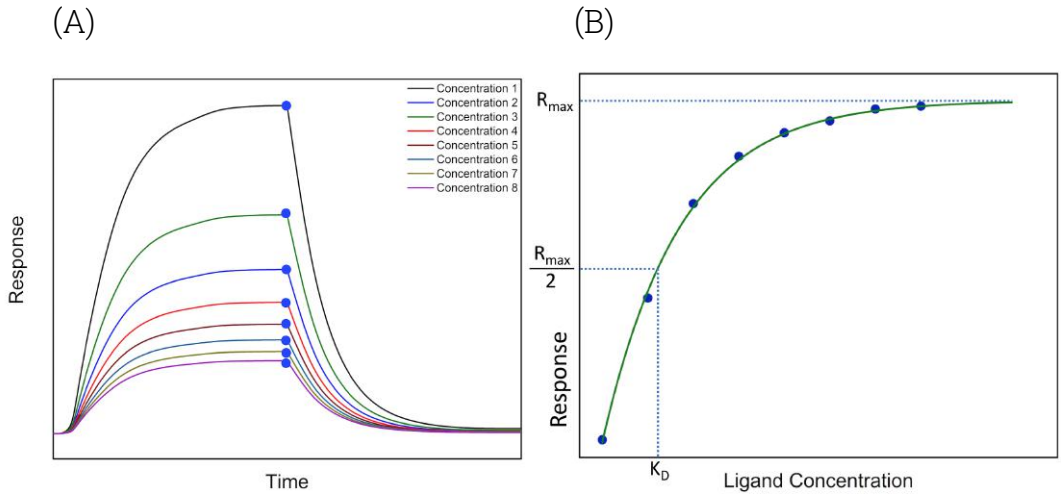


Figure 37: (A) example of SPR curves in experiment to obtain affinity values; (B) K_D calculation using data extracted from SPR curves.

The kinetic strategy for calculating K_D is not appropriate for the SPR dataset obtained in the 'Yes/No experiments, since the conditions for dissociation had not been optimised.

The K_D calculation using the thermodynamic approach, however, can be performed using only the association segment of an SPR curve. Table 3 display the K_D values obtained for the most relevant proteins tested (data regarding the remaining domains are in the Annexes). These should only be taken as indicative values.

Table 3: K_D calculations of some of the tested proteins.

Ligand	Calculate K_D/µM
PABP1	0.3
PABP1(A)	1.1
PABP1(C)	0.4
PABP1(G)	0.4
PABP1(J)	0.8

The data obtained suggest that the affinity between eIF4E4(iv) and PABP1 is in the low µM range. However, the unusual behaviour regarding protein dissociation suggest that complementary K_D calculations using alternative techniques may be required to increase confidence in these results.

Similar experiments were also performed, using eIF4E4(iv)-SBP constructs and SA chips (streptavidin coated chip, that allows, for example, the immobilization of SBP-tagged proteins, where SBP stands for Streptavidin Binding Protein), but it was not possible to immobilize the protein to satisfactory levels.
NTA chips (Nitrilotriacetic acid coated chips, that allow the immobilization of His-tagged proteins) were also attempted, but the ligand (in this case, eIF4E4(iv)-SBP) displayed a significant unspecific binding to the chip surface (including in the blank, control channels), preventing the use of this strategy.

More detailed investigations were subsequently undertaken to optimise the conditions for SPR in the case of eIF4E4(iv) vs PABP1. For this, the flow rate was increased to 60µL/min, and temperature set to 30˚C.

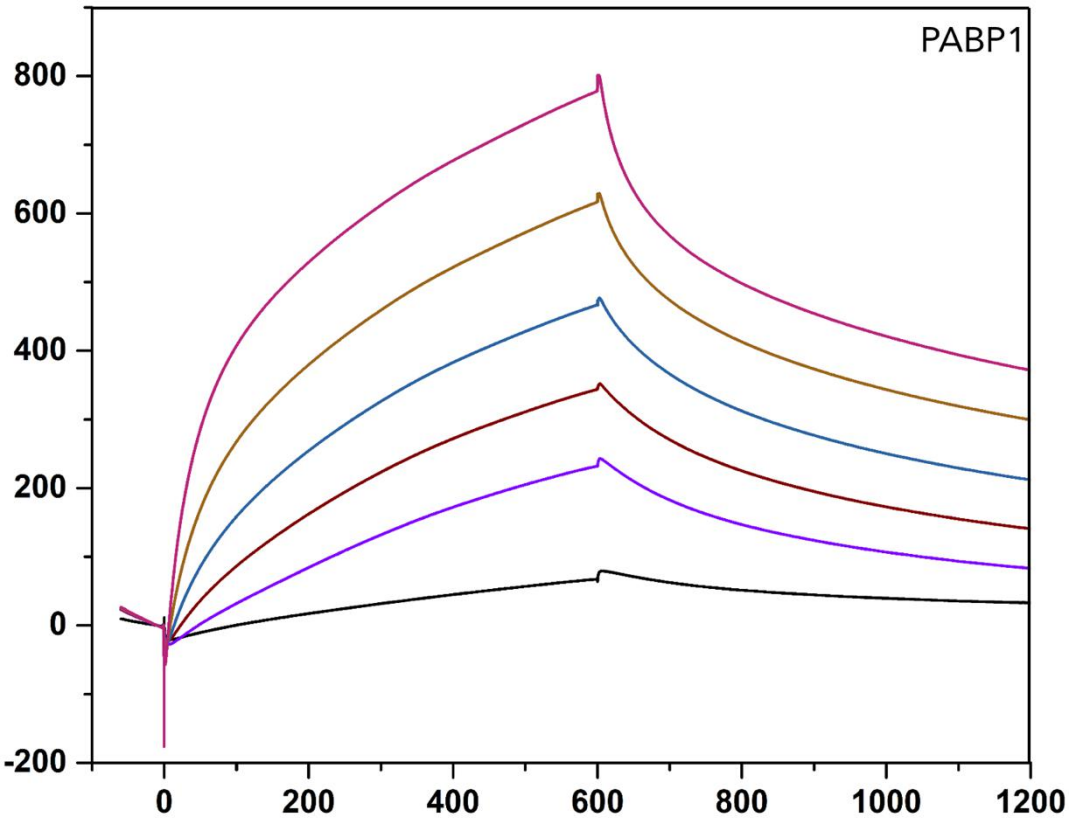


Figure 38: Binding affinity sensorgram of eIF4E4(iv) against PABP1 full length, under optimized conditions.

The results revealed that it is possible to achieve a more balanced association and dissociation curves (Figure 38). However, the initial SPR experiments had served their purpose in that they provided a guide to the locations of key binding sites, and attention now turned to the utilization of a parallel technique.

Chapter 5 – Microscale Thermophoresis

5.1: Binding Affinity Experiments

MST binding experiments use a fluorescent-tagged target molecule and a ligand. For this, the target molecule has its concentration constant, while the ligand is tested in various concentrations (typically 16 different concentrations $[L]_1$ to $[L]_{16}$, in a standard MST run). Glass capillaries are filled with target molecule/ligand solutions, and then undergo MST analysis.

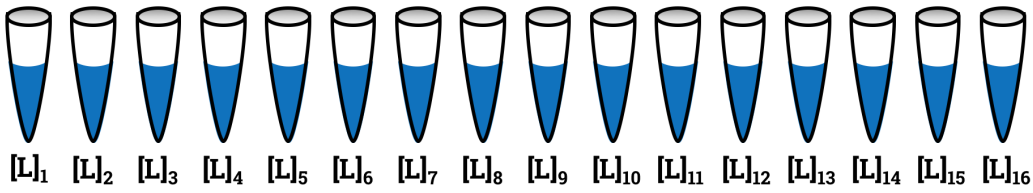


Figure 39: MST sample preparation, the target molecule concentration is kept constant, while the ligand concentration, [L], is varied, with $[L]_1 > [L]_2 > \ldots > [L]_{16}$.

MST experiments rely on the thermal diffusion of a complex. Before thermophoresis starts (initial state), the molecules as distributed randomly in the bulk of the solution contained in the capillaries. When an IR laser is turned on, a temperature gradient builds up in the solution, and typically, molecules move away from the laser focusing point (although it is also possible that molecules move towards the temperature gradient, in a process known as negative thermophoresis). Since the target protein is fluorescent, their local concentration near the temperature gradient is reduced, and this leads to a decrease in the fluorescence signal measured at this spot. When the IR laser is turned off,

the temperature gradient dissipates, and the molecules then can return to the initial area. Since the mass of a molecule affects its mobility in the temperature gradient, there is a difference between the free target molecule mobility and the target/ligand complex mobility. This variation in motion is used in MST measurements to calculate affinity, since the different ratios of target/ligand contained in the capillaries allow the calculation of affinity curves.

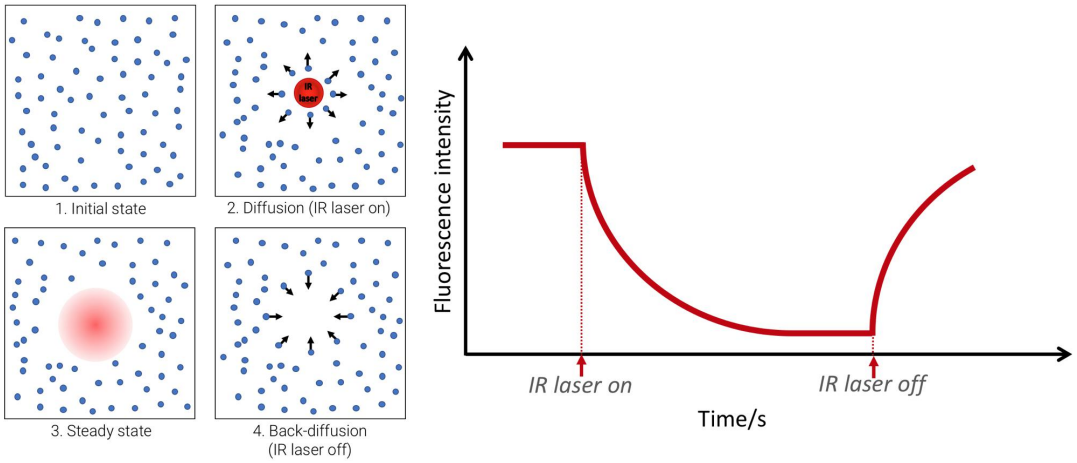


Figure 40: Microscale thermophoresis schematic (based on Jerabek-Willemsen, 2014). There are four main stages during the MST measurement, 1: Initial state; 2: Diffusion; 3: Steady state and 4: Back-diffusion. The effect of this stages is reflected in the fluorescence intensity over time, as observed in a typical MST curve.

In this study, PABP1 domains were tested (at least in duplicates) against eIF4E4(iv), the fluorescent-labelled target protein; their relative affinities are shown in Figure 41.

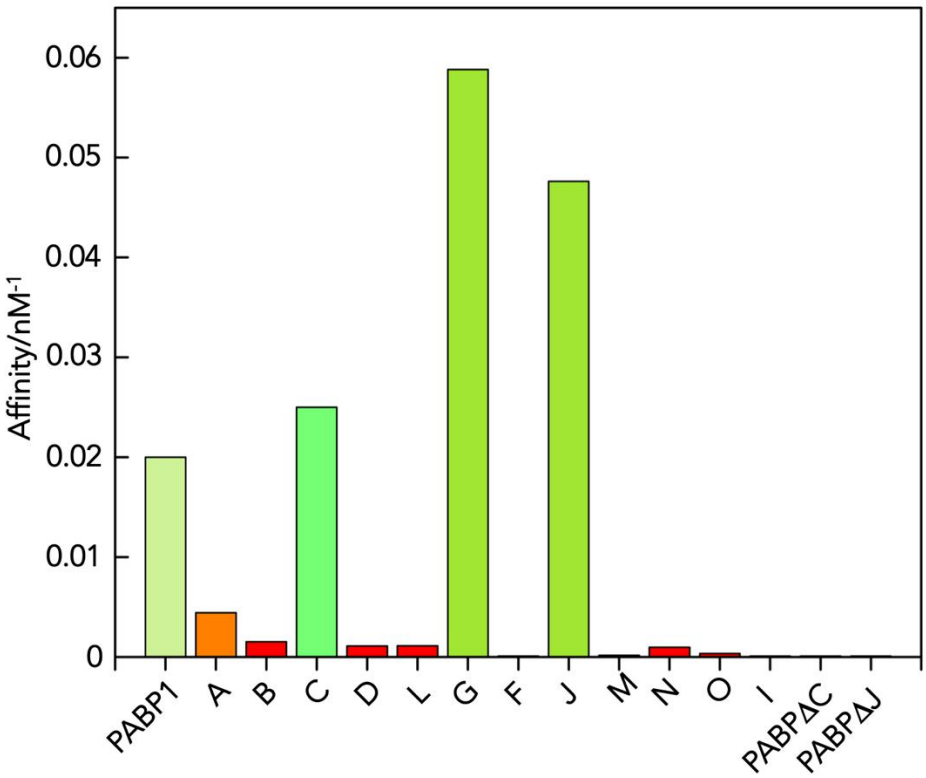


Figure 41: Relative affinity of PABP1 domains against eIF4E4(iv).

The MST data show a higher affinity for PABP1 C-terminal domains (PABP1(C), PABP1(G) and PABP1(J)), and for full-length PABP1. PABP1(A) shows weak binding, significantly smaller than observed for the full-length protein and the C-terminal domains. These results suggest that the C-terminal region of PABP1 is involved in binding with eIF4E4.

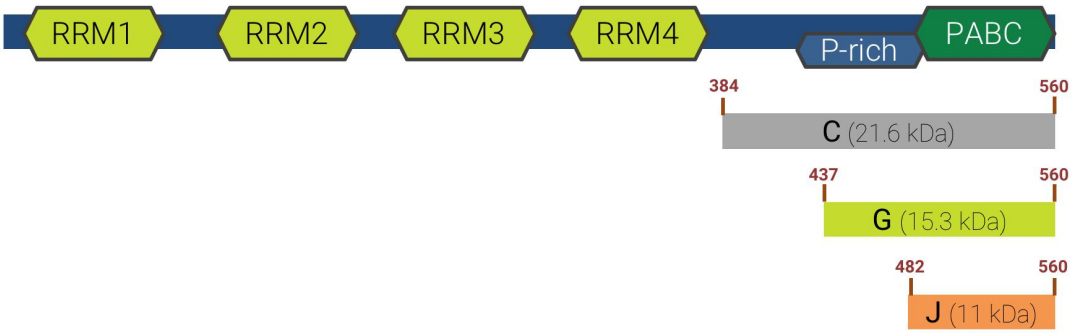


Figure 42: PABP1 full length and C-terminal domains with higher affinities for eIF4E4(iv).

In order to further investigate the importance of PABP1 C-terminal in binding to eIF4E4(iv), we used deletion constructs, where different parts from PABP1 C-terminal domains were removed (Figure 43 illustrates these deletion domains).

Figure 43: Representation of PABP1 C-terminal deletion constructs.

We expected that if the PABP1 C-terminal region interacts with eIF4E4(iv), no binding curve would be observed, giving additional evidence that PABP1 C-terminal is involved in binding with eIF4E4.

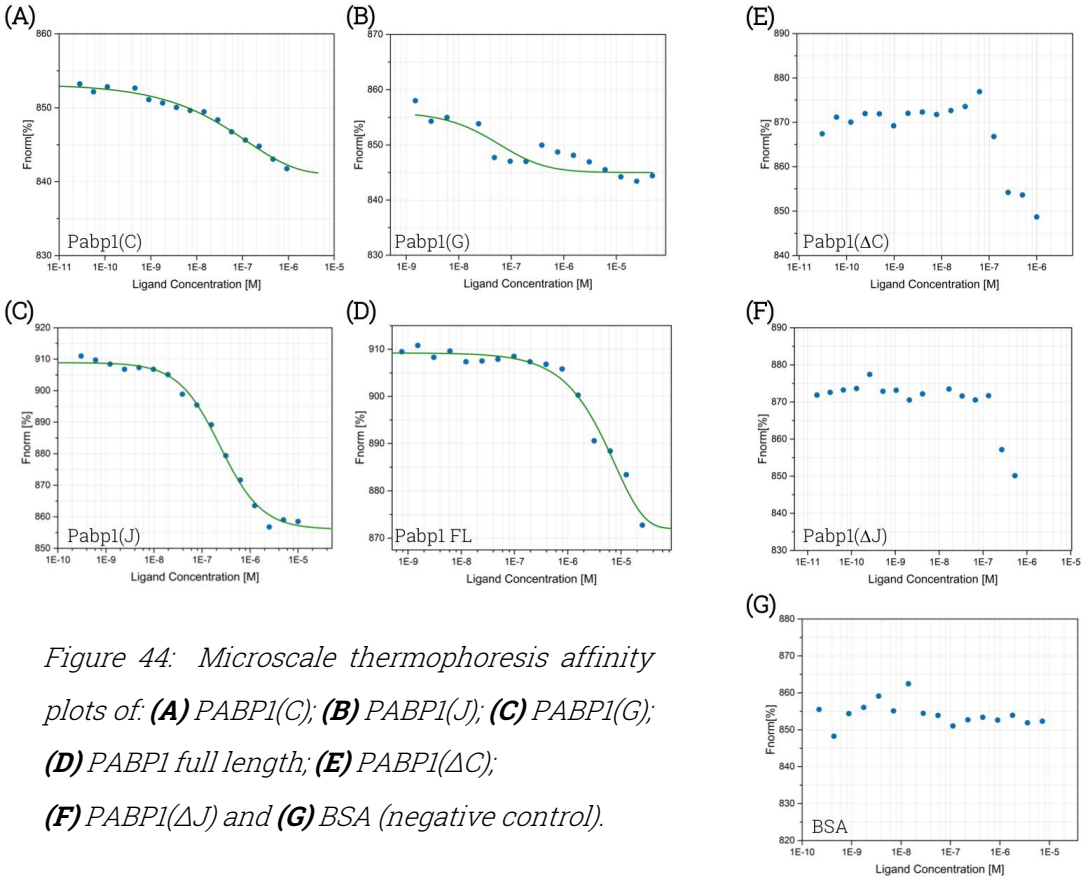


Figure 44: Microscale thermophoresis affinity plots of: ***(A)*** *PABP1(C);* ***(B)*** *PABP1(J);* ***(C)*** *PABP1(G);* ***(D)*** *PABP1 full length;* ***(E)*** *PABP1(ΔC);* ***(F)*** *PABP1(ΔJ) and* ***(G)*** *BSA (negative control).*

The affinity values obtained for the PABP1 C-terminal domains were in the nM range, and a summary of the obtained values are found on Table 4.

Table 4: PABP1 domains/eIF4E4(iv) K_D calculations by Microscale Thermophoresis.

Protein	K_D / nM	Protein	K_D / nM
PABP1(A)	226	PABP1(L)	891
PABP1(B)	659	PABP1(M)	5860
PABP1(C)	40	PABP1(N)	1020
PABP1(D)	900	PABP1(O)	2800
PABP1(F)	No binding	PABP1 FL	50
PABP1(G)	17	PABP1(ΔC)	No binding
PABP1(I)	No binding	PABP1(ΔC)	No binding
PABP1(J)	21	BSA	No binding

The MST data also allow the estimation of the stoichiometry involved in eIF4E4/PABP1 binding, through the calculation of Hill coefficients. The data fitting to the Hill Equation (GOUTELLE, 2008) (using a Logistic non-linear curve fit) revealed a Hill Coefficient close to 1 (Table 5), suggesting a noncooperative, independent binding mode.

Table 5: Hill coefficient for the C-terminal domains of PABP1, when binding to eIF4E4(iv).

Target	Ligand	Hill Coefficient
eIF4E4(iv)	Full length PABP1	0.85
eIF4E4(iv)	PABP1(C)	0.84
eIF4E4(iv)	PABP1(G)	1.08
eIF4E4(iv)	PABP1(J)	0.95

These data suggest that the binding ratio between eIF4E4(iv) and PABP1 is 1:1, in such way that the data indicate that only one molecule of each protein is required for the complex formation.

The MST experiments pass through a number of quality checks that verify several parameters that are used to assess the measurement reliability. For example, it is important to verify for deformities of the capillaries (what would lead to faulty measurements). This is achieved by measuring the fluorescence intensity along length of the capillaries inside the inside the equipment; good, well-shaped capillaries display a Gaussian shaped fluorescence curve symmetrically centered on the capillary axis. Capillary evaluation also provides information regarding the occurrence of protein adsorption inside the capillary wall (affecting measurement accuracy, since adsorption of the target fluorescent-tagged protein leads to fluorescence intensity inhomogeneity). The occurrence of adsorption is identified by the occurrence of "split" peak on the capillary peak format, and it is verified at the start and end of each measurement.

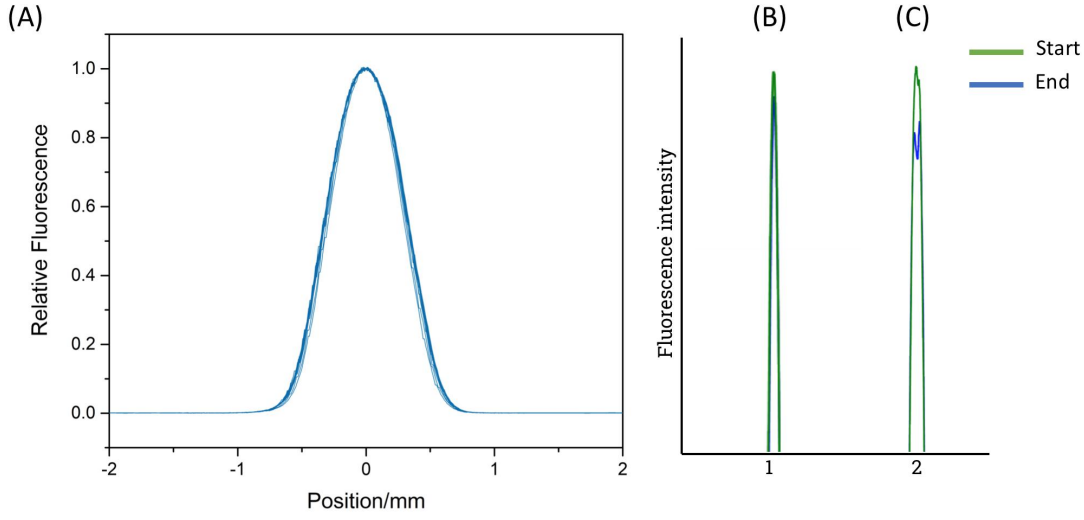


Figure 45: (A) Capillaries shape quality check; (B) capillary without protein adsorption; (C) capillary with protein adsorption, with higher degree of adsorption observed at the end of measurement.

The presence of aggregated protein can also affect the accuracy of MST measurements. An ideal MST measurement displays smooth curves that follow the typical thermophoretic profile; however, if protein aggregates are present, these curves show bumps and a chaotic behaviour, deviating from the typical profile.

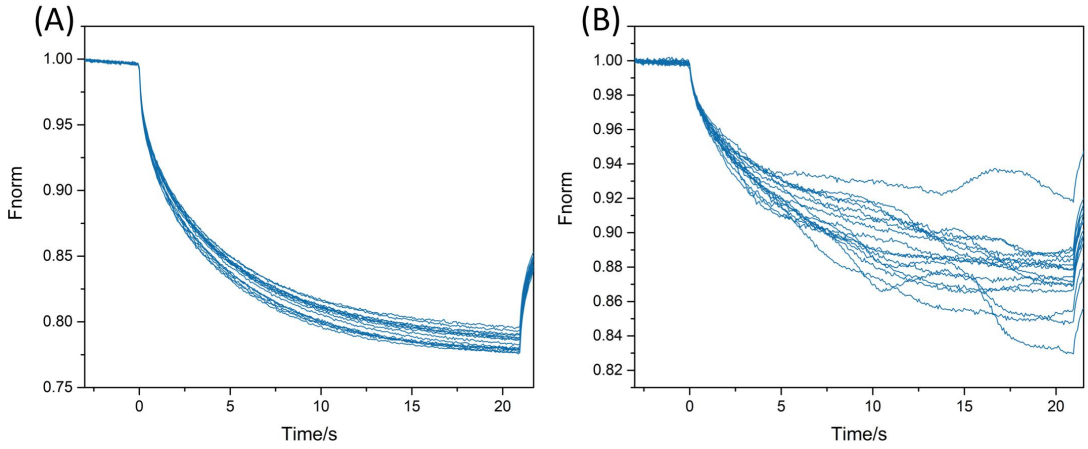


Figure 46: (A) MST curves of an aggregate-free sample, displaying smooth curve; (B) MST curves of an aggregate-containing sample, displaying bumps and erratic behaviour.

Another important parameter that needs to be evaluated is the occurrence of ligand-induced photobleaching. The MST experiment measures fluorescence intensity in two moments during thermophoresis; the initial measurement during the "cold" stage (show by the blue bars in Figure 47), and the second measurement during the "hot" stage (show by the red bars in Figure 47). The fluorescence levels during the "cold" stage must be similar along the different capillaries (and therefore in the different ligand concentrations). Variations in this stage indicate that the ligand is inducing fluorescent-tagged target photobleaching, and although the data obtained can provide qualitative data (i.e. the occurrence or not of binding), the accurate calculation of affinity constants is difficult.

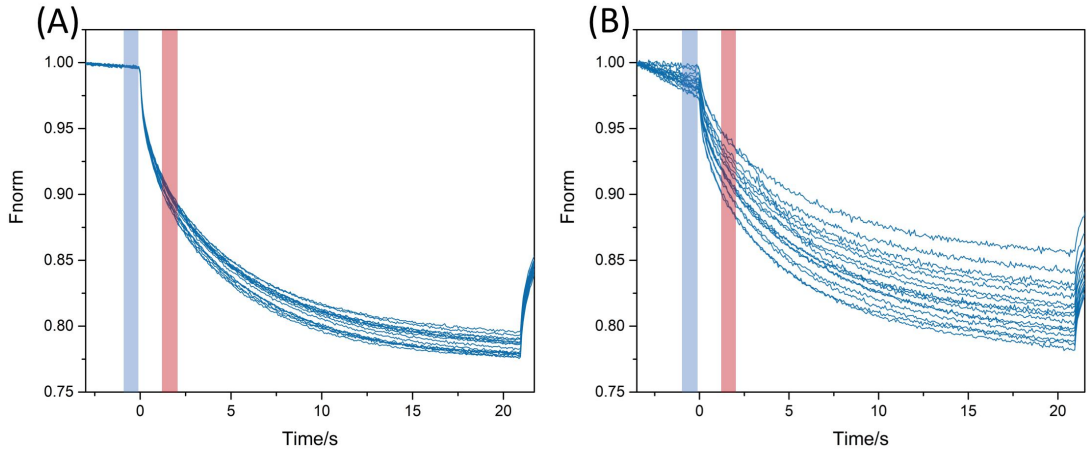


Figure 47: (A) MST curves without the occurrence of ligand induced photobleaching, observed by constant fluorescence levels in the blue bar area; (B) MST curves with the occurrence of ligand induced photobleaching, observed by varied fluorescence levels in the blue bar area.

The implementation of quality checks on different measurement parameters give greater confidence in the data obtained by MST.

Supplementary to the eIF4E4/PABP1 studies, the interactions of these proteins with eIF4G3 were also investigated. There are a few previous reports in the literature (DA COSTA LIMA, 2010; FREIRE, 2017) describing the interaction between PABP1 and eIF4G3. However, this interaction is described as being very weak, or even non-existent (FREIRE, 2017; DE MELO NETO, 2015), thus leaving uncertainty as to whether these two proteins really bind to each other. For this reason, another eIF4E4 derivative construct was prepared, eIF4E4(v) (Figure 48).

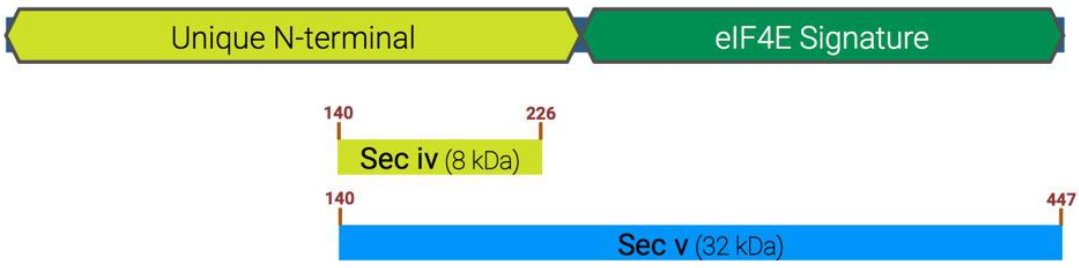


Figure 48: eIF4E4 full length and sectors eIF4E4(iv) and eIF4E4(v).

eIF4E4(v) is a longer construct, composed by the whole eIF4E4(iv), and extends until the eIF4E4 C-terminal. eIF4E4(v) was designed to have this architecture in order to contain both the PAM2 motif, responsible for PABP1 binding, as described in this work, but also eIF4G3 domains, described to be located within the eIF4E4 signature domain (FREIRE, 2017).

It was not possible to obtain eIF4E4(v) as an isolated protein, but its production as a co-expressed complex with PABP1(J) was successful (complex prepared by Dr Maja Firczuk).

This investigation involved the MST analysis of eIF4G3 with eIF4E4(iv), PABP1, PABP1(J) and the complex {eIF4E4(v)/PABP1(J)}.

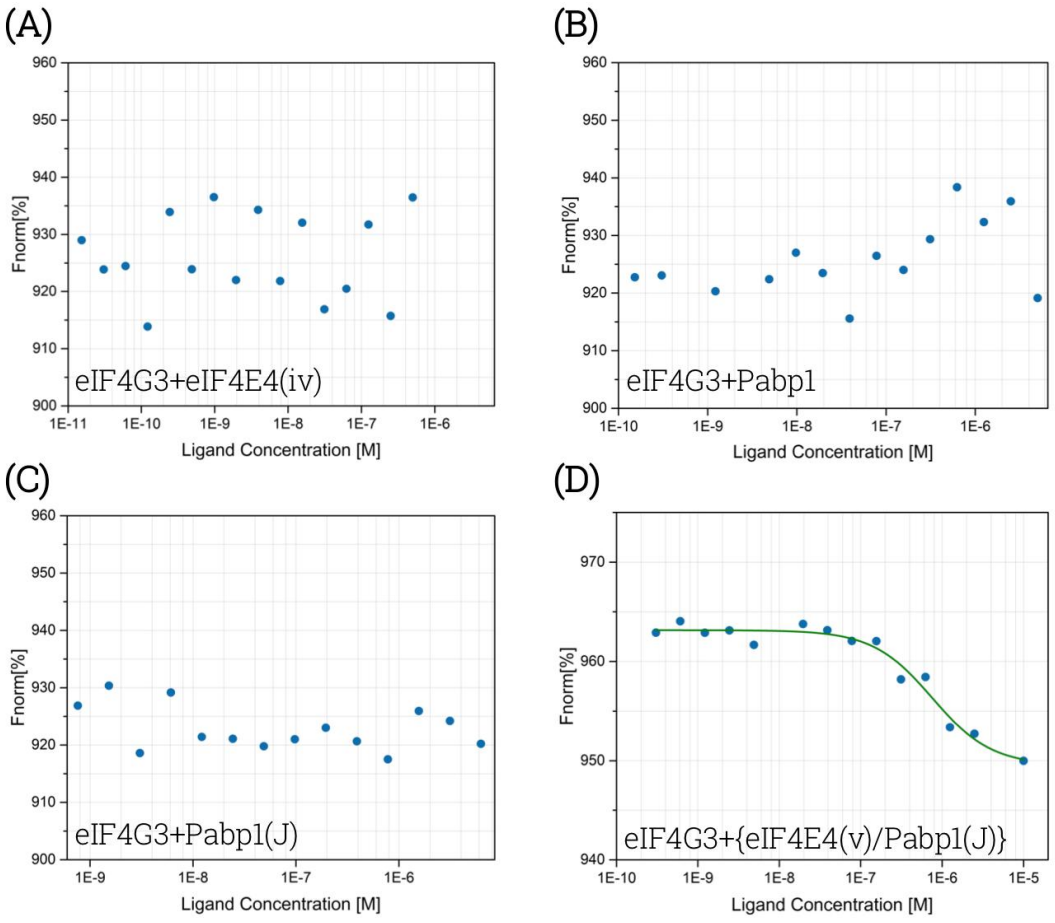


Figure 49: MST data of eIF4G3 and ***(A)*** *eIF4E4(iv);* ***(B)*** *PABP1 FL;* ***(C)*** *PABP1(J) and* ***(D)*** *{eIF4E4(v)/PABP1(J)}.*

The titration of eIF4G3 with eIF4E4(iv), PABP1 and PABP1(J) did not reveal evidence of binding. The absence of binding for eIF4E4(iv) was expected, since this construct does not contain the region responsible for eIF4G4 binding, located in the eIF4E4 signature domain (FREIRE, 2017).
The titration of eIF4G3 with PABP1 and PABP1(J) was not able to detect binding between these proteins. These data also suggest that there might be no binding between eIF4G3 and PABP1 in trypanosomatids, or that the interaction is very weak.

The titration of eIF4G3 with eIF4E4(v)/PABP1(J) complex, however, detected binding between the proteins (Figure 49).

It is important to highlight that although it was possible to detect binding between eIF4G3 and eIF4E4(v)/PABP1(J), this three-components complex was not very stable, with the rate of aggregation increasing with time. This might indicate that additional factors may be necessary to stabilize the whole complex, and therefore, further studies will be necessary.

The set of results obtained for eIF4G3 and eIF4E4(iv), PABP1(J), PABP1 and eIF4E4(v)/PABP1(J) complex suggest that, in the trypanosomatid translation initiation complex, the interaction between eIF4G3 and PABP1 is very weak or absent. This behaviour suggests the hypothesis that the extended eIF4E4 N-terminal might be responsible for the "scaffolding" of the eIF4G3-eIF4E4-PABP1 complex, in such way that it replaces the eIF4G3/PABP1 interaction.

Chapter 6 – NMR and Crystallographic analysis of the PABP1 PAM2 domain

6.1: Nuclear Magnetic Resonance - NMR

The NMR experiments provided information regarding the binding interactions between *Leishmania* eIF4E4(iv) and PABP1(J).

The initial NMR experiments were used to obtain qualitative information regarding protein folding, through the observation of spectral characteristics in the ^{1}H spectra.

Figure 50 shows the proton spectra of the PABP1(J). There are two major areas of interest, the first being in the 9-6 ppm region (left-hand side of the spectrum, in the dashed box), and the second located in the 1-0 ppm region (right-hand side of the spectrum, in the dashed box).

Folded proteins tend to have sharp, well defined and dispersed peaks (PAGE, 2005) in the 9-6 ppm region (amide and aromatic region) and intense, sharp peaks in the 1-0 ppm region (methyl group region (PAGE, 2005)). Observing PABP1(J) domain spectrum, it is possible to clearly see these characteristics, giving a good qualitative evidence that the protein is well folded.

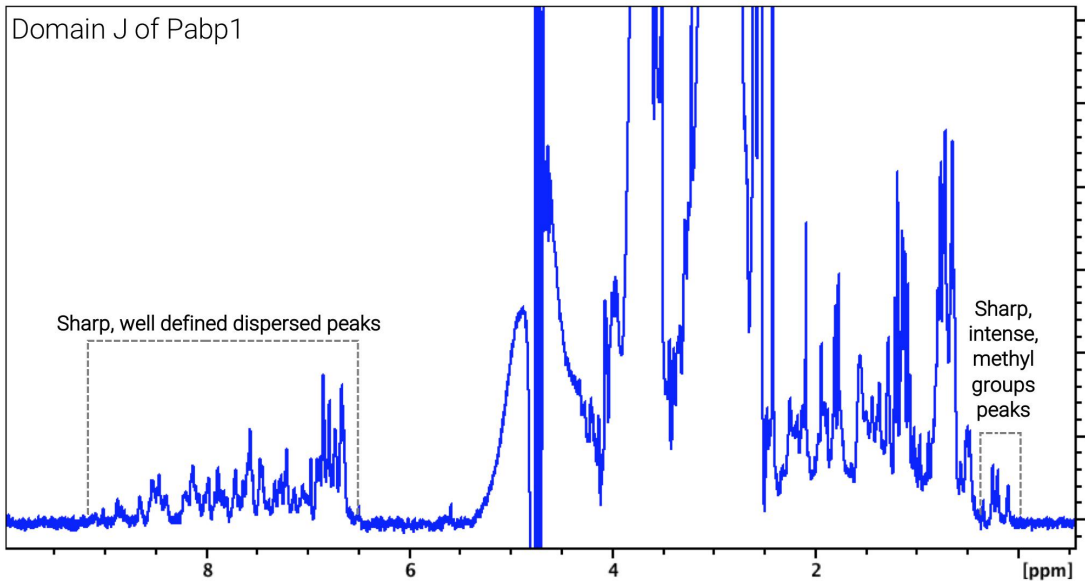


Figure 50: PABP1(J) proton spectrum, with characteristic aspects of a folded protein.

In contrast, Figure 51 shows the proton spectrum for eIF4E4(iv), where the opposite situation is observed, when compared with the PABP1(J) domain. The 9-6 ppm region is poorly spread, with non-sharp peaks, and the typical methyl peaks in the 0-1 ppm region can barely be seen. These conditions suggest that eIF4E4(iv) is significantly unstructured, result that agree to the expected lack of significant structure for eIF4E4(iv), since this protein is rich in prolines (that commonly contribute to destabilization protein structures).

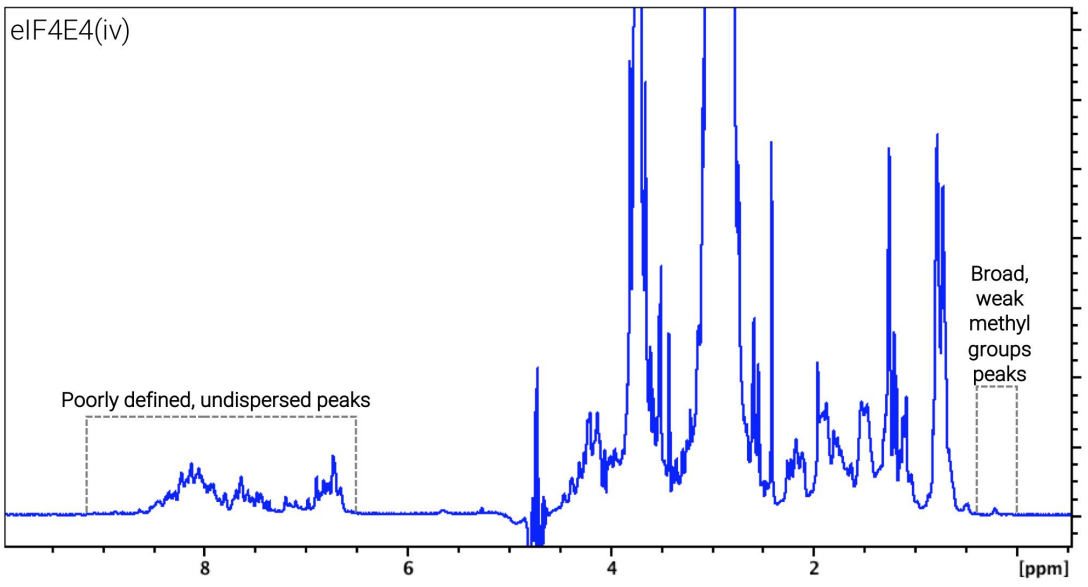


Figure 51: eIF4E4(iv) proton spectrum, with characteristic aspects of an unfolded protein.

NMR techniques were used to verify binding between *Leishmania* eIF4E4(iv) and PABP1(J), and if binding is detected, identify the regions/motifs involved in binding. We used bidimensional experiments, {^{1}H-^{15}N}-HSQC, where each peak corresponds to an amino acid residue.

We have done titrations with eIF4E4(iv) and PABP1(J), initially titrating isotopically labelled eIF4E4(iv) with unlabelled PABP1(J), and then, used isotopically labelled PABP1(J), that was then titrated with unlabelled eIF4E4(iv).

The PABP1(J) spectrum was obtained in good resolution, with peaks well resolved and dispersed along both dimensions. The peaks spread is an indication of protein folding (in a similar way as described above for the proton spectra).

Amino acid residues assignment was performed using triple resonance experiments, as described in the Material and Methods. The peaks were identified and assigned, and this information was used in the identification of the binding motif, after the titration experiments.
The HSQC NMR spectrum of the titration of isotopically labelled PABP1(J) with unlabelled eIF4E4(iv) is shown below:

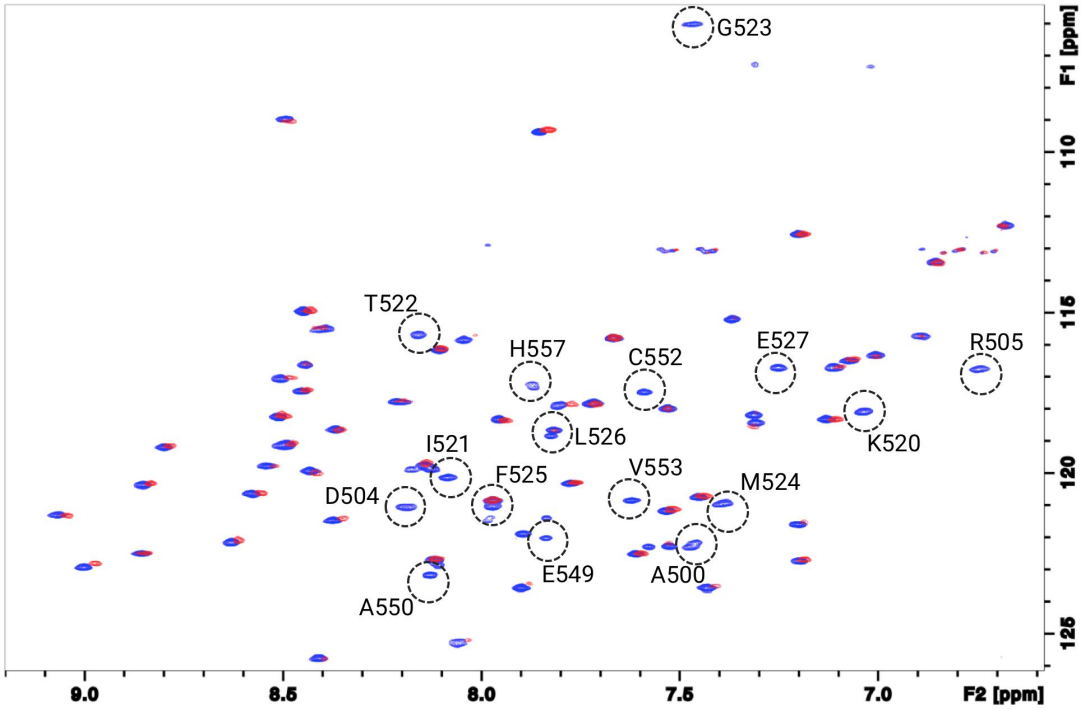


Figure 52: HSQC spectra of PABP1(J) Domain before (in blue) and after (in red) the titration with eIF4E4(iv). The residues that had alterations in their peaks, indicating binding, are highlighted by dashed circles.

After eIF4E4(iv) was added to PABP1(J), a number of peaks had their intensities dramatically reduced (this is observed when the interactions are in a range between slow (tight binding) and fast exchange (weaker binding) (WILLIAMSON, 2013)).

This behaviour is a strong indication that binding occurs between the two proteins. The residues identified as involved in binding are highlighted in PABP1(J) sequence below:

```
482
LPPITPQELESMSPQEQRAALGDRLFLKVYEIAPELAPKITGMFLEMKPKEAYELL
                        560
NDQKRLEERVTEALCVLKAHQTA
```

The mapped residues that the experiments suggest being involved in binding comprise the sequence with high homology to the MLLE domain, through the sequence KITGMFLE. This motif, along with its partner will be discussed in more detail in the following chapter.

The next step was to repeat the experiment, in reversed order, using isotopically labelled eIF4E4(iv), and then titrating with unlabelled PABP1(J). The eIF4E4(iv) spectrum was also obtained in good resolution, with peak well resolved, but displaying smaller dispersion along ^{1}H and ^{15}N dimensions. The concentration of peaks is a characteristic of unfolded protein, agreeing to what was described for the correspondent proton spectra. The HSQC NMR spectrum of the titration of isotopically labelled eIF4E4(iv) with unlabelled PABP1(J) is shown below:

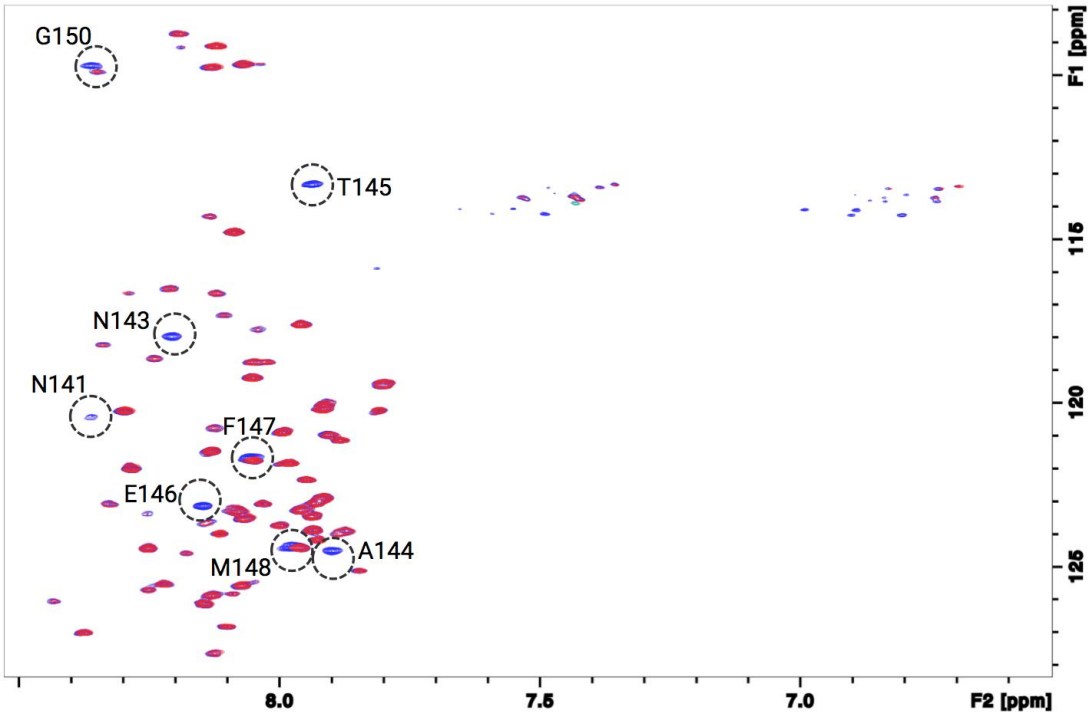


Figure 53: HSQC spectra of eIF4E4(iv) before (in blue) and after (in red) the titration with PABP1(iv). The residues that had alterations in their peaks, indicating binding, are highlighted by dashed circles.

After eIF4E4(iv) was titrated with PABP1(J), peaks also had their intensities reduced, disappearing in a similar way as observed in the PABP1(J) experiment. Again, the behaviour is an indication of binding is between the two proteins. The residues identified in the experiment are highlighted in eIF4E4(iv) sequence below:

```
140
MNPNATEFMPGRRNGPDGGLEALPTSTADMELAKTPAGAAAAAVHAPSLPGAVRRS
                                225
LQNSPIIQPSRLSVKSASEIEAISKNSALNA
```

The mapped residues suggested to be involved in binding are correspondent to PAM2 motifs, that are known to bind MLLE domain, through the sequence (L/P/F)x(P/V)xAxx(F/W)xP. This motif, along with MLLE domain will be discussed in more detail in the following chapter.
In order to confirm the binding motif, small peptides were synthesized, corresponding to the region that is suspected to be involved in binding.

The peptides were synthesized by our collaborators in the University of Leeds, in the group of Prof. Andrew Wilson. In this new set of experiments, we performed titrations in a way similar to the one used with the protein pairs, but now isotopically labelled PABP1(J) was titrated with unlabelled eIF4E4(iv) peptide, also called PAM2 peptide (Ac-HHMNPNATEFMPGR-NH_2).

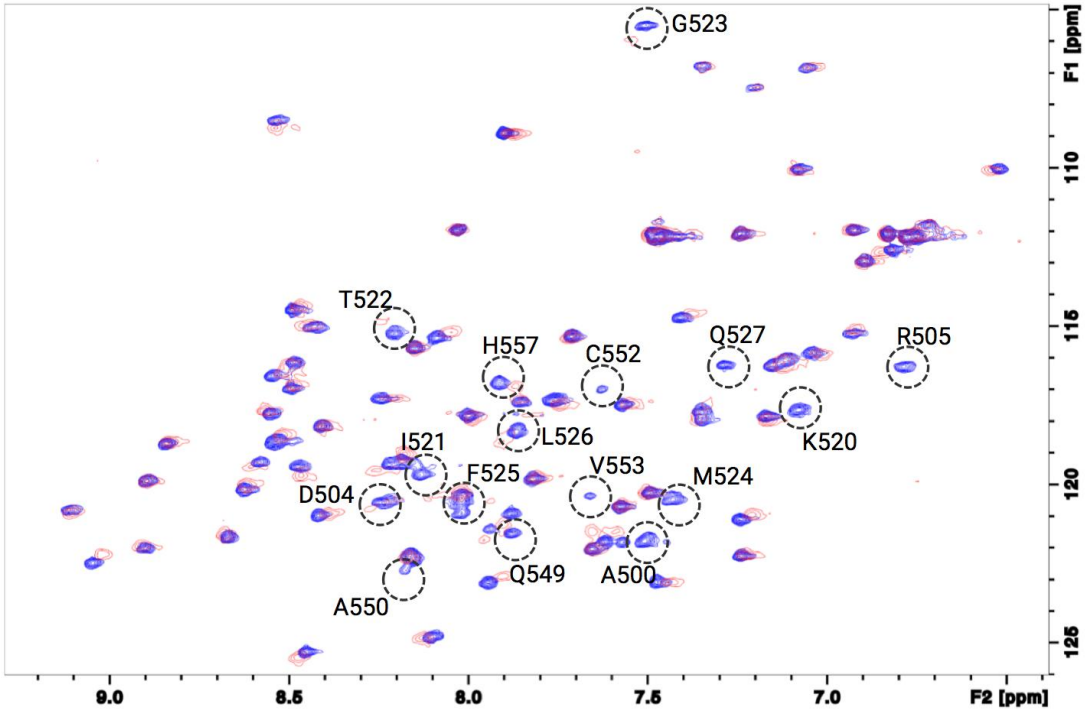


Figure 54: HSQC spectra of PABP1(J) Domain before (in blue) and after (in red) the titration with eIF4E4(iv) peptide. The residues that had alterations in their peaks, indicating binding, are highlighted by dashed circles.

After the peptide was added to PABP1(J), it was observed that the peaks that have disappeared were the same peaks that had disappeared in the experiment where PABP1(J) was titrated with eIF4E4(iv). This result gave greater confidence in the indication that binding occurs between the two proteins. The peptide experiment not only confirmed the existence of binding between eIF4E4(iv) and PABP1(J), but also strengthened the hypothesis that the PAM2 motif in eIF4E4(iv) is involved in binding. The realization of an experiment involving labelled eIF4E4(iv) and a PABP1(J) peptide would not be feasible, since the peptide that would contain the whole binding motif would be very large and challenging to be synthesized (the peptide would have almost the same length of that PABP1(J)).

6.2: Crystallographic data

Crystallographic experiments were also attempted in order to obtain more information regarding eIF4E4(iv)/PABP1(J) binding. These experiments were performed by Dr Maja Firczuk, from McCarthy's group, in collaboration with Prof. Alex Cameron group, also from the School of Life Sciences of the University of Warwick.

The crystallization experiments were successful in obtaining high resolution crystals of PABP1(J), as shown in Figure 55. In this section, I will analyze the crystallographic data, and then combine it with NMR results.

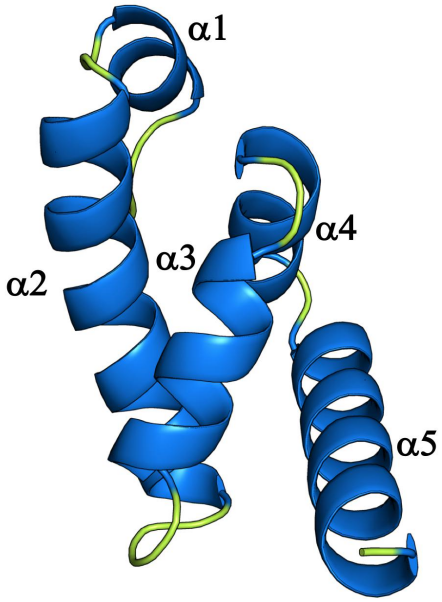


Figure 55: PABP1(J) crystal structure, with 5 α-helices assigned.

Figure 56: PABP1(J) α-helices location within its primary sequence.

The PABP1(J) crystal structure reveals the typical structure for PABC domains that contain MLLE region (XIE, 2014), composed by five

consecutive α-helices. PABP1(J) dimensions are x=26.5Å, y=28.5Å and z=46.0Å, and a summary of some of PABP1(J) structural properties is presented in Table 6:

Table 6: PABP1(J) structural properties obtained by crystallography.

Helix	length/Å
α1	8.9
α2	26.0
α3	14.6
α4	12.8
α5	24.5

	Angle between consecutive helices/°			
	α2	**α3**	**α4**	**α5**
α1	138.98			
α2		139.59		
α3			132.26	
α4				117.45

In order to obtain additional details for binding motif mapping, the NMR data was combined with PABP1(J) crystal structure, highlighting the residues (in blue) that had their signals abolished in the NMR data (Figure 57).

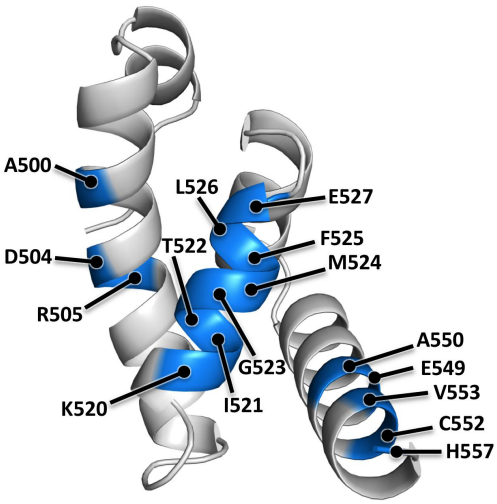


Figure 57: PABP1(J) crystal structure, with binding residues identified by NMR highlighted in blue.

The binding residues identified by the NMR experiments are located in the α-helices α2, α3 (majority) and α5. The comparison of the residues

identified shows a remarkable similarity with data from the literature (XIE, 2014), as displayed in Figure 58:

(A) (B)

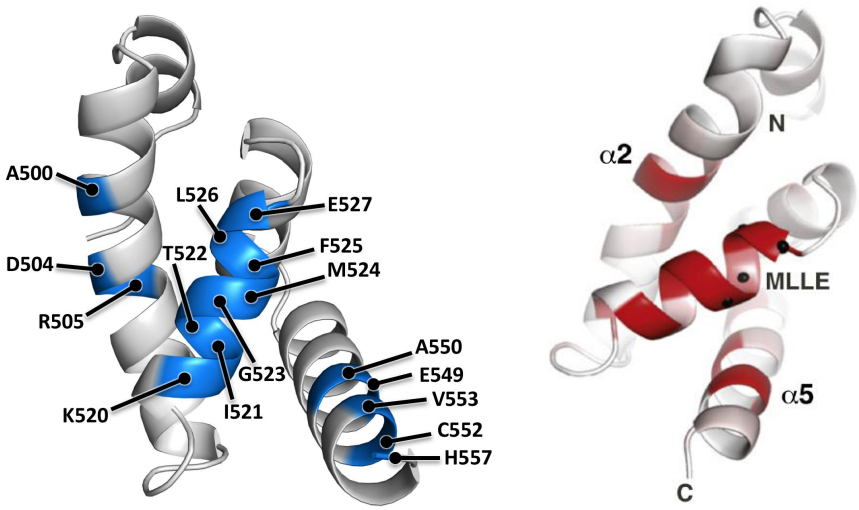


Figure 58: (A) PABP1(J) crystal structure with binding motif highlighted in blue; (B) human PABC1 MLLE domain, with binding residues in red, as described in the literature (XIE, 2014).

Additionally, a crystal structure of a complex of PABP1(J) and the peptide that contains the eIF4E4(iv) PAM2 binding motif was obtained (Figure 59). The amino acid residues from the peptide are positioned in such way that they interact in the MLLE domain in PABP1(J), agreeing with the NMR data.

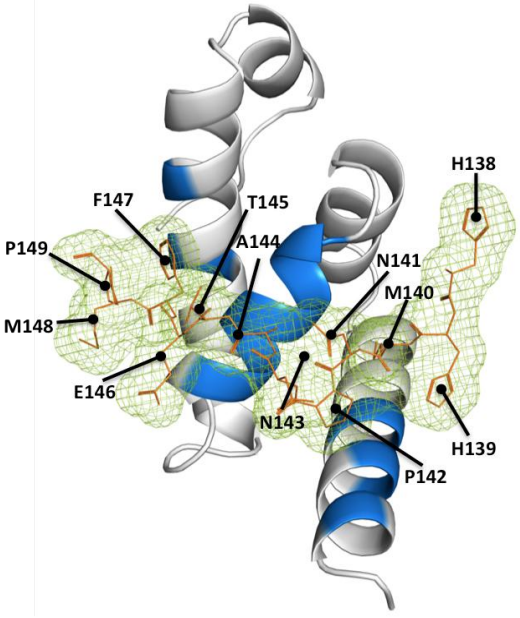


Figure 59: PABP1(J) crystal structure, with binding residues identified by NMR highlighted in blue and PAM2 peptide in orange sticks/green mesh.

In order to better display the interaction partners (*i.e.* amino acid residues that interact with each other), a map was prepared, as shown in Figure 60.

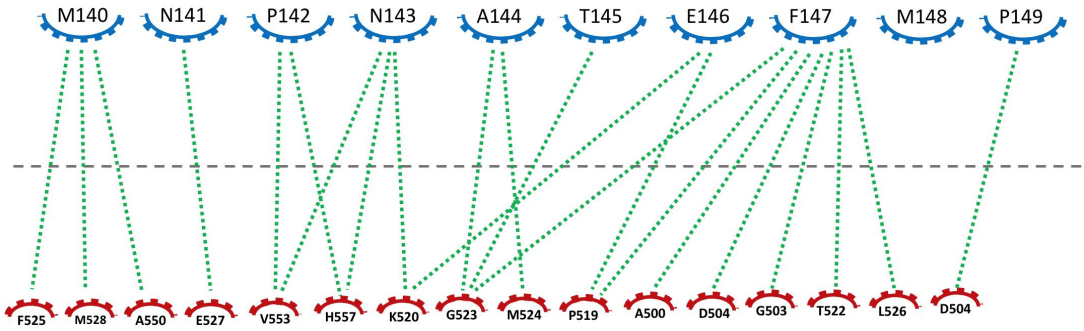


Figure 60: PABP1(J) and eIF4E4(iv) peptide interaction map, with eiF4E4(iv) peptide residues represented on the top (blue); PABP1(J) residues on the bottom (red). Interactions are indicated by the green dashed lines.

In order to obtain a comprehensive biophysical evaluation of these bindings, electrostatic and hydrophobicity maps were prepared.

(A) (B)

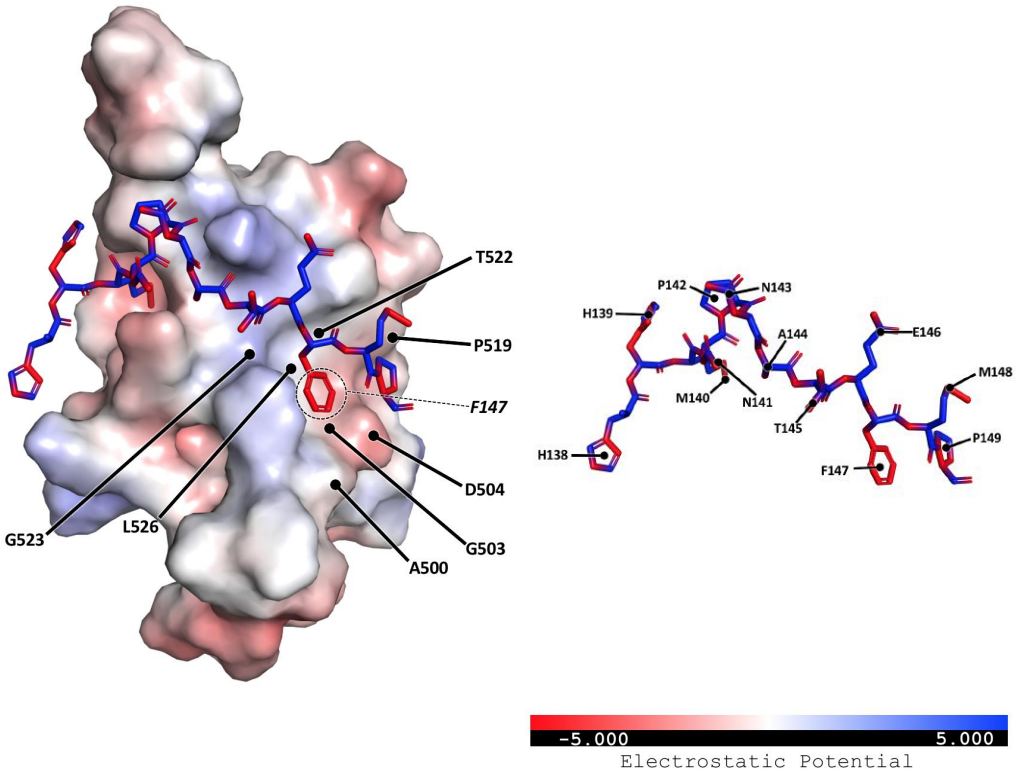


Figure 61: Electrostatic mapping of (A) PABP1(J) and (B) eIF4E4(iv) peptide.

The electrostatic map (Figure 61) does not suggest that the binding between PABP1(J) and eIF4E4(iv) takes place via electrical attraction, since there is no clear indication that this type of interaction is occurring in this protein/ligand complex. However, the evaluation of the hydrophobicity maps offers an alternative explanation regarding the nature of binding, as displayed in Figure 62.

(A) (B)

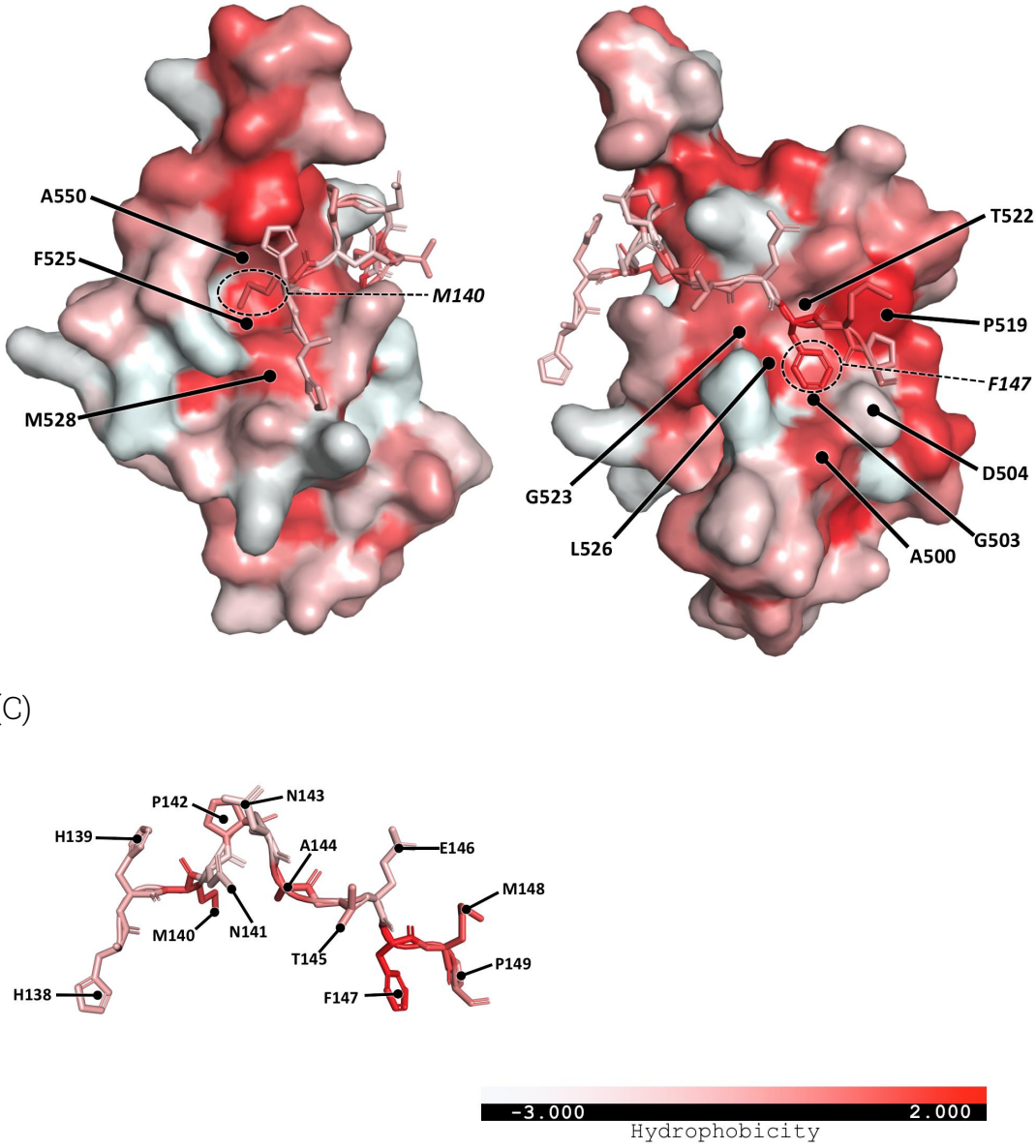


Figure 62: Hydrophobicity mapping of (A) PABP1(J) M140 pocket, (B) PABP1(J) F147 pocket, and (C) eIF4E4(iv) peptide.

The hydrophobicity map reveals multiple points of interaction based on hydrophobicity between PABP1(J) and eIF4E4(iv). It is important to highlight the eIF4E4(iv) residues M140 and F147, both very hydrophobic. These two residues are located within two hydrophobic pockets, known to be involved in PABC binding to several ligands (XIE, 2014; KOZLOV, 2010).

More detailed information can be found in the Annexes Chapter.

6.3: PAM2 motifs

We also investigated the presence of PAM2 motifs and MLLE domains in the other eIF4E and PABP1 isoforms of *L. major*.

Among the eIF4E isoforms, only eIF4E3 would contain an amino acid sequence that reasonably matches a PAM2 motif, but with significant difference from the PAM2 sequence observed in eIF4E4. The remaining eIF4E isoforms are significantly shorter than eIF4E4, not containing an equivalent N-terminal amino acid sequence.

```
                    121                                                168
L. major eIF4E1     --------------------------------------------------
L. major eIF4E2     --------------------------------------------------
L. major eIF4E3     AA----AVAKPPSTQPATKLSAAAEPFVPGGPKQMSATSTHVDPKATTE
L. major eIF4E4     VAARSVPTRFSPATVPRHHMNPNATEFMPGRRNGPDGGLEA-LPTSTAD
L. major eIF4E5     --------------------------------------------------
L. major eIF4E6     --------------------------------------------------
```

The three PABP1 isoforms have the MLLE domain in their primary structure. However, the (L→ F) substitution is only present in PABP1, while PABP2 and PABP3 display the standard MLLE sequence. It is also important to note that the KITG sequence that precedes the MLLE/MFLE is also conserved in the other PABP1 isoforms, while the remaining PABP1 residues that are involved in binding in with eIF4E4 are only marginally conserved in the other isoforms.

```
                        519                                      560
L. major PABP1          PKITGMFLEMKPKEAYELLNDQKRLEERVTEALCVLKAHQTA
L. major PABP2          AKITGMLLEMDNAEILNMLDSPTMLDSKIAEAQDVLNRHMSV
L. major PABP3          AKITGMLLEMSREEIFEILADHFALLSKIQEANAVLQQHTGN
```

The PAM2 motif in *Leishmania's* eIF4E4 is not present in human eIF4E. However, the PABP1 MLLE domain is present in both species, with *Leishmania* PABP1 having an amino acid substitution (L→F) when compared to the standard MLLE motif (MLLE → MFLE).

```
                        119                                      158
Leish major eIF4E4      PAVAARSVPTRFSPATVPRHHMNPNATEFMPGRRNGPDGG
Homo sapiens eIF4E4     ----------------------------------------

                        520                                      560
Leishmania major PABP1  KITGMFLEMKPKEAYELLNDQKRLEERVTEALCVLKAHQT
Homo sapiens PABP1      KITGMLLEIDNSELLHMLESPESLRSKVDEAVAVLQAHQA
```

In order to investigate if the PAM2/MLLE interaction might be present in other trypanosomatids, a number of species were chosen, and their eIF4E4 and PABP1 sequence aligned. The relevant section of these alignments is presented below.

eIF4E4

```
                        119                                  154
Leishmania major        PAVAARSVPTRFSPATVP----RHHMNPNATEFMPGRRNG
Leishmania panamensis   ------------------------MSPNATEFVPGRSNG
Leishmania guyanensis   PAVAARSVPTRFSPATVP----R-HMSPNATEFVPGRSNG
Leishmania braziliensis PAVAARSVPTRFSPATVP----R-HMNPNATEFVPGRSNG
Leishmania mexicana     ------------------------MNPNATEFMPGRRNG
Leishmania infantum     ------------------------MNPNATEFMPGRRNG
Trypanosoma brucei      -PVSTHVIPTRMSPVHAPS--AAFHMSPNAVSYVPRGA--
Trypanosoma theileri    -PTAALRFPTRMSPMHAPFPPVSSTMNPDAKDFIPHLS--
Trypanosoma cruzi Brener -PTTALRLPTRMSPMHAPFPSVSISMNPNATDFVPHLT--
```

PABP1

```
                          520                                       560
Leishmania major          KITGMFLEMKPKEAYELLNDQKRLEERVTEALCVLKAHQTA
Leishmania panamensis     KITGMFLEMNPKEAYELLNDQKRLEERVTEALCVLKAHQTV
Leishmania guyanensis     KITGMFLEMNPKEAYELLNDQKRLEERVTEALCVLKAHQTA
Leishmania braziliensis   KITGMFLEMNPKEAYELLNDQKRLEERVTEALCVLKAHQTV
Leishmania mexicana       KITGMFLEMKPKEAYELLNDQKRLEDRVTEALCVLKAHQTT
Leishmania infantum       KITGMFLEMKPKEAYELLNDQKRLEERVTEALCVLKAHQTA

Trypanosoma brucei        KITGMFLEMNPKEALALLSNPKLMHEKVTEALCVLKVHASS
Trypanosoma theileri      KITGMFLEMDLKEAFTLLYNQKLLHEKVTEALCVLKAHGTT
Trypanosoma cruzi Brener  KITGMFLEMDLKEAFTLLTNQRLLQEKVIEALCVLKAHEST
```

Both the PAM2 motif and MLLE domain are very conserved in trypanosomatids, suggesting an analogous assembly of translation initiation factors across the species, at least regarding the eIF4E4/PABP1 interaction. It is also notable that not only the MLLE domain is conserved, but also the (L→F) substitution in the various trypanosomatid species.

Chapter 7 – Discussion

Leishmaniasis is an important disease that affects a large number of people and, as explained in the Introduction, the identification of potential new therapeutic targets is therefore very important.

The trypanosomatid translation initiation complex displays characteristics that are not observed elsewhere, making it a very interesting object of study. The trypanosomatid translation initiation factor eIF4E4 has a unique N-terminal sequence whose interaction with PABP1 has not previously been described at the molecular level. The fact that this interaction is not observed in other eukaryotes makes it an interesting potential target for new therapeutic approaches.

The research described in this thesis demonstrates that synthetic peptides can form tight complexes with the trypanosomatid translation factors (eIF4E4 and eIF4G3). This demonstrates the potential value of peptides or peptidomimetics as drug discovery tools. By determining the molecular architecture of the complexes formed by these factors, we have a source of information that could, in future, inform the design of peptide-based inhibitors. One way of approaching this would be to screen peptide derivatives that mimic protein interaction sites in quantitative assays – for example, plate-reader assays utilizing, in each case, a fluorescein-derivatized peptide copy of one of the interaction sequences/domains in a binding pair, thus maximizing the change in anisotropy measured upon complex formation by the binding partners. It will be possible to test peptides of different lengths so as to establish a range of affinities for competition assays to be used in screening.

Using this proof-of-concept platform, it will then be possible to explore a range of modified synthetic peptides with additional properties (e.g. membrane-permeating and/or proteolysis-resistant) that are expected to be advantageous in the in vivo context. Peptidomimetics are useful drug discovery tools with high structural diversity that can be used to test target validity and, in cases like this, to explore the structure of relevant intermolecular interfaces. Bioinformatic analysis coupled to modelling of intermolecular docking and experimental structural interaction studies would also be used to refine the design of peptidomimetic derivatives of binding-domain amino acid sequences. One promising approach here would be the implementation of peptide stapling to lock peptides into conformations that have a high affinity for the target-binding domain. Stapling (SCHAFMEISTER,2000; VERDINE, 2012) is an effective helix-stabilizing method and can be used to build additional functionality into the peptide, including increased hydrophobicity, conferring membrane-permeability (ZHANG, 2008). The potential value of this would be that peptidomimetics might be developed that can enter into living cells.

This strategy has been used previously to design and synthesize eIF4E-binding peptidomimetics based on the binding-domain sequences of human 4E-BP1 and eIF4G (TAIT, 2010). Of course, in the present case, the objective would be to develop peptidomimetics that inhibit protein-protein interactions in Leishmania, but not in the human translation machinery.

Another strategy for developing potential anti-Leishmania drugs would be to screen for small-molecule inhibitors. At least one effective small molecule competitive inhibitor (called 4EGI-1) has previously been identified that targets the human eIF4E-eIF4G interaction (identified from the Chembridge DiverSet E Library (MOERKE, 2007)). Screening could again be based on fluorescence anisotropy assays as described

above. In parallel, the small molecule 'ligandability' of potential drug targets can be tested via NMR-based fragment screening (EDFELDT, 2011). Once small molecule inhibitory leads have been identified that specifically target Leishmania protein-protein interactions within promastigotes or amastigotes, molecular modelling work can be used to pursue a refine and re-test cycle in order to optimise the inhibitory properties of the selected ligands.

As with the peptidomimetics, the small molecule inhibitors will be tested in both in vitro and in vivo assays. As the number of structurally analysed protein-protein interactions increases, the breadth of potential inhibitor molecules will be expanded, thus increasing the likelihood of finding at least one inhibitor that might form the basis of proper drug development.

Trypanosomatids have multiple isoforms for some of its translation initiation factors (*e.g.* it has 6 eIF4E, 5 eIF4G and 3 PABP isoforms). The existence of such a high number of isoforms is likely to be related to Leishmania's adaptation to its complex life cycle, in which it is exposed to different host environments. For example, eIF4E1 is found at significantly higher levels in amastigotes, and seems to be involved in the amastigote phase of the parasite (YOFFE, 2006), when it lives inside the mammalian host, and faces higher temperatures. The expression levels of the other eIF4E isoforms, however, do not seem to be significantly affected by temperature variations. eIF4E5 and eIF4E6 (FREIRE, 2011), on the other hand, are involved in cell motility during the promastigote phase.

In this study, several methods were used in order to investigate the interactions between *Leishmania*'s eIF4E4 and PABP1 proteins. The overproduction of the full-length versions of these proteins was found to

be difficult. A number of strategies were explored in attempts to obtain these proteins, via the use of different tags (His-tag, GST, MBP, among others), and several cultivation conditions. The production and purification of PABP1 was successful, although at moderate yield, while little progress was achieved with the other full-length proteins.

Full-length eIF4E4 could not be obtained in soluble form on its own. However, it was found possible to generate it in *E. coli* as part of a complex with PABP1. The literature (ZINOVIEV, 2011) has previously provided qualitative evidence of an analogous interaction between these two proteins in *L. amazonensis* on the basis of pull-down data.
In order to map accurately the potential binding regions, PABP1 was also divided into several domains, whereby identification of the respective subdomains was guided by bioinformatics analysis of the predicted motifs (for example, RNA Recognition Motifs).

The initial investigations used Surface Plasmon Resonance as a means to scan domains of PABP1 for binding to eIF4E4(iv). For these experiments, *Leishmania* eIF4E4(iv) was immobilized on the SPR chip surface using amine coupling. The initial qualitative observations using Yes/No Binding assays provided an indication of which PABP1 domains would bind to eIF4E4(iv). The Yes/No Binding experiments suggested that the PABP1 C-terminal domains (C, G and J), as well as the full-length protein, bind to eIF4E4(iv). There was also a positive signal for one of the N-terminal domains (A), but the data obtained were not unequivocally clear. Under the conditions used in these initial scanning experiments, the sensorgrams lacked a satisfactory dissociation phase. This problem can arise from nonspecific binding and subsequent, more detailed, investigations led to significant improvements in this part of the binding

profile. The optimization experimental conditions (flow rate, buffers used, temperature) reduced the contribution of non-specific binding, although it did not improve significantly the dissociation phase.

MST experiments were used seeking to further investigate the binding between eIF4E4(iv) and the various PABP1 domains. The MST data provided more detailed information regarding protein binding, and the results agreed with what was observed in the SPR experiments, with the full-length PABP1 and its C-terminal domains displaying binding to eIF4E4(iv).

In order to further test for protein-protein interactions, deletion constructs were prepared (PABP1ΔA, PABP1ΔC and PABP1ΔJ), where the C-terminal domains were absent. It was predicted on the basis of the earlier experiments that deletion of the C-terminal region would abolish binding between the deletion constructs and eIF4E4(iv), and this hypothesis was confirmed experimentally.

The interaction data obtained in the MST experiments gave a strong indication that the PABP1 C-terminal is involved in binding to eIF4E4(iv). In order to investigate the structural basis of this binding, NMR and crystallographic experiments were performed. As mentioned in the results chapter, the crystallographic work was performed by Dr Maja Firczuk in collaboration with Prof Alex Cameron's group.

Initially, the first NMR experiments attempted involved the acquisition of 2D {^{1}H-^{15}N}-HSQC spectra. Titration of an isotopically labelled target with an unlabelled ligand provided a means to determine if binding was taking place.

The experiments involved first the titration of isotopically labelled PABP1(J) with unlabelled eIF4E4(iv). These experiments showed very clearly that a number of peaks disappeared (shifted) in response to the titration, indicating that binding was happening between the two proteins. Moreover, identification of the chemical shift perturbations (CFPs) indicated which amino acid residues were involved in binding. Similarly, isotopically labelled eIF4E4(iv) was titrated with unlabelled PABP1, and the occurrence of CSPs again confirmed that specific interactions had taken place.

In this form of experiment, two extreme types of behaviour can be observed. Where binding is weak and there is fast exchange, a single peak appears that is a weighted average of the two states, and this progressively moves position over the range of the titration (Figure 63). Where there is strong binding, there is slow exchange between bound and free states, giving rise to changes in the relative intensity of two distinct positions as the ligand concentration is titrated (Figure 64). Where these relatively simple types of behaviour are observed, it is possible to obtain information on the system affinity. However, in the case of the eIF4E4(iv)/PABP1(J) pair, an intermediate type of behaviour is observed, in which the peaks disappear, and reappear close to ligand saturation, making it difficult to calculate affinity on the basis of the titration data.

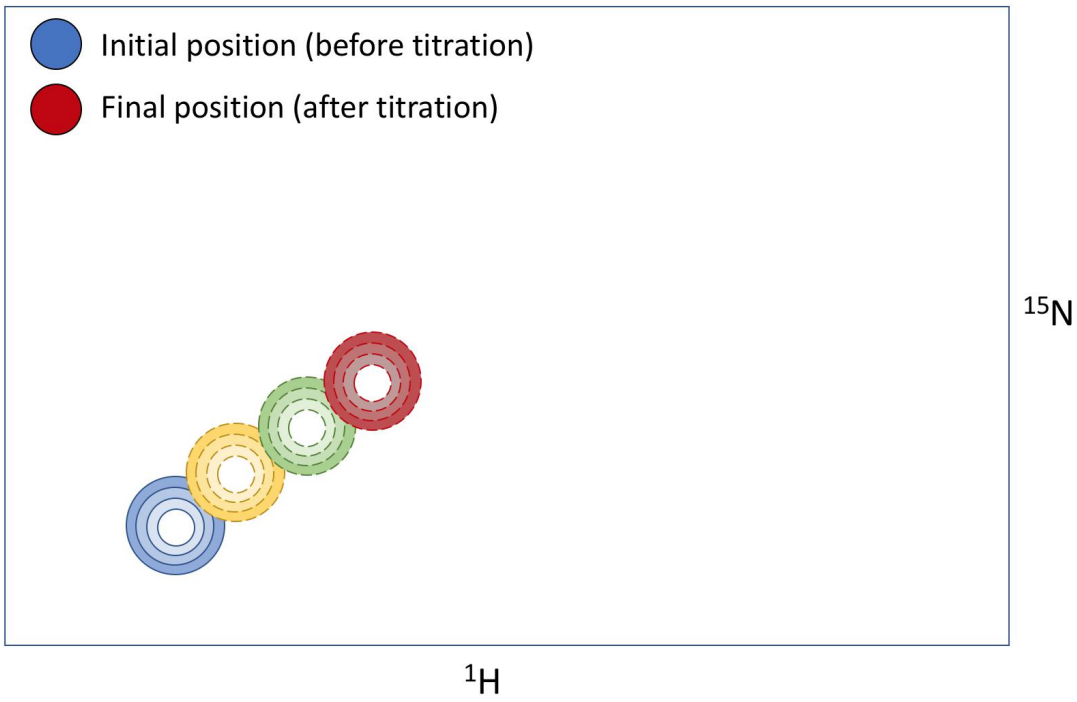


Figure 63: Illustration of chemical shift perturbation in a fast-exchange system, with peak slowly changing its position.

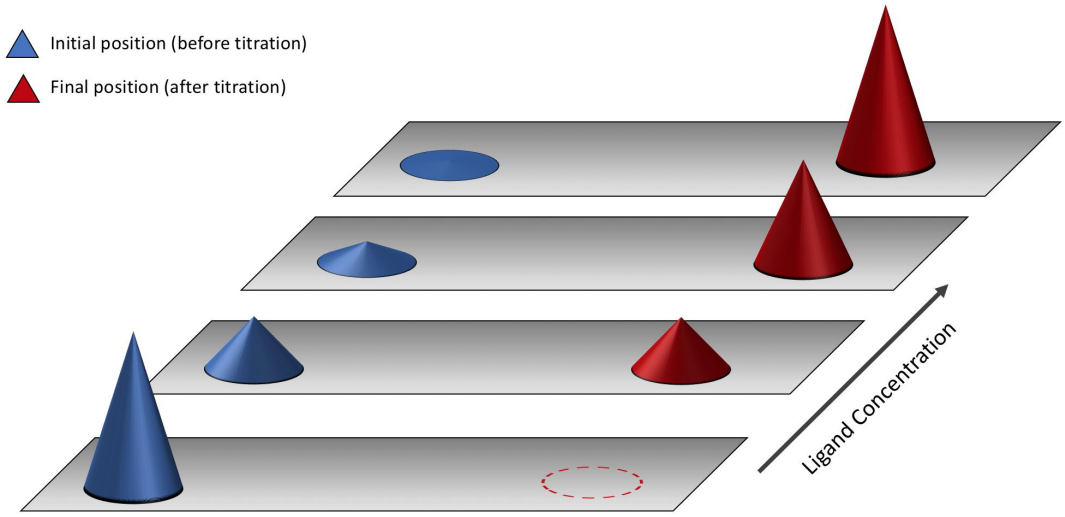


Figure 64: Illustration of chemical shift perturbation in a slow-exchange system, with peak intensity decreasing in the initial position; and increasing in a new, final position.

Even though no accurate quantitative data could be obtained from the HSQC data, it was possible to determine the identity of each peak via triple resonance experiments, thus leading to the characterization of

amino acid residues that are either directly involved in, or whose environment is affected by, binding interactions.

The spectra assignment revealed that the peaks that disappeared in the titration experiments corresponded to the MLLE domain and PAM2 motif, for PABP1 and eIF4E4(iv), respectively.

By combining the NMR data with crystallographic data, it was possible to observe that both agreed in assignment of the binding motifs involved in eIF4E4(iv)/PABP1 binding. The MLLE/PAM2 type of interaction has been described previously as mediating interactions between other pairs of proteins. Indeed, PABC domains in other proteins have been described previously (KOZLOV 2010; XIE, 2014) as MLLE-containing domains that interact with several partners. However, the interaction between eIF4E4 and PABP1 has not previously been described in detail. The presence of PAM2 motif in the unique N-terminal sequence of eIF4E4 highlights the importance of this segment in the interaction with PABP1 and suggests an explanation for its existence. The present study has also revealed that only one of the PAM2 or PAM2-like sequences in eIF4E4 is involved in PABP1 binding.

Table 7: eIF4E4 aminoacid residues and corresponding PABP1 aminoacid residues involved in the PAM2/MLLE interaction.

eIF4E4 a.a. residue	Interacting PABP1 a.a. residue
M140	F525, M528, A550
N141	E527
P142	V553, H557
N143	V553, H557, K520
A144	G523, M524
T145	G523
E146	K520, P519
F147	G523, P519, A500, D504, G503, T522, L526,
M148	-
P149	D504

It was also possible to obtain information about the molecular principles underpinning binding between eIF4E4 and PABP1. Analysis of the NMR and crystallographic data, using electrostatic and hydrophobicity maps, indicates that the main driving force for complex formation is through the association of hydrophobic regions of multiple amino acid side chains, with two hydrophobic pockets in PABP1 being particularly important.

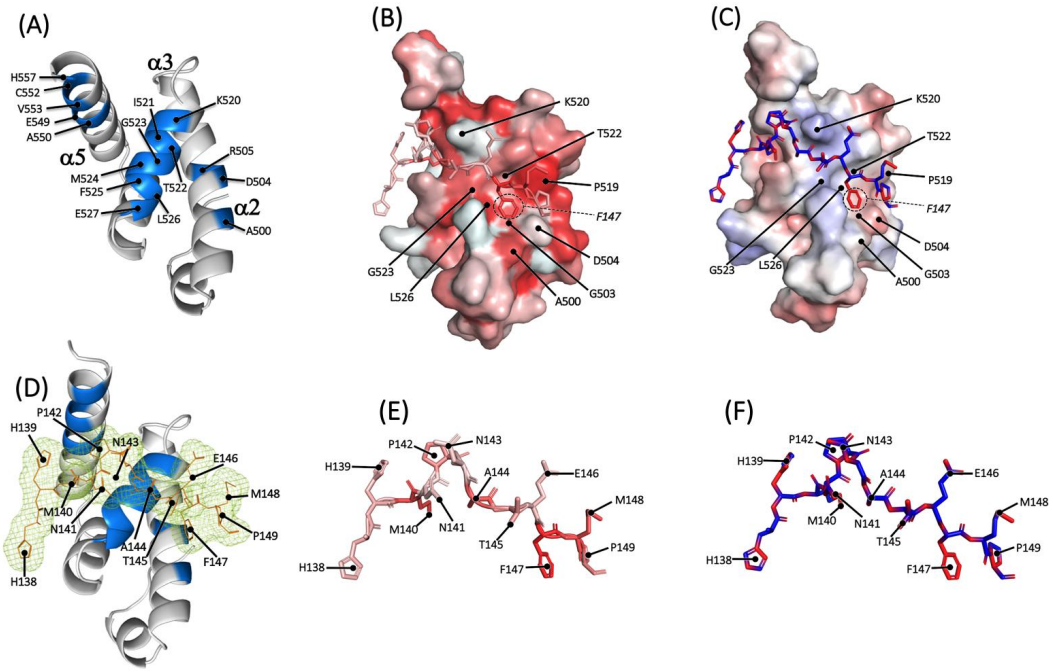


Figure 65: (A) PABP1(J) crystal structure with residues involved in binding to eIF4E4(iv) highlighted in blue; (B) PABP1(J)/eIF4E4(iv) peptide hydrophobicity map; (C) PABP1(J)/eIF4E4(iv) peptide electrostatic map (D) PABP1(J)/eIF4E4(iv) crystal structure with eIF4E4(iv) residues involved in binding to PABP1 highlighted; (E) eIF4E4(iv) peptide hydrophobicity map and (F) eIF4E4(iv) peptide electrostatic map.

Our investigation also looked into the eIF4G3-eIF4E4-PABP1 interaction. MST data were obtained using eIF4G3, eIF4E4(iv), PABP1 FL, PABP1(J) and the {eIF4E4(v)/PABP1(J)} complex. The rationale for these choices lies in the respective protein architectures. eIF4E4(iv) contains only the unique N-terminal that we found to bind to PABP1, so it should not bind to eIF4G3, serving as a negative control. Similarly, PABP1(J) contains only the PABC domain that binds to eIF4E4(iv), and which is not expected to bind to eIF4G3. eIF4E4(v) contains a region that is predicted on the basis of bioinformatics to contain potential eIF4G3-binding sites, and PABP1 FL retains all of the relevant natural binding sites. It was not found possible to express and purify eIF4E4(v) on its own, but its co-production with PABP1(J) was successful.

The MST experiments did not detect binding between eIF4G3 and eIF4E4(iv), or between eIF4G3 and PABP1 FL. The experiments

investigating binding between eIF4G3 and PABP1(J) also did not detect any interaction. However, it was possible to observe binding between eIF4G3 and {eIF4E4(v)/PABP1(J)}. Considering the fact that no binding could be detected between eIF4G3 and PABP1(J), eIF4E4(v) must be responsible for the interaction between eIF4G3 and eIF4E4(v)/PABP1(J).

These data indicate that in Trypanosomatids, the translation initiation complex displays an alternative architecture, with eIF4E4 binding to PABP1 and eIF4G3, but with no (or very weak) interaction between PABP1 and eIF4G3 (Figure 66).

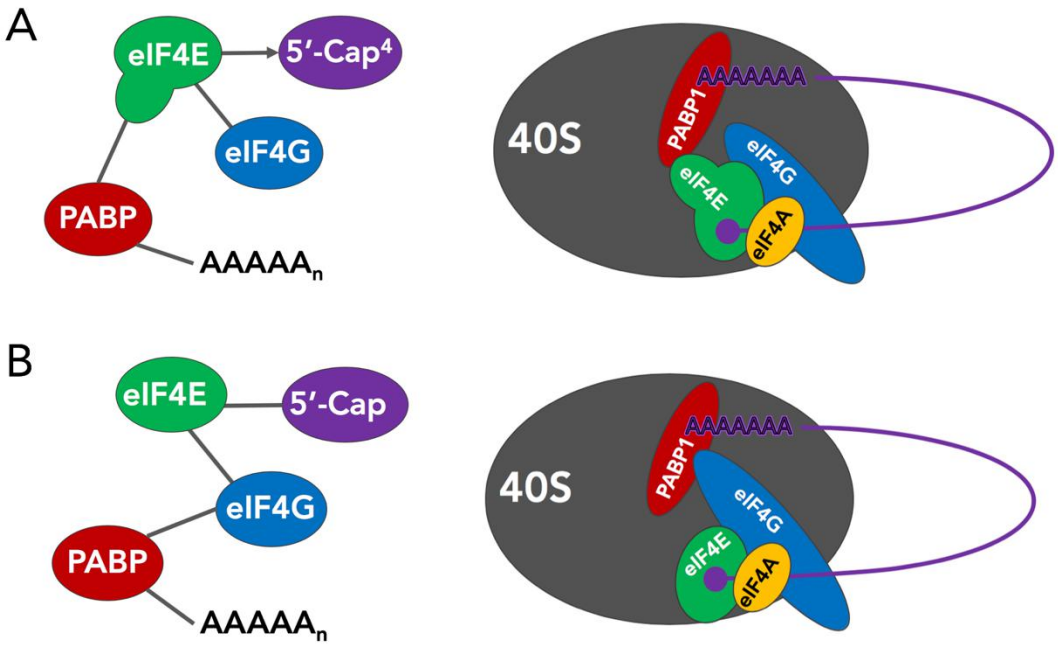


Figure 66: In trypanosomatids, eIF4E4 is the "scaffolding" component in the cap-binding complex. (A) The data obtained in this work indicate that Leishmania eIF4E4 interacts in specific manner with PABP1 and eIF4G3, while we were not able to detect interaction between PABP1 and eIF4G3. (B) In other known eukaryotic cap-binding complexes, eIF4G is the main scaffolding protein.

Finally, this study has focused on key interactions and structures underpinning the unique architecture of the *Leishmania* eIF4E4-PABP1-eIF4G3 biomolecular chain that links the 5′ and 3′ ends of trypanosomatid mRNAs. Future work will be required to increase our understanding of

the range of interactions involving other translation factors in trypanosomatid translation initiation, such as the DEAD-box RNA-binding protein eIF4A.

Chapter 8 – Conclusions

This study set out to primarily to investigate and analyze the interaction between two translation initiation factors from the trypanosomatid *Leishmania major*, *i.e.* eIF4E4 and PABP1.

The initial strategy involved the investigation of the unique N-terminal sequence from eIF4E4, and its interaction with different domains of PABP1. Within the eIF4E4 N-terminal, a short 86 amino acid sequence was described (ZINOVIEV, 2011) to be required for binding to PABP1, and this sequence, denominated in this work eIF4E4(iv), was used as a starting point.

PABP1 was divided into several shorter domains that correspond to different parts of its architecture, and these domains were used in interaction experiments with eIF4E4(iv).

The SPR experiments gave initial, although not completely unequivocal results regarding the binding regions in PABP1, however they pointed towards involvement of the C-terminal region of PABP1.

MST experiments provided additional information, displaying the relative affinities of the different PABP1 domains, most importantly pointing to the involvement of the PABP1 C-terminal in binding to eIF4E4(iv). Experiments with PABP1 constructs in which the C-terminal was deleted showed that binding to eIF4E4(iv) is abolished with the removal of PABP1 C-terminal.

NMR experiments were performed in order to further investigate binding and also identify the binding motif, using isotopically labelled eIF4E4(iv) and PABP1(J). 2D NMR data revealed that the proteins undergo a specific interaction, whereby triple resonance experiments mapped potential binding motifs.

Motif mapping revealed that the binding takes place through a MLLE/PAM2 interaction, with the PABP1 PABC domain containing the MLLE domain, and the eIF4E4 N-terminal the PAM2 motif.

The combination of NMR data with the data obtained through X-ray crystallography provided mutually consistent results, revealing that the MLLE domain corresponds to the typical structure (XIE, 2014; KOZLOV 2010) for a MLLE/PAM2 interaction. These results have therefore described in detail the characteristics of this unique Trypanosomatid eIF4E4 N-terminal region and its interactions with PABP1. The NMR data revealed that eIF4E4 N-terminal is naturally disordered, but its interaction with PABP1 induces at least a partial folding in this region.

Additionally, the interaction of eIF4G3 and eIF4E4 and PABP1 was investigated by MST. Our MST results revealed interaction between eIF4G3 and eIF4E4, but we were not able to detect interaction between eIF4G3 and PABP1 (indicating that this interaction might not exist or be very weak). These data suggest that in Trypanosomatids, the translation initiation complex might display an alternative architecture, with eIF4E4 binding to PABP1 and eIF4G3, but without interaction between PABP1 and eIF4G3 (see Figure 67).

Chapter 9 – Annexes

The following data is presented in this Chapter:

-Primers List

-Plasmid maps and gene blocks

-Proteins and domains sequences

-SPR sensorgrams and protein concentrations used

-Microscale thermophoresis data

-Nuclear Magnetic Resonance Data

-Crystallographic data

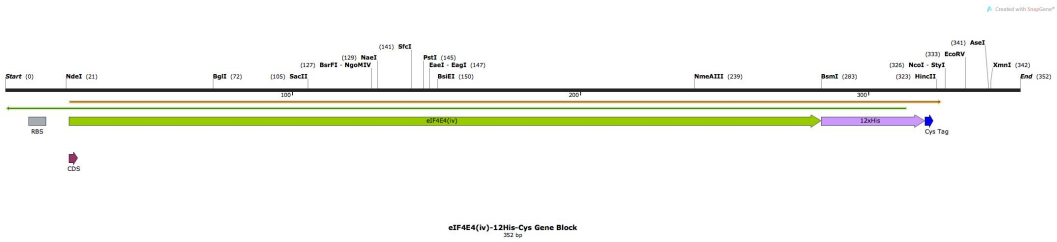
PRIMERS LIST

Name	Primer Sequence	Use
663_pET22b_NdeI_F	ATATACATCATATGAAAATAAAAACAGGTGCACGCATC	eIF4E4-12His in pEt22b vector Forward primer.
663_pET22b_HindIII_R	ATATACAAGCTTTGGGGTTAACCTTAATAGCGACGACGG	eIF4E4-12His in pEt22b vector Reverse primer.
883_pET22b_NdeI_F	ATATATCCATATGATGAAAATAAAAACAGGTGCACGC	PABP1-12His in pEt22b vector Forward primer.
883_pET22b_HindIII_R	ATATATCAAGCTTTGGGGTTAACCTTACGCAGTCTGG	PABP1-12His in pEt22b vector Reverse primer.
663_MBP_BamHI_F	ATATACGGATCCATGAGCACTCCGTTAGACGTTCGT	eIF4E4-6His in MBP fusion vector Forward primer.
663_MBP_HindIII_R	ATATACAAGCTTTGGGGTTAACCTTAATAGCGACGACGG	eIF4E4-6His in MBP fusion vector Reverse primer.
883_MBP_BamHI_F	ATATACGGATCCATGGCAGCAGCAGTCCAAGAAGCA	PABP1-6His in MBP fusion vector Forward primer.
883_MBP_HindIII_R	ATATATCAAGCTTTGGGGTTAACCTTACGCAGTCTGG	PABP1-6His in MBP fusion vector Reverse primer.

PLASMID MAPS AND GENE BLOCKS

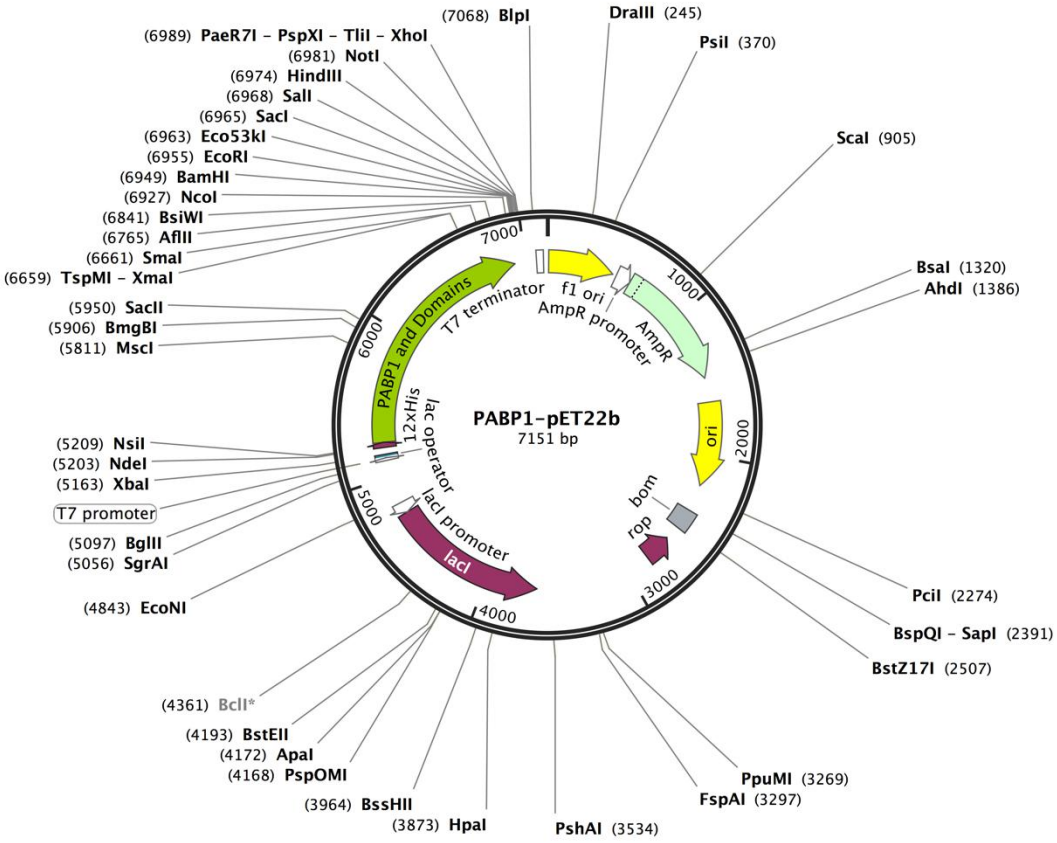
>eIF4E4(iv) Plasmid map – pET22b

>eIF4E4(iv) Gene Block

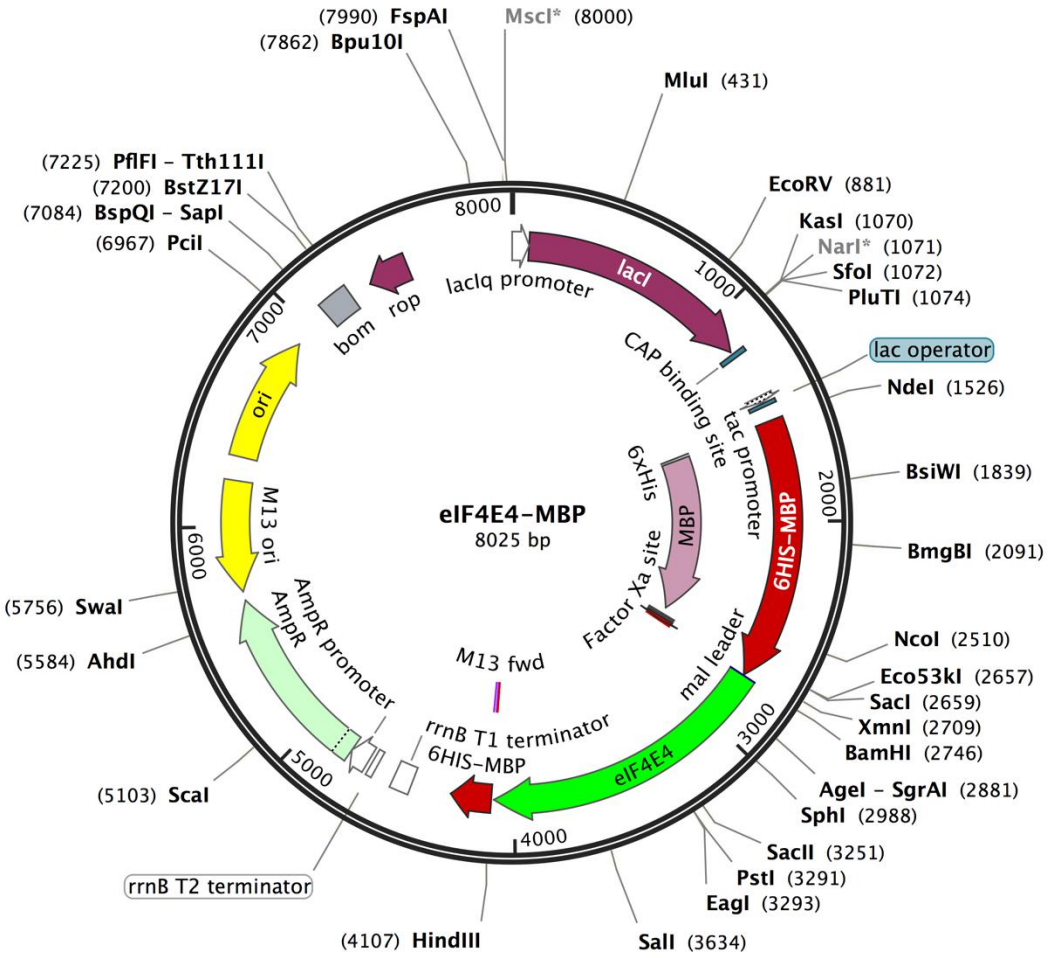


ACTTTAAGAAGGAGATATACATATGAATCCGAACGCAACCGAGTTCATGCCGGGTC
GTCGTAACGGCCCAGATGGCGGCCTGGAAGCCCTGCCAACCAGCACCGCGGACATG
GAATTGGCAAAAACGCCGGCTGGTGCGGCTGCAGCGGCCGTTCATGCGCCGTCCCT
GCCGGGTGCGGTGCGTCGTTCGCTGCAAAACAGCCCGATCATTCAGCCGAGCCGTC
TGAGCGTCAAAAGCGCCAGCGAGATCGAAGCGATTAGCAAAAACAGCGCCCTGAAT
GCACATCACCACCACCACCATCACCACCACCATCACCACTGTTGACCATGGATATC
GGAATTAATTCGGATC

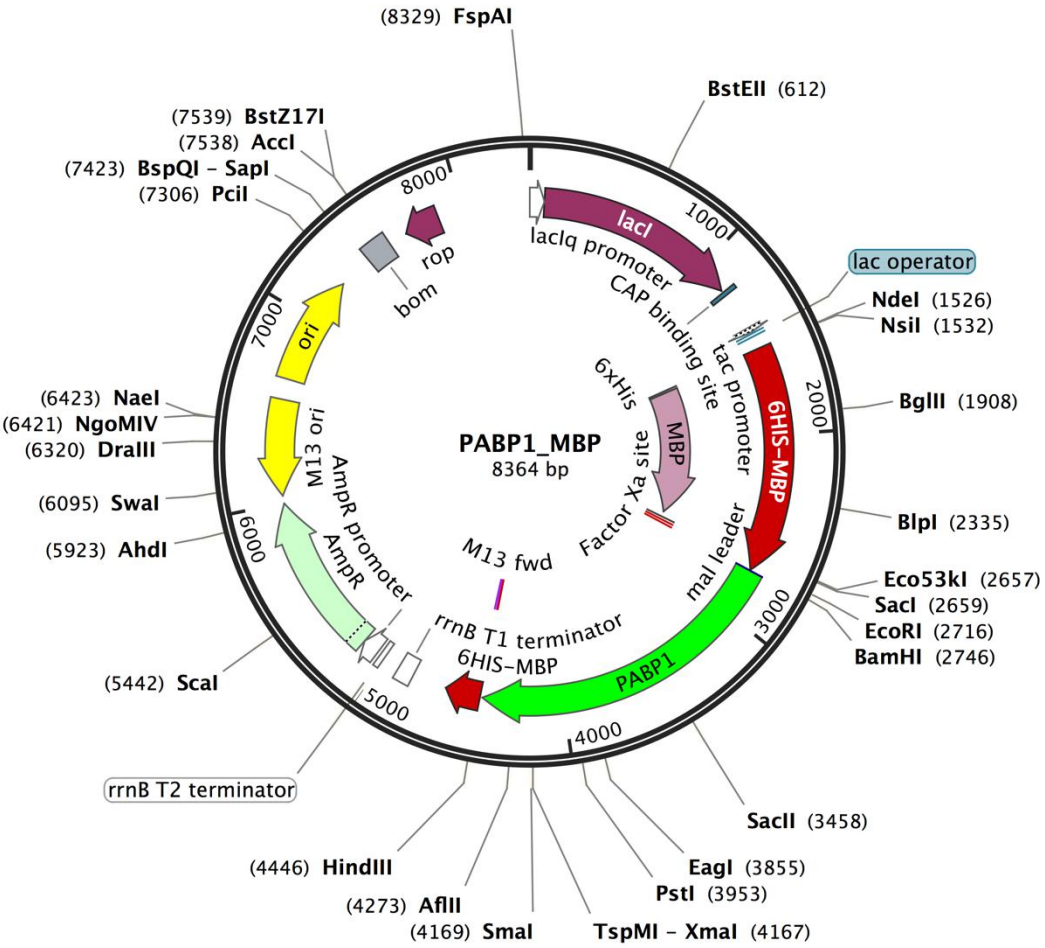
>PABP1 and Domains Plasmid map – pET22b



>eIF4E4-MBP Plasmid map



>PABP1-MBP Plasmid map



PROTEINS AND DOMAINS SEQUENCES

>Leish major PABP1 Full length

MAAAVQEAAAPVAHQPQMDKPMQIASIYVGDLDATINEPQLVELFKPFGTILNVRV
CRDIITQRSLGYGYVNFDNHDSAEKAIESMNFKRVGDKCVRLMWQQRDPALRYSGN
GNVFVKNLEKDVDSKSLHDIFTKFGSILSCKVMQDEEGKSRGYGFVHFKDETSAKD
AIVKMNGAADHASEDKKALYVANFIRRNARLAALVANFTNVYIKQVLPTVNKDVIE
KFFAKFGGITSAAACKDKSGRVFAFCNFEKHDDAVKAVEAMHDHHIDGITAPGEKL
YVQRAQPRSERLIALRQKYMQHQALGNNLYVRNFDPEFTGADLLELFKEYGEVKSC
RVMVSESGVSRGFGFVSFSNADEANAALREMNGRMLNGKPLIVNIAQRRDQRYTML
RLQFQQRLQMMMRQMHQPMPFVGSQGRPMRGRGGRQQLGGRAQGHPMPMPSPQQPQ
APAQPQGFATPSAVGFVQATPKHSPGDVPETPPLPPITPQELESMSPQEQRAALGD
RLFLKVYEIAPELAPKITGMFLEMKPKEAYELLNDQKRLEERVTEALCVLKAHQTA

>Leish major PABP1(A)

MAAAVQEAAAPVAHQPQMDKPMQIASIYVGDLDATINEPQLVELFKPFGTILNVRV
CRDIITQRSLGYGYVNFDNHDSAEKAIESMNFKRVGDKCVRLMWQQRDPALRYSGN
GNVFVKNLEKDVDSKSLHDIFTKFGSILSCKVMQDEEGKSRGYGFVHFKDETSAKD
AIVKMNGAADHASEDKKALYVANFIRRNARLA

>Leish major PABP1 (B)

ALVANFTNVYIKQVLPTVNKDVIEKFFAKFGGITSAAACKDKSGRVFAFCNFEKHD
DAVKAVEAMHDHHIDGITAPGEKLYVQRAQPRSERLIALRQKYMQHQALGNNLYVR
NFDPEFTGADLLELFKEYGEVKSCRVMVSESGVSRGFGFVSFSNADEANAALREMN
GRMLNGKPLIVNIAQR

>Leish major PABP1 (C)

RDQRYTMLRLQFQQRLQMMMRQMHQPMPFVGSQGRPMRGRGGRQQLGGRAQGHPMP
MPSPQQPQAPAQPQGFATPSAVGFVQATPKHSPGDVPETPPLPPITPQELESMSPQ
EQRAALGDRLFLKVYEIAPELAPKITGMFLEMKPKEAYELLNDQKRLEERVTEALC
VLKAHQTA

>Leish major PABP1(D)

MAAAVQEAAAPVAHQPQMDKPMQIASIYVGDLDATINEPQLVELFKPFGTILNVRV
CRDIITQRSLGYGYVNFDNHDSAEKAIESMNFKRVGDKCVRLMWQQRDPALR

>Leish major PABP1(E)

YSGNGNVFVKNLEKDVDSKSLHDIFTKFGSILSCKVMQDEEGKSRGYGFVHFKDET
SAKDAIVKMNGAADHASEDKKALYVANFIRRNARLAALVANFTNVYIKQVLPTVNK

DVIEKFFAKFGGITSAAACKDKSGRVFAFCNFEKHDDAVKAVEAMHDHHIDGITAP
GEKLYVQRAQPRSERLIALRQ

>Leish major PABP1(F)

KYMQHQALGNNLYVRNFDPEFTGADLLELFKEYGEVKSCRVMVSESGVSRGFGFVS
FSNADEANAALREMNGRMLNGKPLIVNIAQRRDQRYTMLRLQFQQRLQMMMRQMHQ
PMPFVGSQGRPMRGRGGRQQLGGRAQGH

>Leish major PABP1(G)

PMPMPSPQQPQAPAQPQGFATPSAVGFVQATPKHSPGDVPETPPLPPITPQELESM
SPQEQRAALGDRLFLKVYEIAPELAPKITGMFLEMKPKEAYELLNDQKRLEERVTE
ALCVLKAHQTA

>Leish major PABP1(H)

RRDQRYTMLRLQFQQRLQMMMRQMHQPMPFVGSQGRPMRGRGGRQQLGGRAQGHPM
PMPSPQQPQAPAQPQGFATPSAVGFVQATPKHSPGDVPETPPLPPITP

>Leish major PABP1(I)

QKYMQHQALGNNLYVRNFDPEFTGADLLELFKEYGEVKSCRVMVSESGVSRGFGFV
SFSNADEANAALREMNGRMLNGKPLIVNIAQRRDQRYTMLRLQFQQRLQMMMRQMH
QPMPFVGSQGRPMRGRGGRQQLGGRAQGHPMPMPSPQQPQAPAQPQGFATPSAVGF
VQATPKHSPGDVPETPPLPPITP

>Leish major PABP1(J)

LPPITPQELESMSPQEQRAALGDRLFLKVYEIAPELAPKITGMFLEMKPKEAYELL
NDQKRLEERVTEALCVLKAHQTA

>Leish major PABP1(L)

RYSGNGNVFVKNLEKDVDSKSLHDIFTKFGSILSCKVMQDEEGKSRGYGFVHFKDE
TSAKDAIVKMNGAADHASEDKKALYVANFIRRNARLA

>Leish major PABP1(M)

LPPITPQELESMSPQEQRAALGDRLFLKVYEIAPELAPKIT

>Leish major PABP1(N)

GMFLEMKPKEAYELLNDQKRLEERVTEALCVLKAHQTA

>Leish major PABP1(O)

RAALGDRLFLKVYEIAPELAPKITGMFLEMKPKEA

>Leish major PABP1(ΔA)

AALVANFTNVYIKQVLPTVNKDVIEKFFAKFGGITSAAACKDKSGRVFAFCNFEKH
DDAVKAVEAMHDHHIDGITAPGEKLYVQRAQPRSERLIALRQKYMQHQALGNNLYV
RNFDPEFTGADLLELFKEYGEVKSCRVMVSESGVSRGFGFVSFSNADEANAALREM
NGRMLNGKPLIVNIAQRRDQRYTMLRLQFQQRLQMMMRQMHQPMPFVGSQGRPMRG
RGGRQQLGGRAQGHPMPMPSPQQPQAPAQPQGFATPSAVGFVQATPKHSPGDVPET
PPLPPITPQELESMSPQEQRAALGDRLFLKVYEIAPELAPKITGMFLEMKPKEAYE
LLNDQKRLEERVTEALCVLKAHQTA

>Leish major PABP1(ΔC)

MAAAVQEAAAPVAHQPQMDKPMQIASIYVGDLDATINEPQLVELFKPFGTILNVRV
CRDIITQRSLGYGYVNFDNHDSAEKAIESMNFKRVGDKCVRLMWQQRDPALRYSGN
GNVFVKNLEKDVDSKSLHDIFTKFGSILSCKVMQDEEGKSRGYGFVHFKDETSAKD
AIVKMNGAADHASEDKKALYVANFIRRNARLAALVANFTNVYIKQVLPTVNKDVIE
KFFAKFGGITSAAACKDKSGRVFAFCNFEKHDDAVKAVEAMHDHHIDGITAPGEKL
YVQRAQPRSERLIALRQKYMQHQALGNNLYVRNFDPEFTGADLLELFKEYGEVKSC
RVMVSESGVSRGFGFVSFSNADEANAALREMNGRMLNGKPLIVNIAQR

>Leish major PABP1(ΔJ)

MAAAVQEAAAPVAHQPQMDKPMQIASIYVGDLDATINEPQLVELFKPFGTILNVRV
CRDIITQRSLGYGYVNFDNHDSAEKAIESMNFKRVGDKCVRLMWQQRDPALRYSGN
GNVFVKNLEKDVDSKSLHDIFTKFGSILSCKVMQDEEGKSRGYGFVHFKDETSAKD
AIVKMNGAADHASEDKKALYVANFIRRNARLAALVANFTNVYIKQVLPTVNKDVIE
KFFAKFGGITSAAACKDKSGRVFAFCNFEKHDDAVKAVEAMHDHHIDGITAPGEKL
YVQRAQPRSERLIALRQKYMQHQALGNNLYVRNFDPEFTGADLLELFKEYGEVKSC
RVMVSESGVSRGFGFVSFSNADEANAALREMNGRMLNGKPLIVNIAQRRDQRYTML
RLQFQQRLQMMMRQMHQPMPFVGSQGRPMRGRGGRQQLGGRAQGHPMPMPSPQQPQ
APAQPQGFATPSAVGFVQATPKHSPGDVPETPPL

>Leish major eIF4E4

MSTPLDVRAAEYSPSFAVTMKKTVAAAPPKSPAPAKSKISVTRTGVNTTYPMPPPM
PEKNYAPFFAEGCQTIAASKASMPPVQPASPLPPMHSAPPTASVVSNSIPPSSPAT
APGERSPAVAARSVPTRFSPATVPRHHMNPNATEFMPGRRNGPDGGLEALPTSTAD
MELAKTPAGAAAAAVHAPSLPGAVRRSLQNSPIIQPSRLSVKSASEIEAISKNSAL
NAAAAAYVPQRTLARVVLTQPSPLALAPSEDPAKNNIEMMLDDLWCLFYLPTTLGE
NIKEEDYNPTLVFRVDSILTFWRVVNNIAAPSELQLSTLYLFRDGIDPKWEDPANR
DGGIVKVKATAAQVDEAWELLLCRTIGDSWSPSVRETVNGVVLKVRERAYWLELWV
TKNSSALQKDLAELWHPILGASFATTYLTHAMMQERSHAAAALAAEKQKKNRRRY

>Leish major eIF4E4(iv)

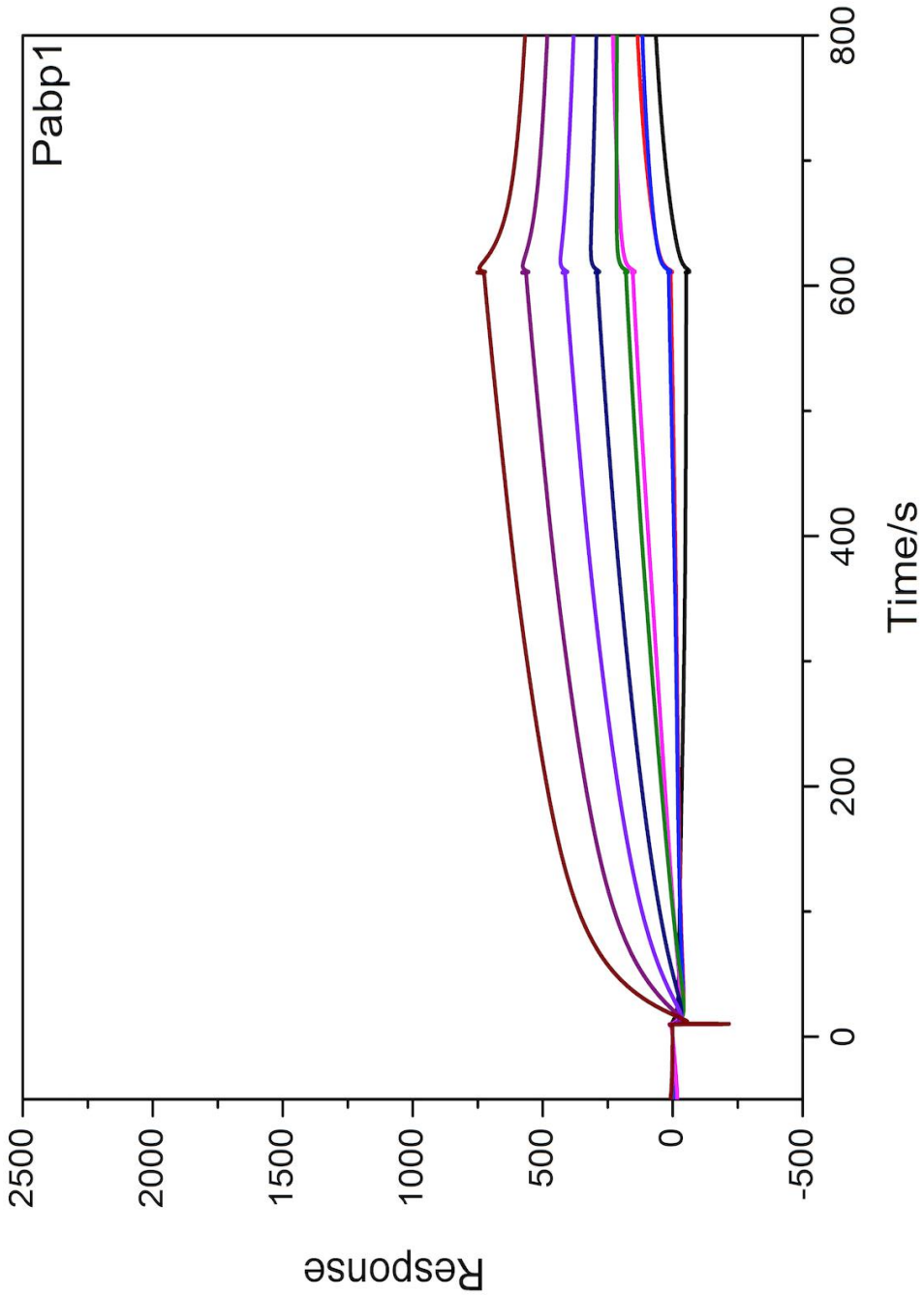
MNPNATEFMPGRRNGPDGGLEALPTSTADMELAKTPAGAAAAAVHAPSLPGAVRRS
LQNSPIIQPSRLSVKSASEIEAISKNSALNA

>Leish major eIF4E4(v)

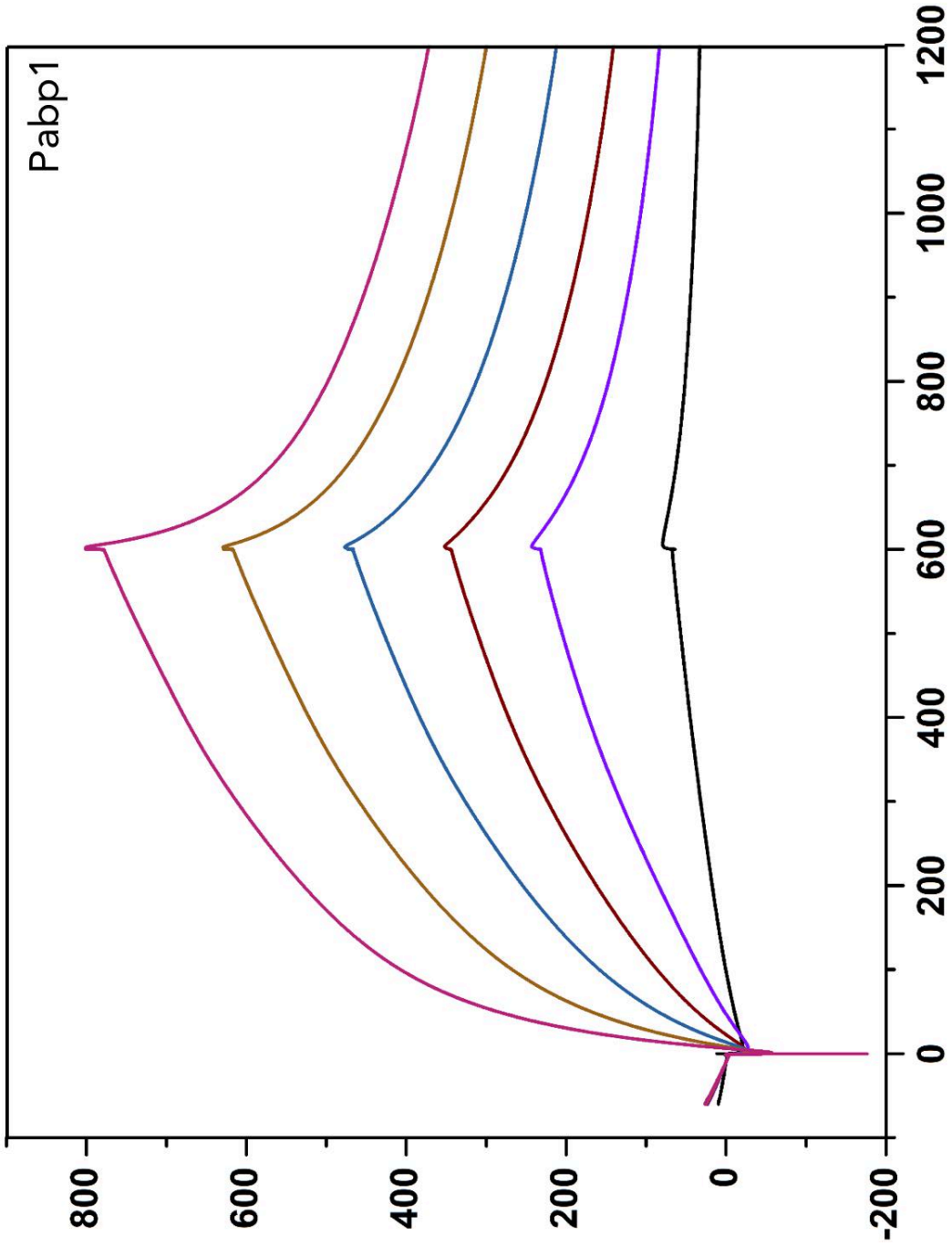
MNPNATEFMPGRRNGPDGGLEALPTSTADMELAKTPAGAAAAAVHAPSLPGAVRRS
LQNSPIIQPSRLSVKSASEIEAISKNSALNAAAAAYVPQRTLARVVLTQPSPLALA
PSEDPAKNNIEMMLDDLWCLFYLPTTLGENIKEEDYNPTLVFRVDSILTFWRVVNN
IAAPSELQLSTLYLFRDGIDPKWEDPANRDGGIVKVKATAAQVDEAWELLLCRTIG
DSWSPSVRETVNGVVLKVRERAYWLELWVTKNSSALQKDLAELWHPILGASFATTY
LTHAMMQERSHAAAALAAEKQKKNRRRY

SPR SENSORGRAMS

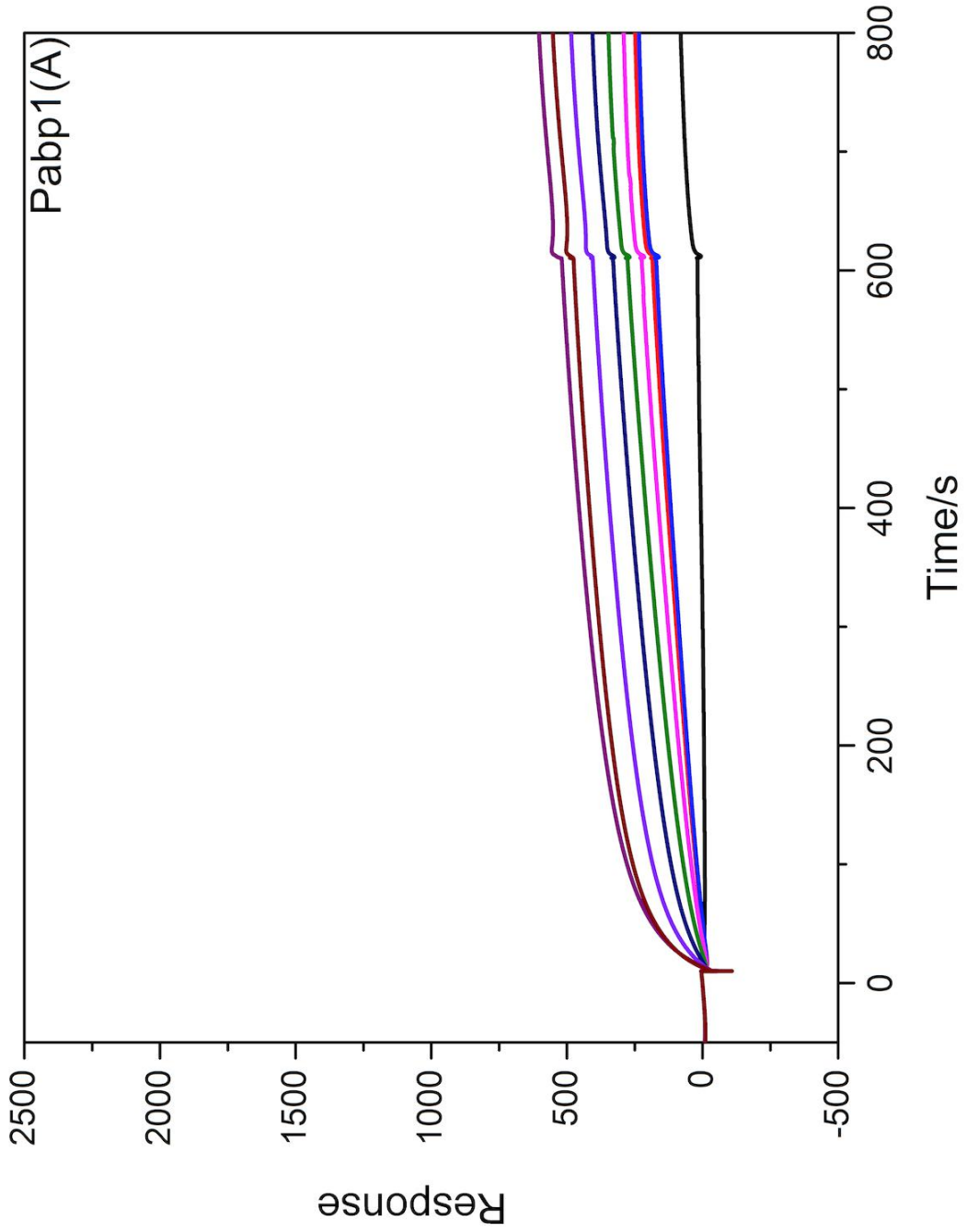
eIF4E4(iv)/ PABP full length – Initial experiment



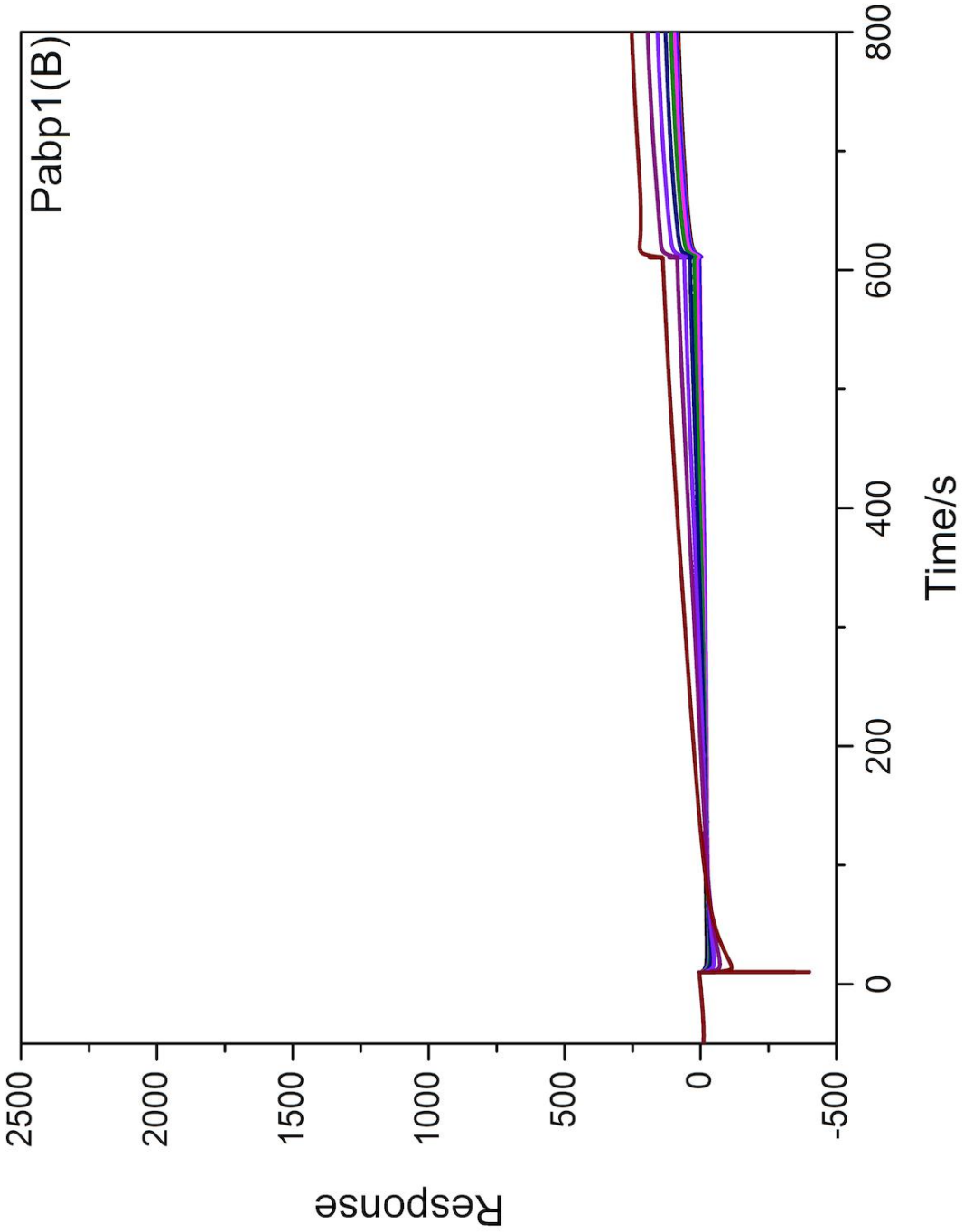
eIF4E4(iv)/ PABP full length – Optimized experiment



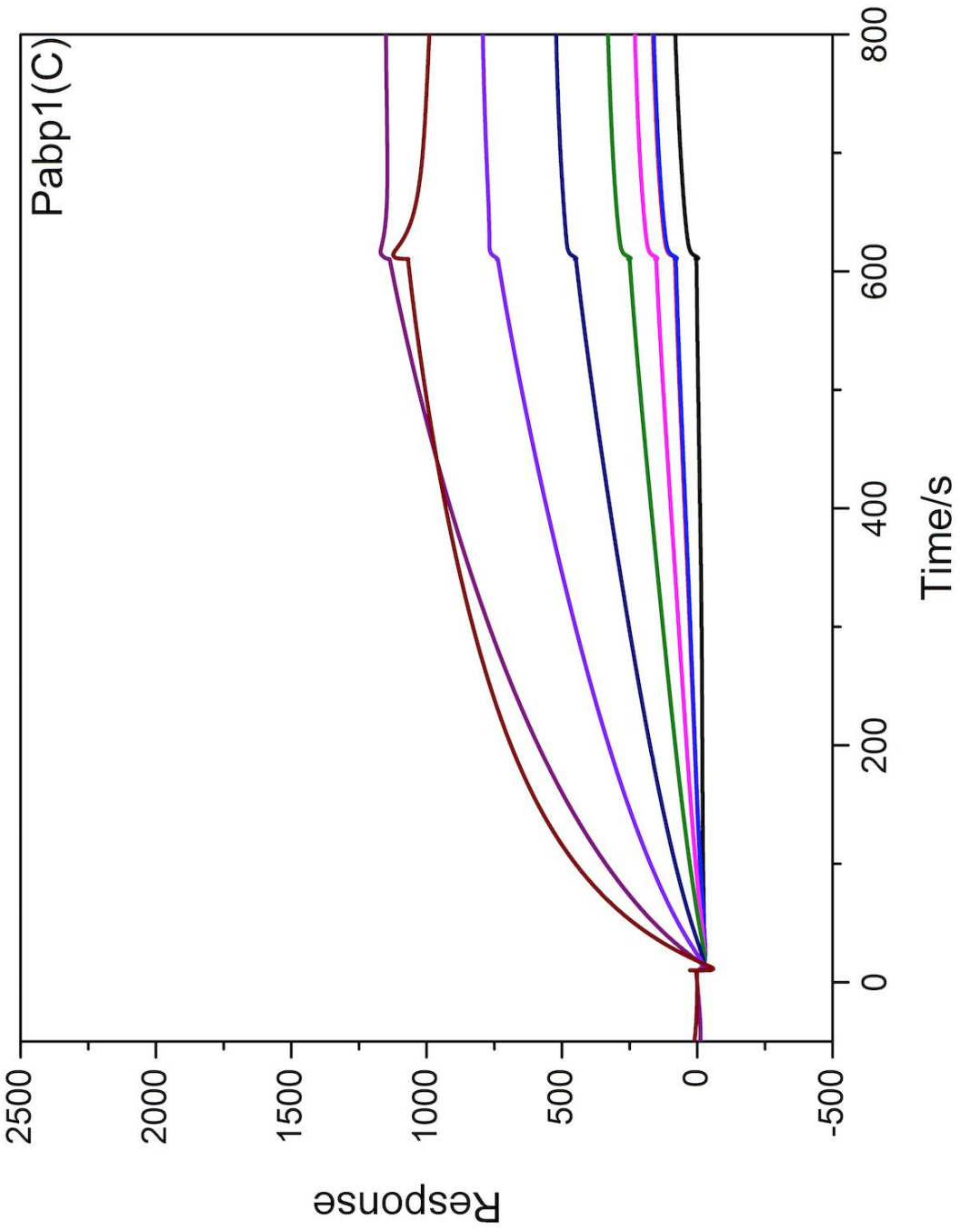
eIF4E4(iv)/ PABP1(A)



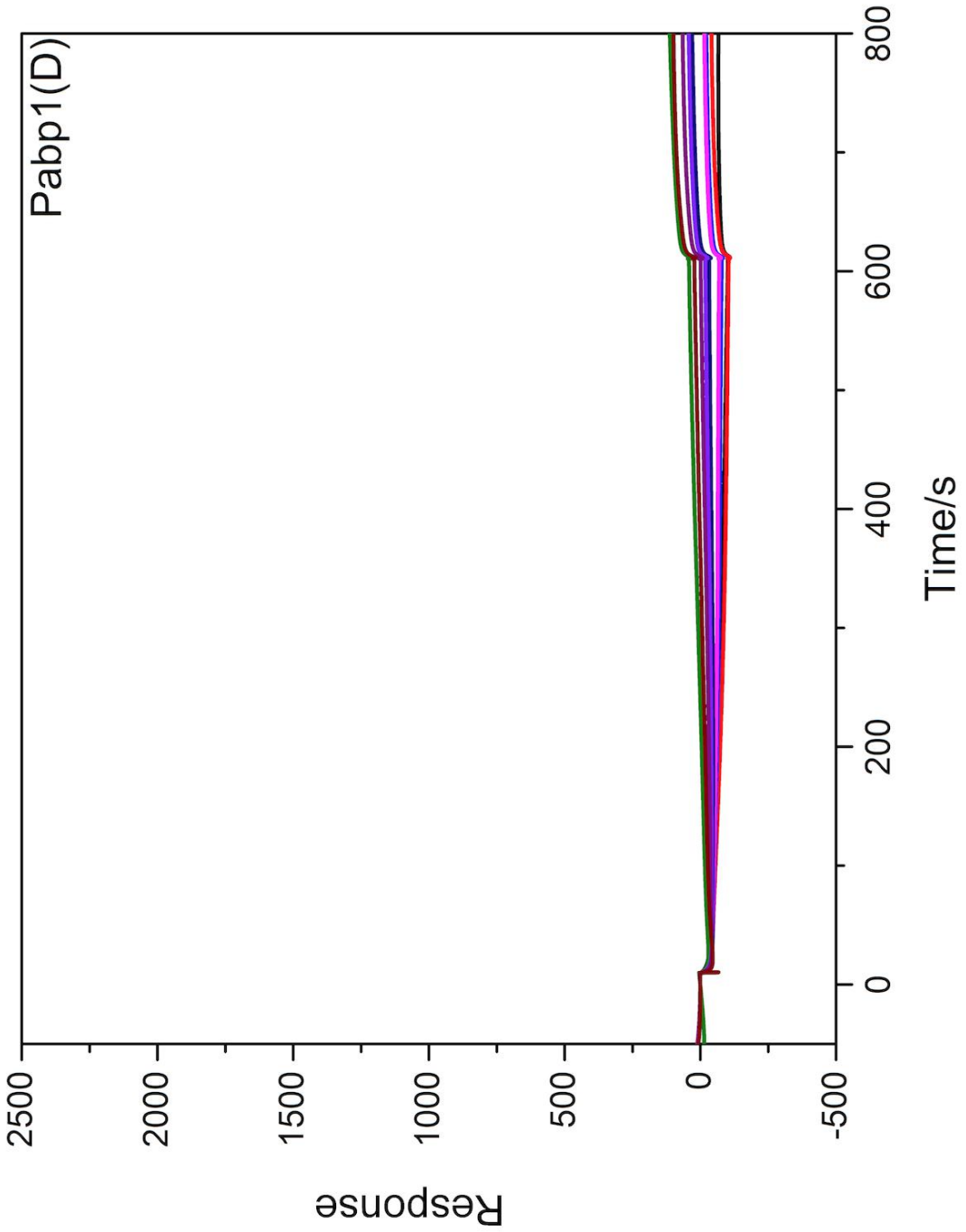
eIF4E4(iv)/ PABP1(B)



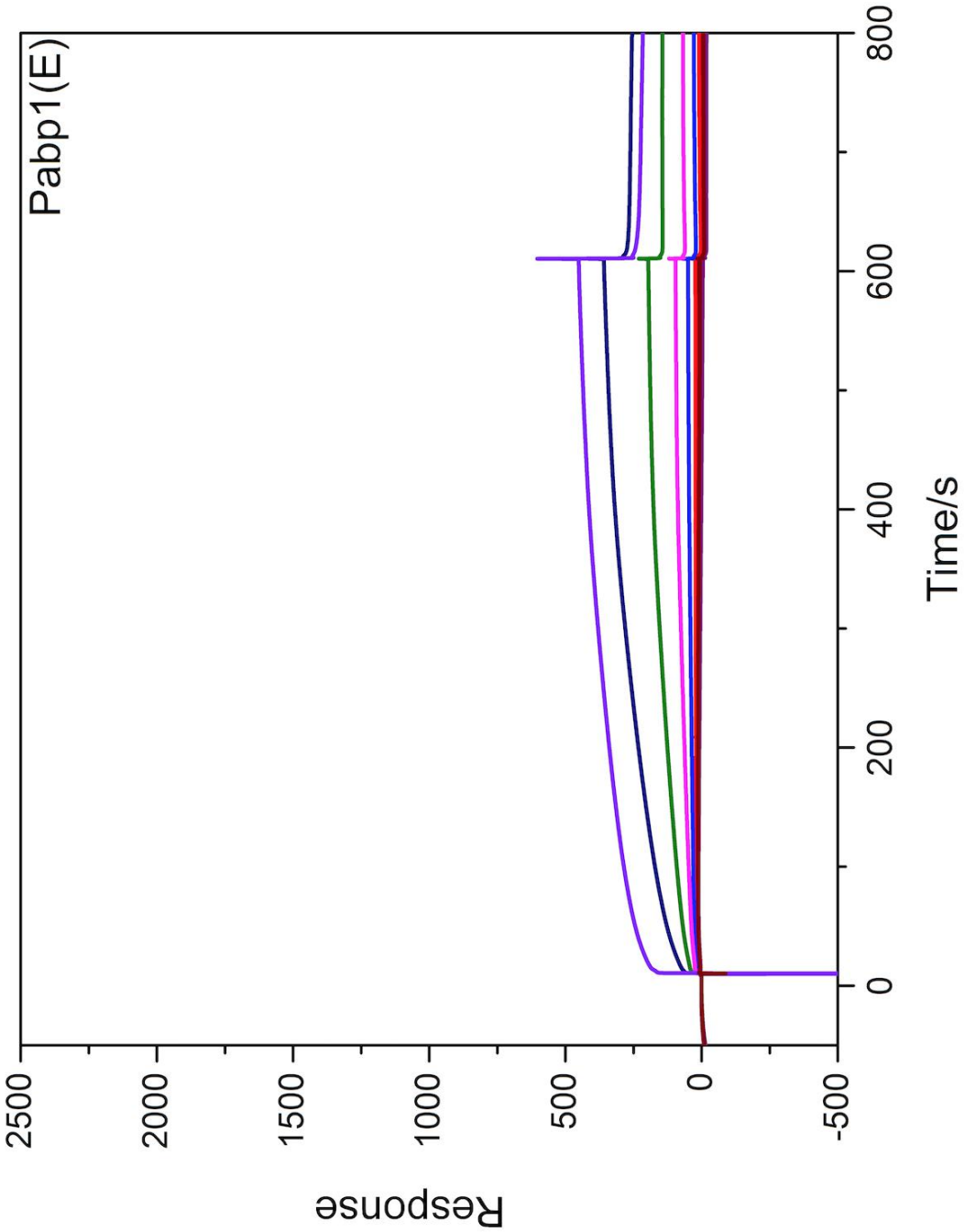
eIF4E4(iv)/ PABP1(C)



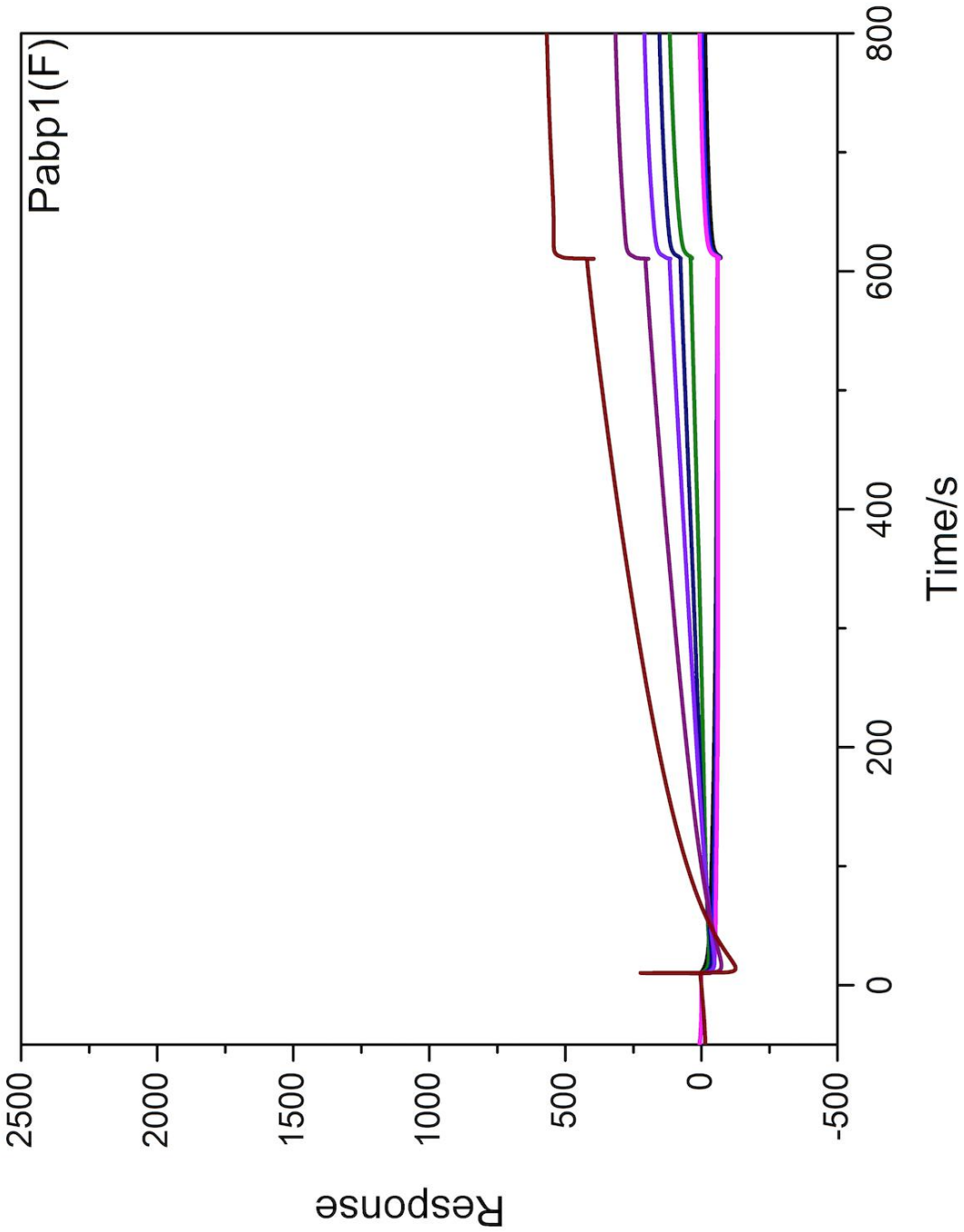
eIF4E4(iv)/ PABP1(D)



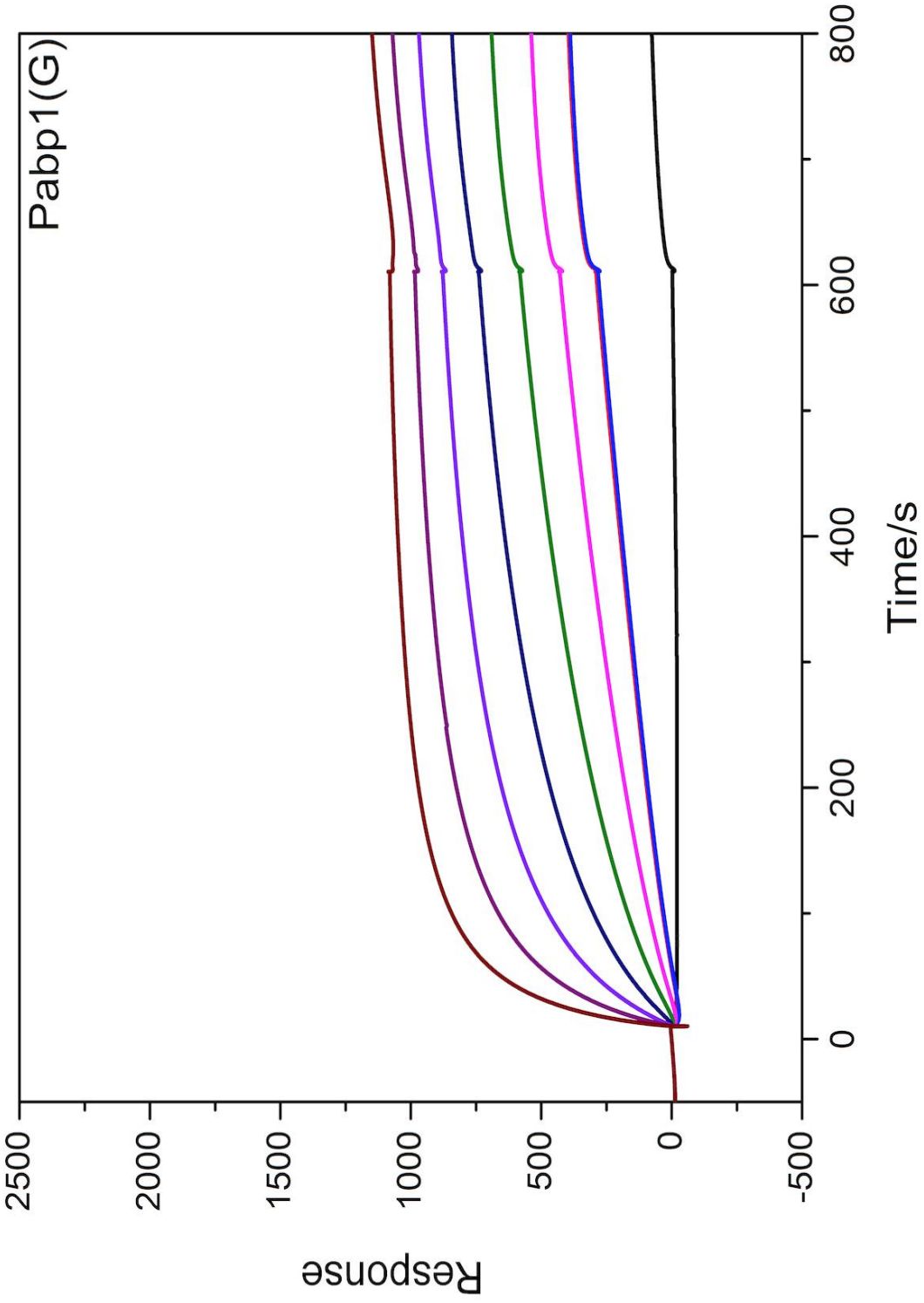
eIF4E4(iv)/ PABP1(E)



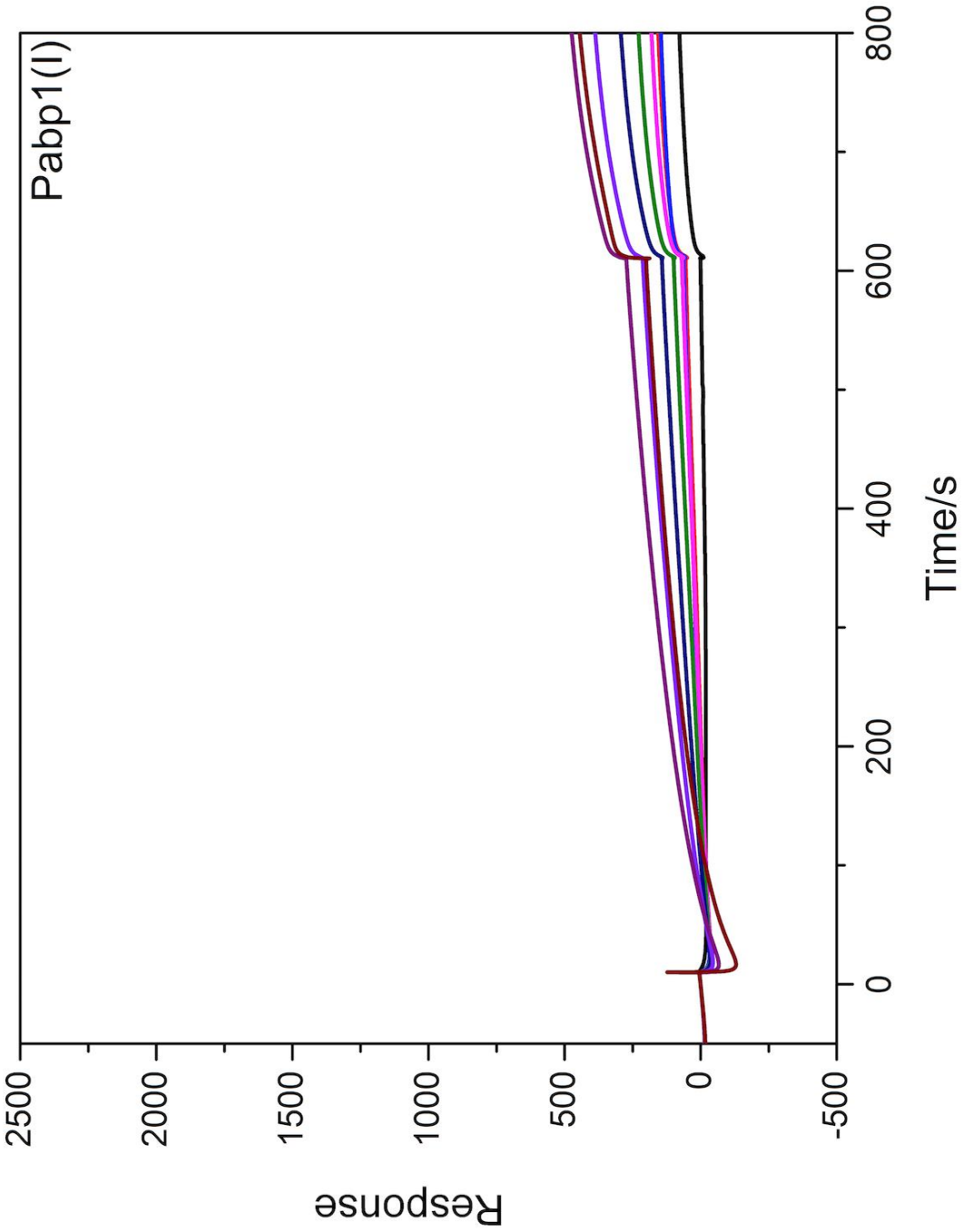
eIF4E4(iv)/ PABP1(F)



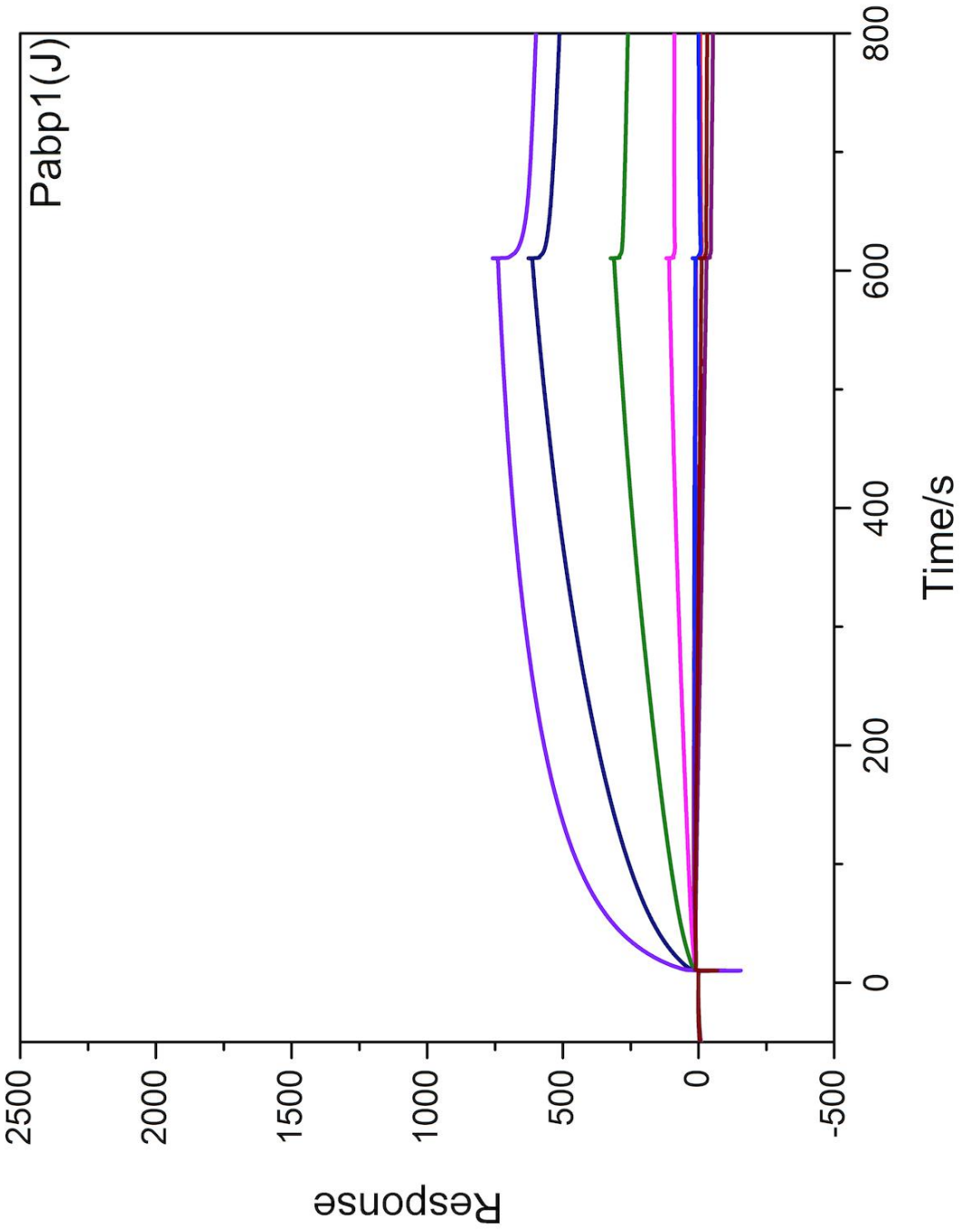
eIF4E4(iv)/ PABP1(G)



eIF4E4(iv)/ PABP1(I)



eIF4E4(iv)/ PABP1(J)



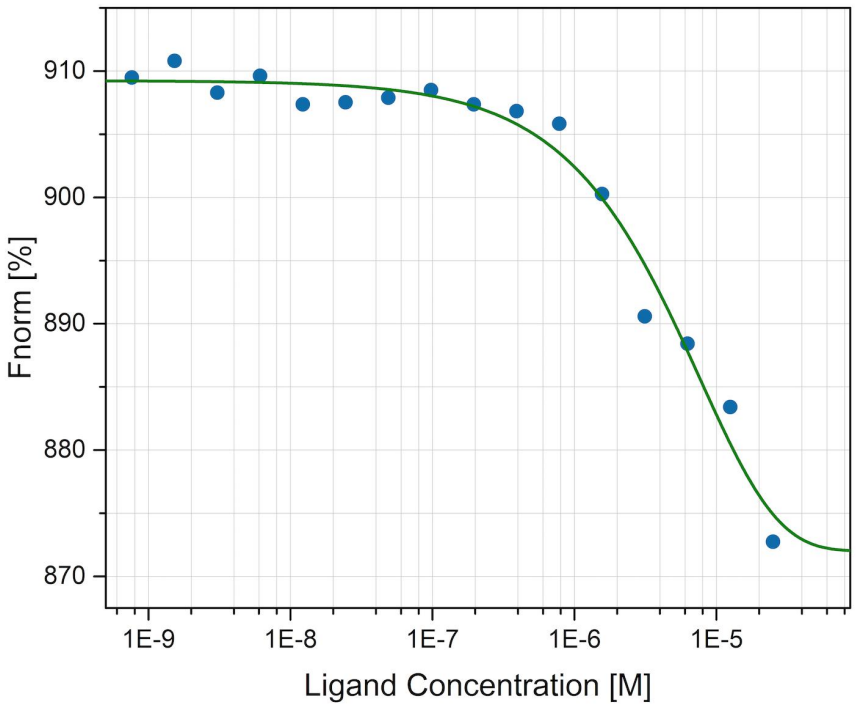
Protein Concentrations for Binding Affinity Experiments

PABP1/μM	PABP1 Optimized/μM	PABP1(A)/μM	PABP1(B)/μM	PABP1(C)/μM	PABP1(D)/μM
1.55	1.49	21.20	1.64	2.31	21.14
0.78	0.74	10.60	0.82	1.16	10.57
0.39	0.37	5.30	0.41	0.58	5.29
0.19	0.19	2.65	0.21	0.29	2.64
0.10	0.09	1.33	0.10	0.14	1.32
0.05	0.02	0.66	0.05	0.07	0.66

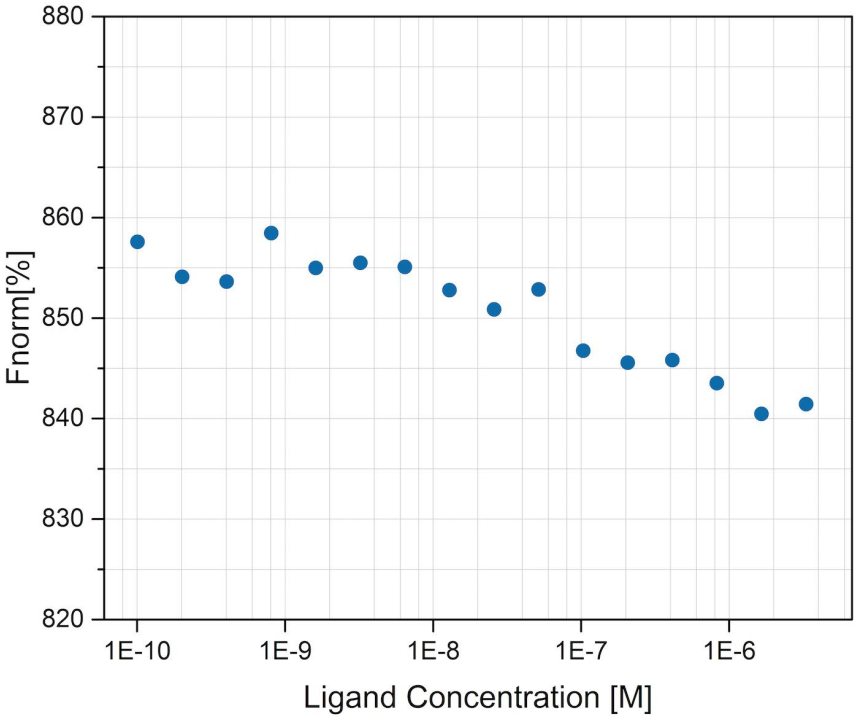
PABP1(E)/μM	PABP1(F)/μM	PABP1(G)/μM	PABP1(I)/μM	PABP1(J)/μM
8.99	21.18	22.08	20.09	21.20
4.50	10.59	11.04	10.05	10.60
2.25	5.30	5.52	5.02	5.30
1.12	2.65	2.76	2.51	2.65
0.56	1.32	1.38	1.26	1.33
0.28	0.66	0.69	0.63	0.66

MICROSCALE THERMOPHORESIS DATA

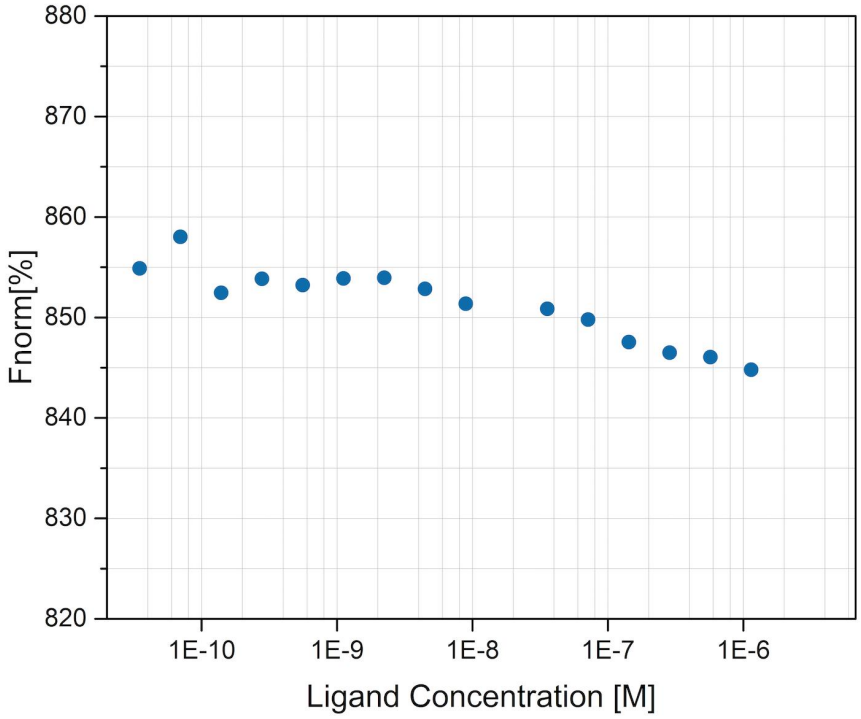
eIF4E4(iv)/ PABP1 full length



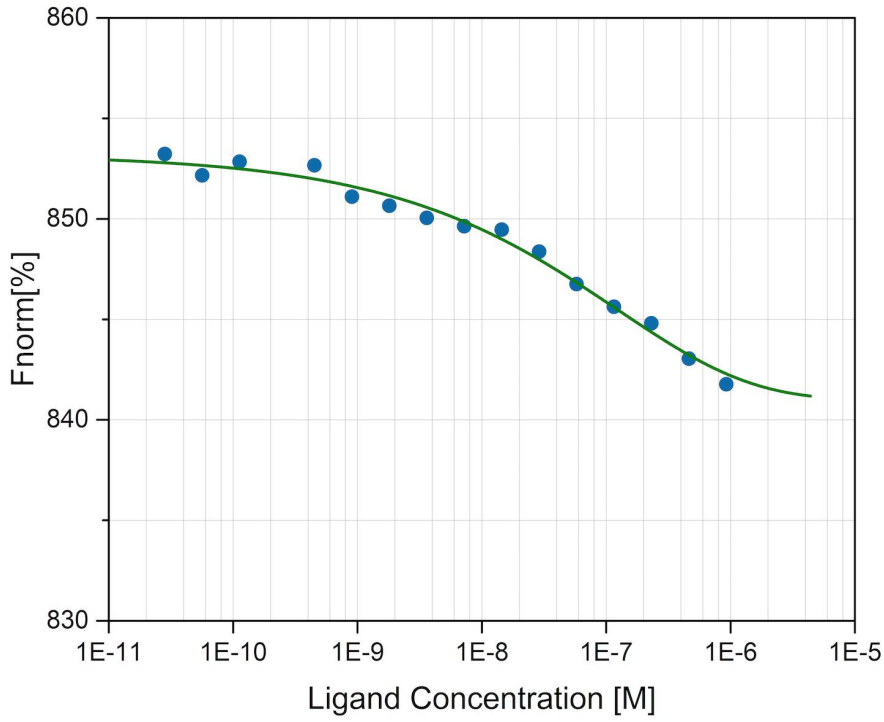
eIF4E4(iv)/ PABP1(A)



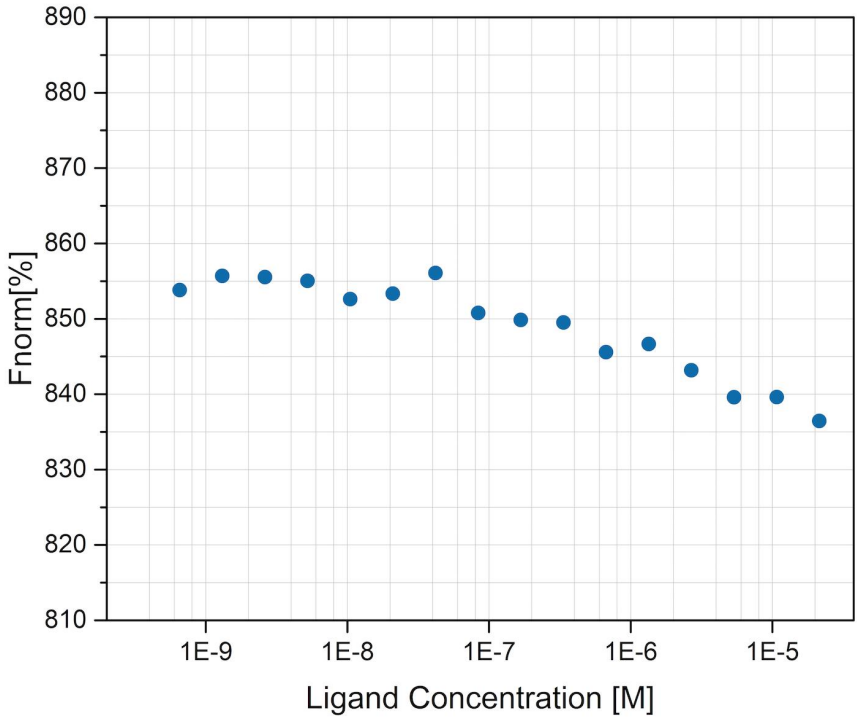
eIF4E4(iv)/ PABP1(B)



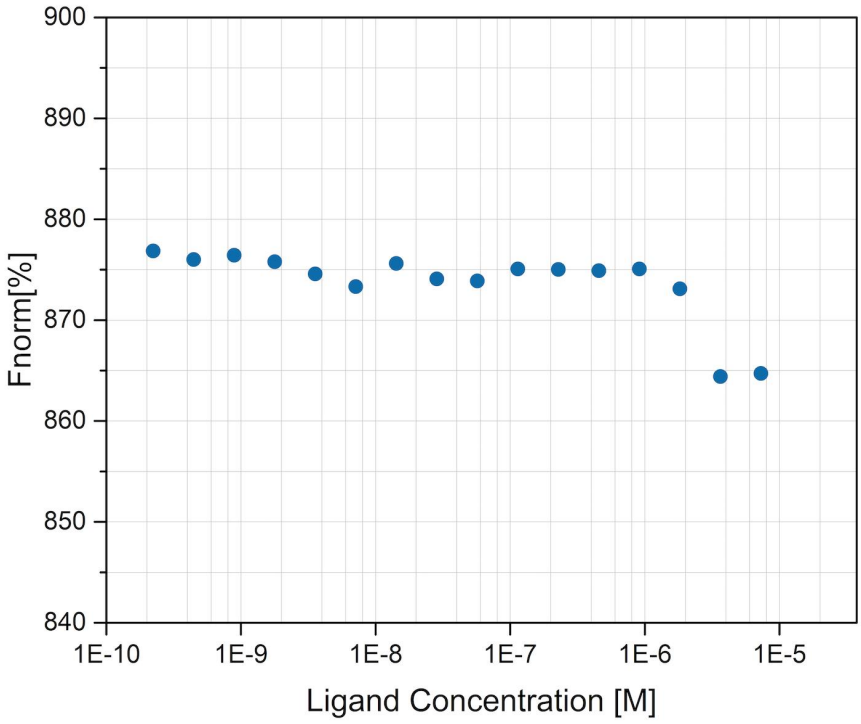
eIF4E4(iv)/ PABP1(C)



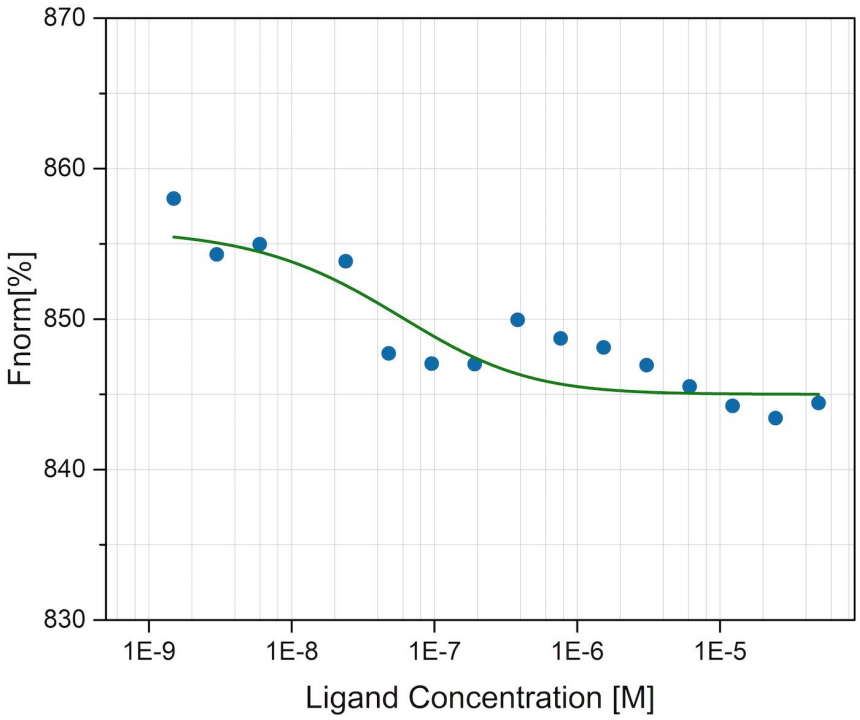
eIF4E4(iv)/ PABP1(D)



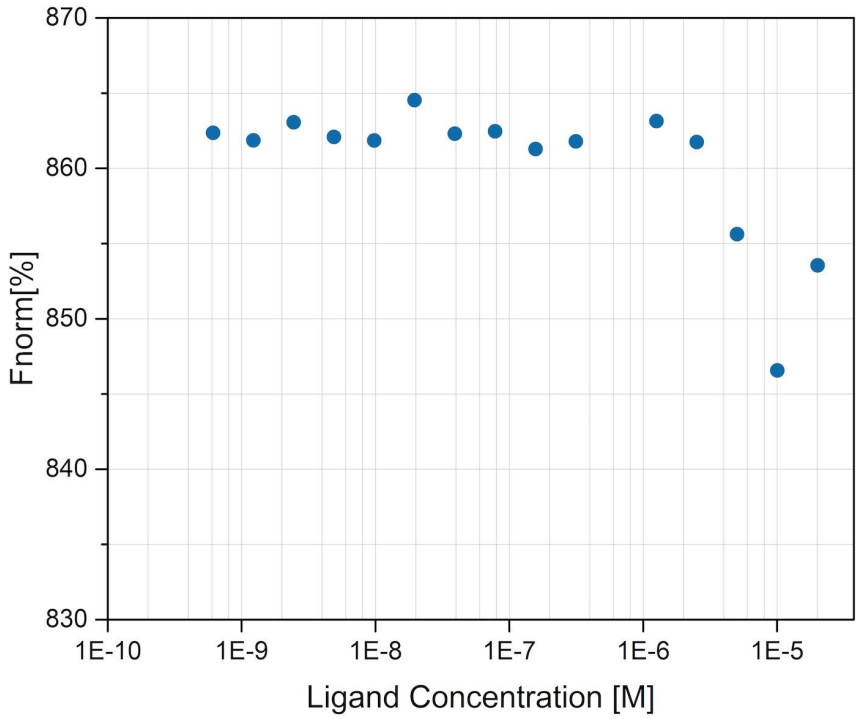
eIF4E4(iv)/ PABP1(F)



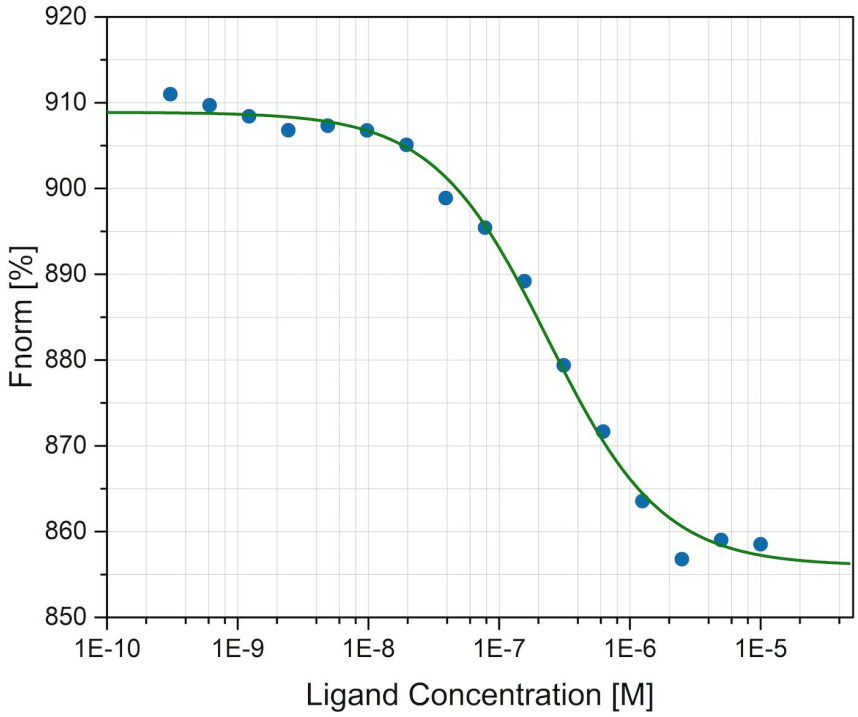
eIF4E4(iv)/ PABP1(G)



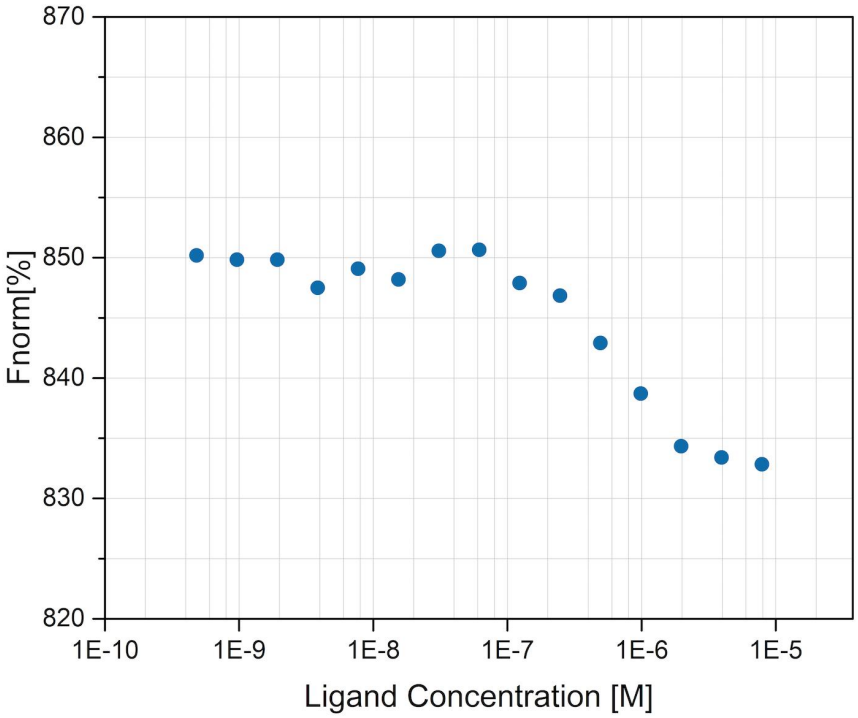
eIF4E4(iv)/ PABP1(I)



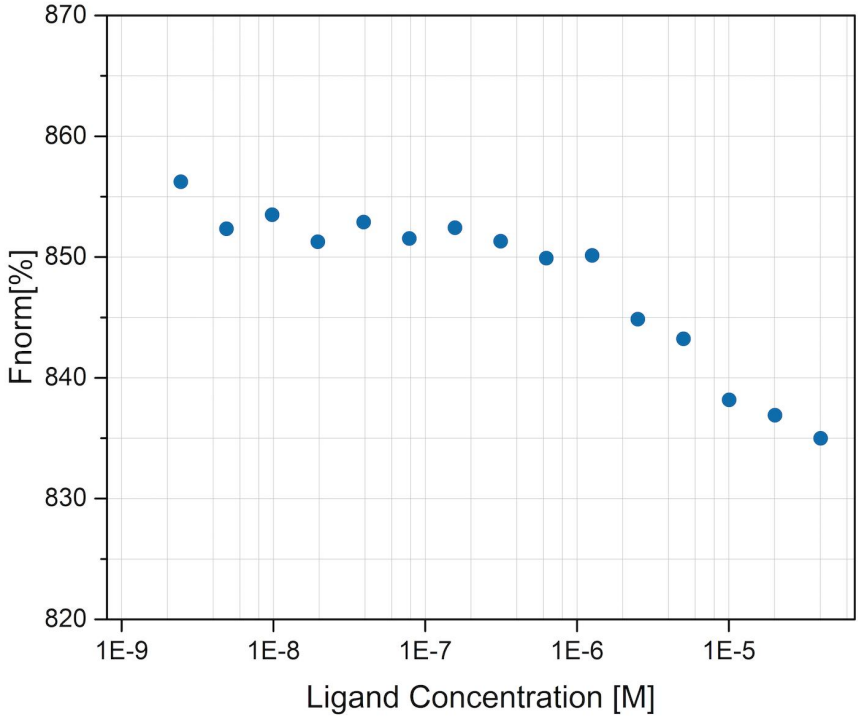
eIF4E4(iv)/ PABP1(J)



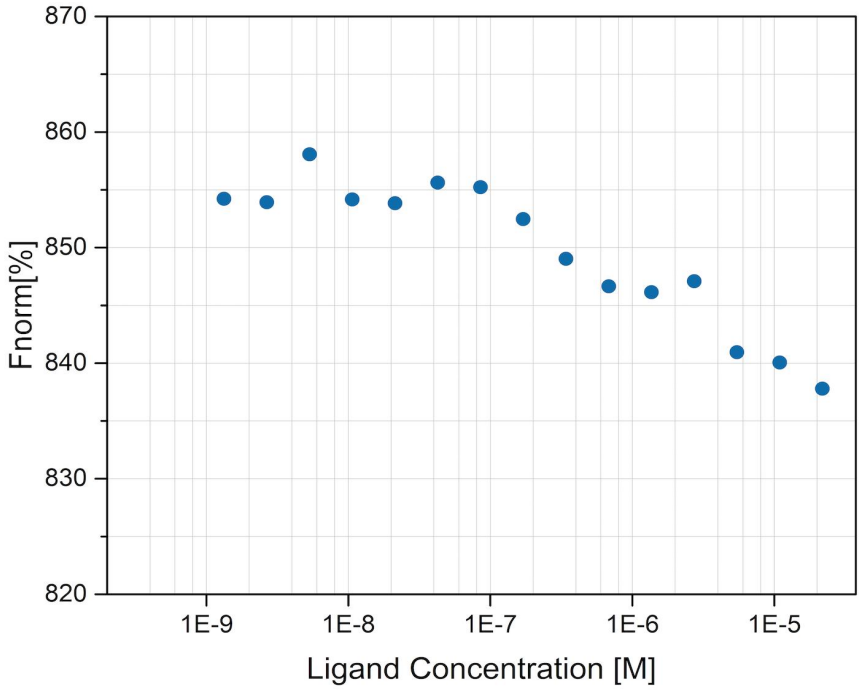
eIF4E4(iv)/ PABP1(L)



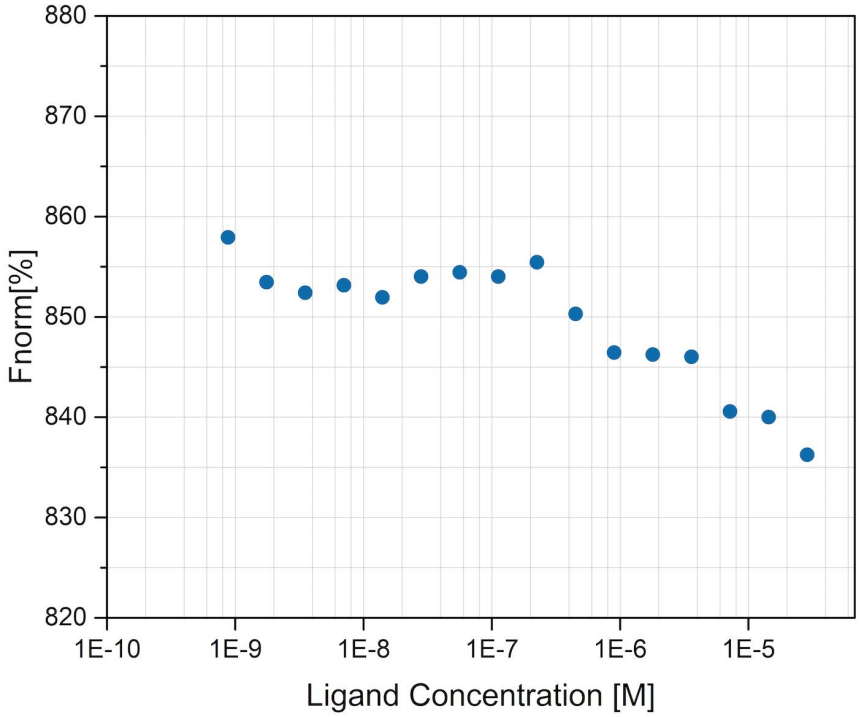
eIF4E4(iv)/ PABP1(M)



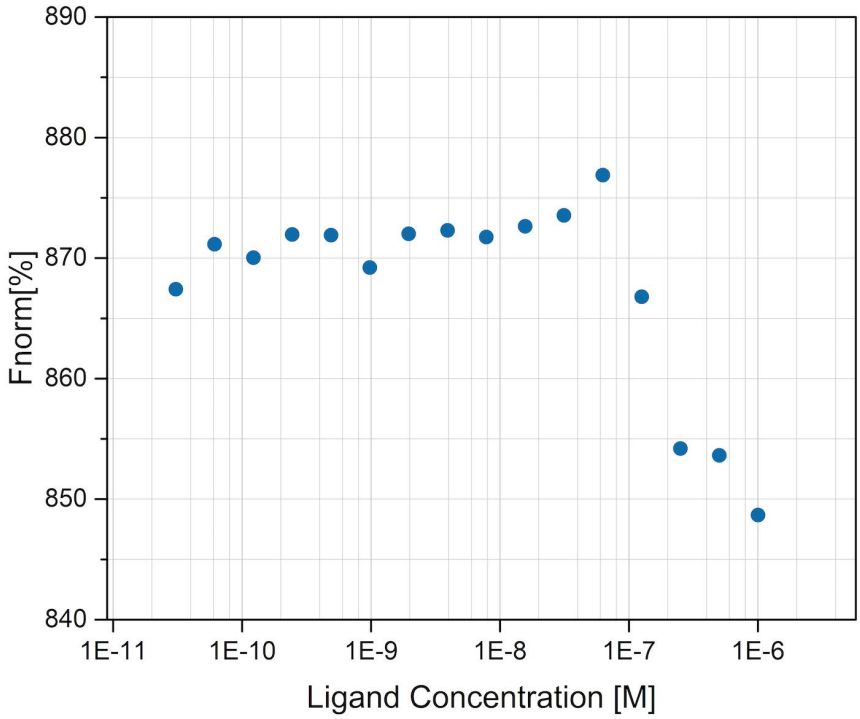
eIF4E4(iv)/ PABP1(N)



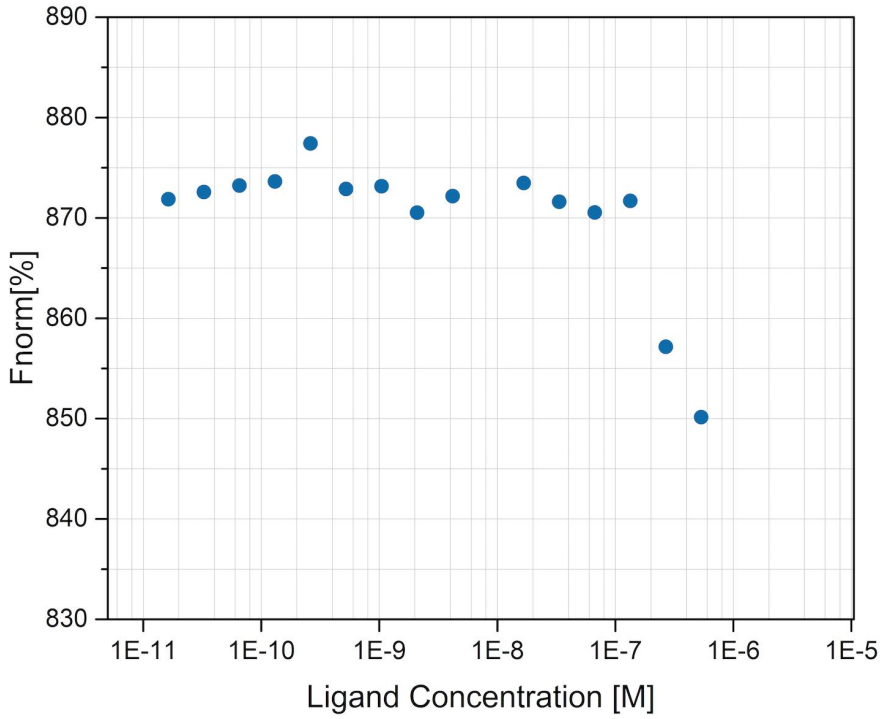
eIF4E4(iv)/ PABP1(O)



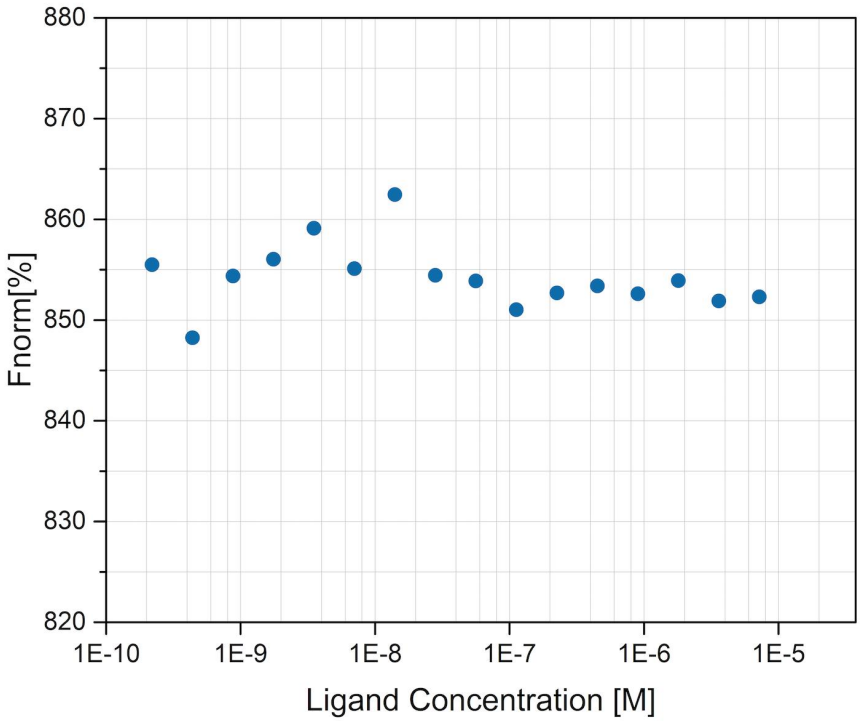
eIF4E4(iv)/ PABP1(ΔC)



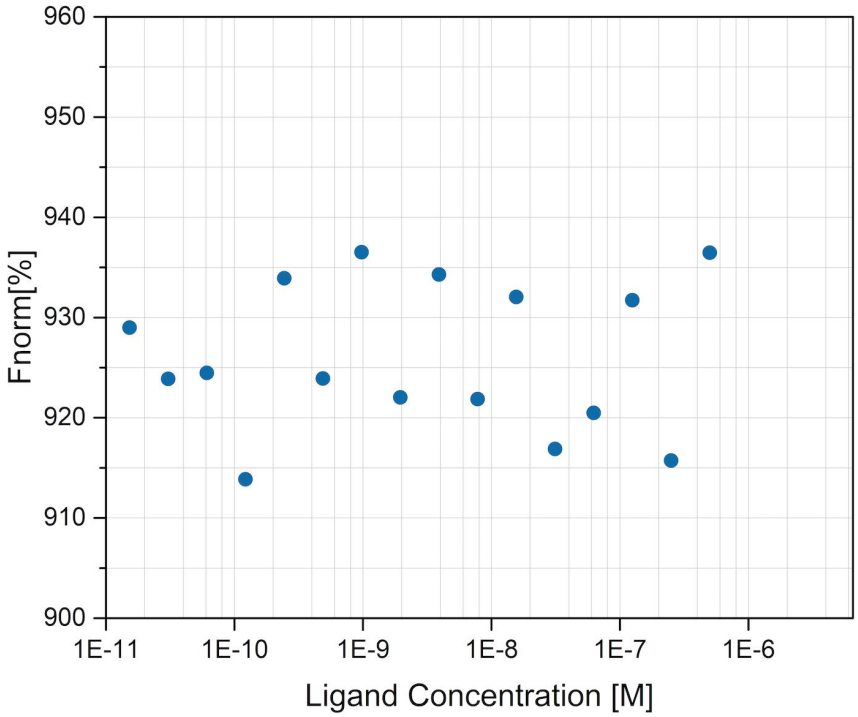
eIF4E4(iv)/ PABP1(ΔJ)



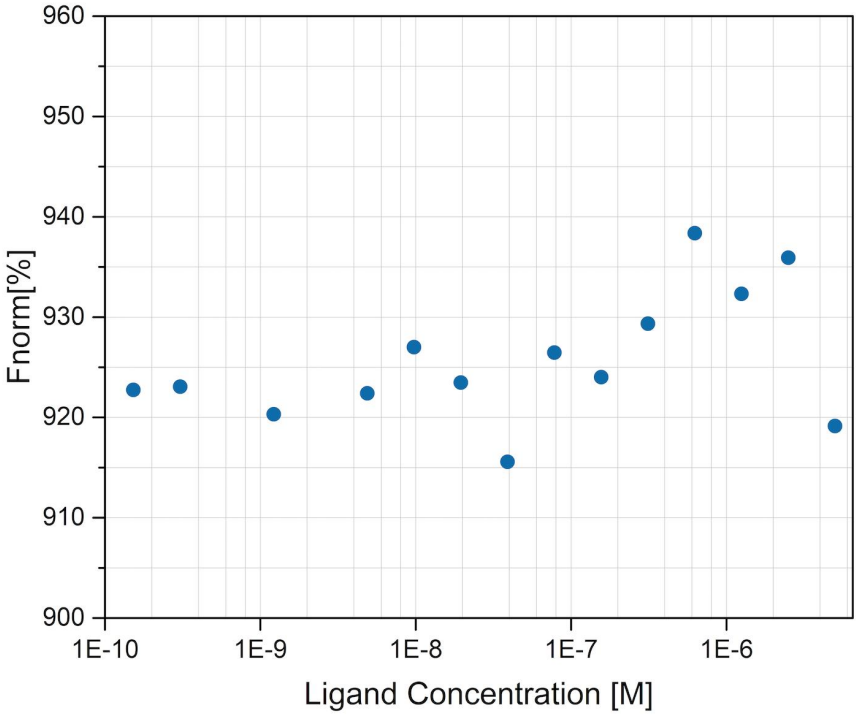
eIF4E4(iv)/BSA



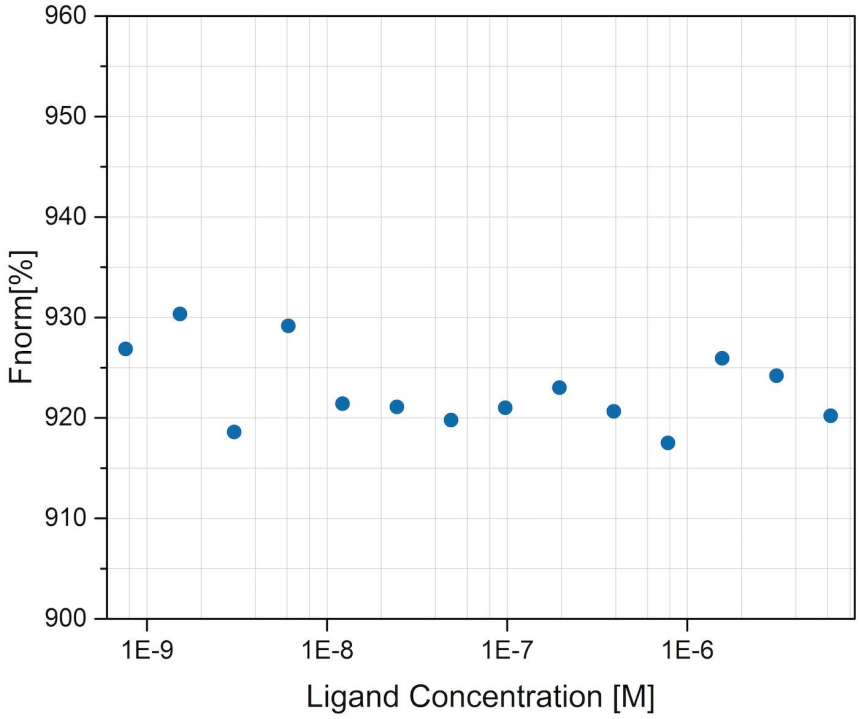
eIF4G3+eIF4E4(iv)



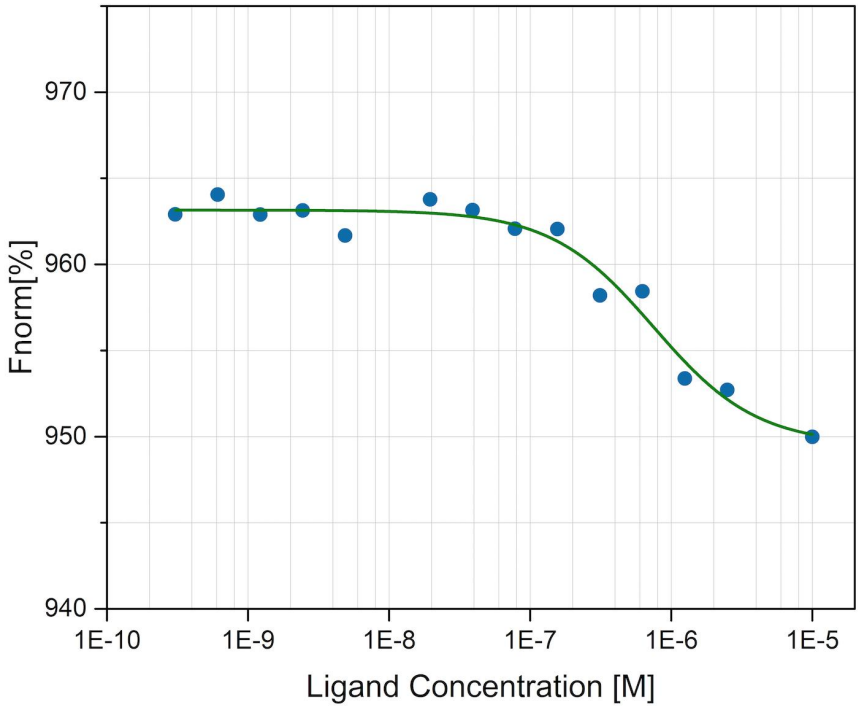
eIF4G3+ PABP1 FL



eIF4G3+ PABP1(J)

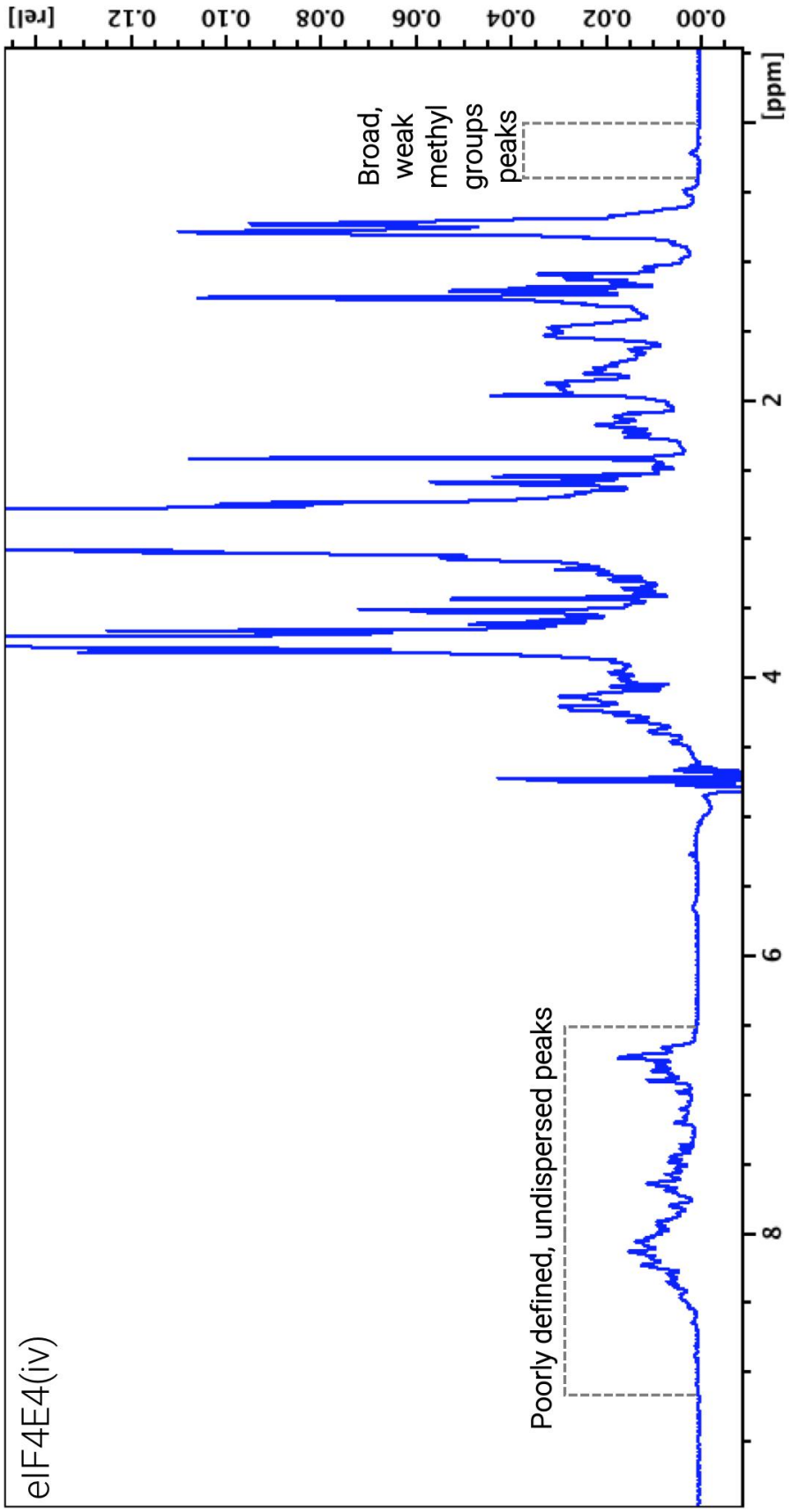


eIF4G3+{eIF4E4(v)/ PABP1(J)} complex

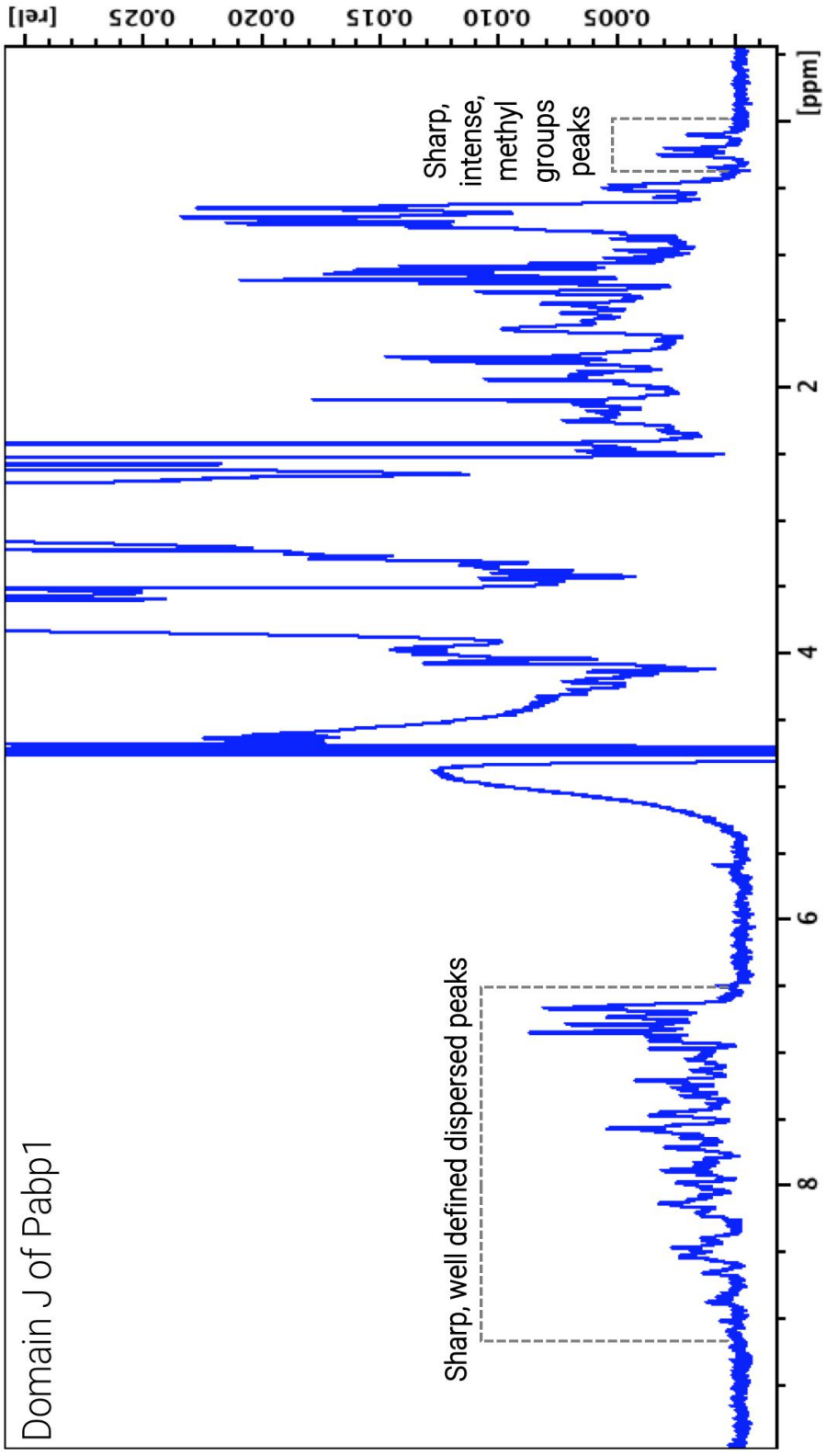


NUCLEAR MAGNETIC RESONANCE SPECTRA

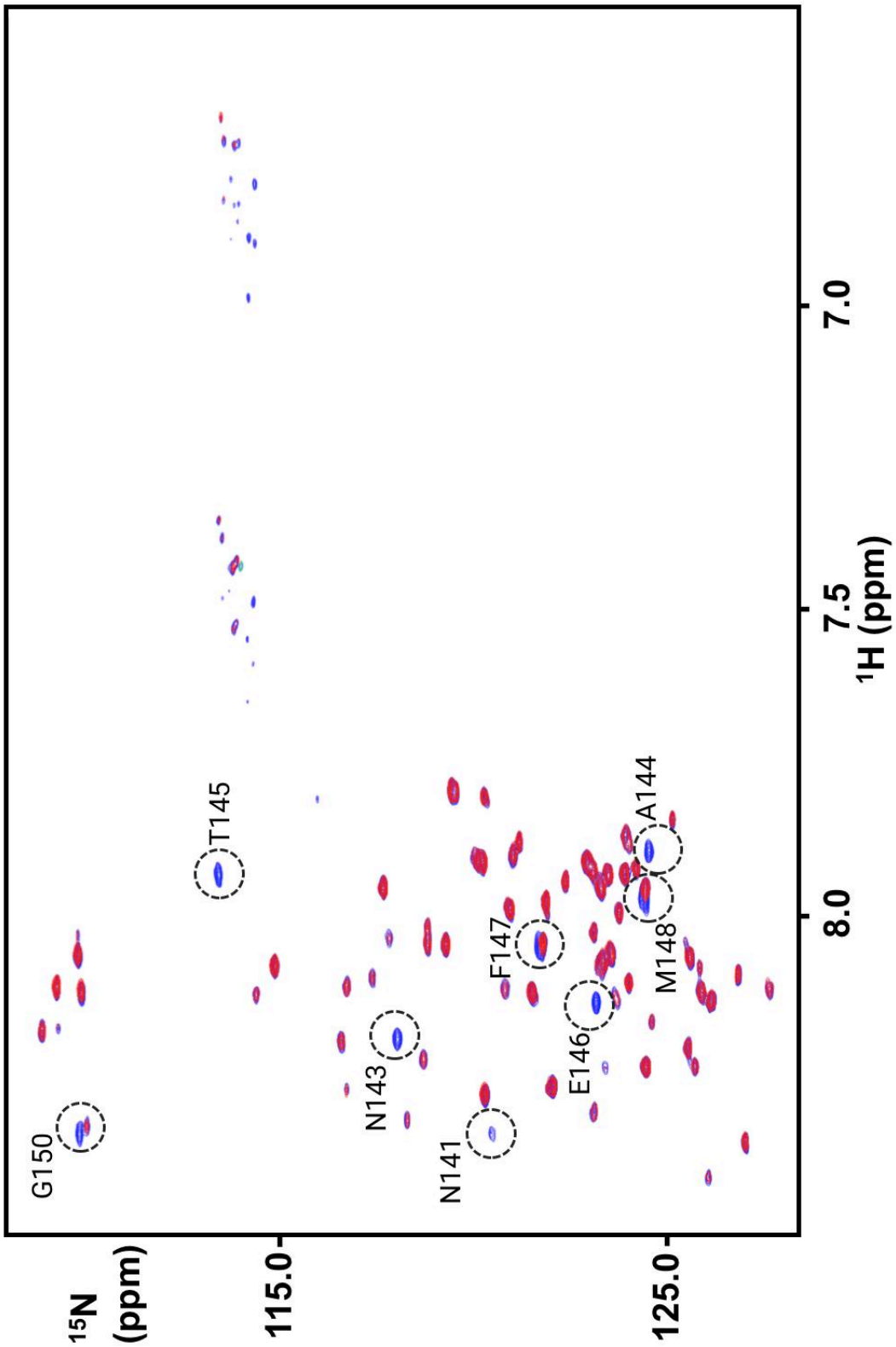
[1]H Isotopically labelled eIF4E4(iv)



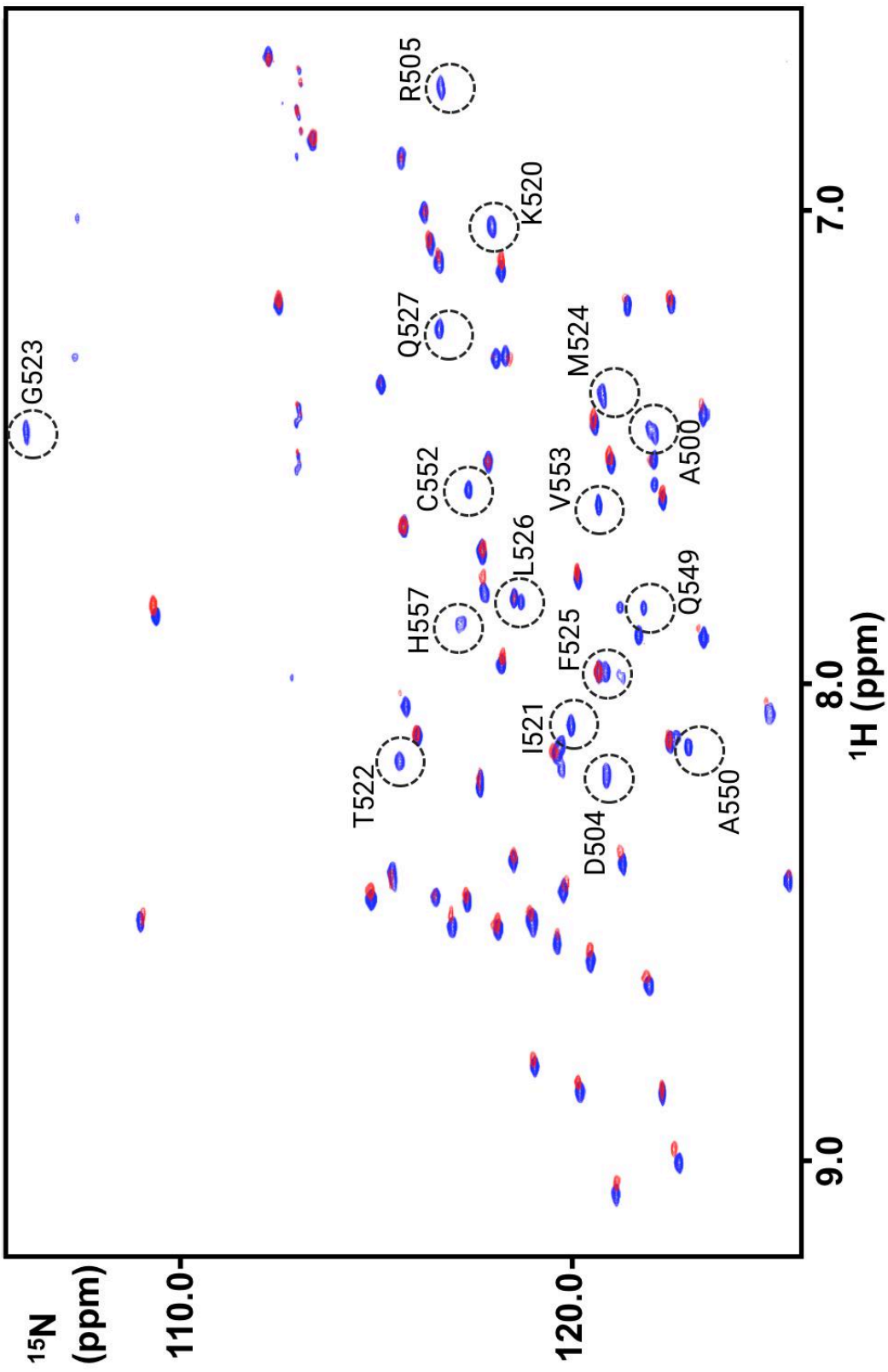
[1]H Isotopically labelled PABP1(J)



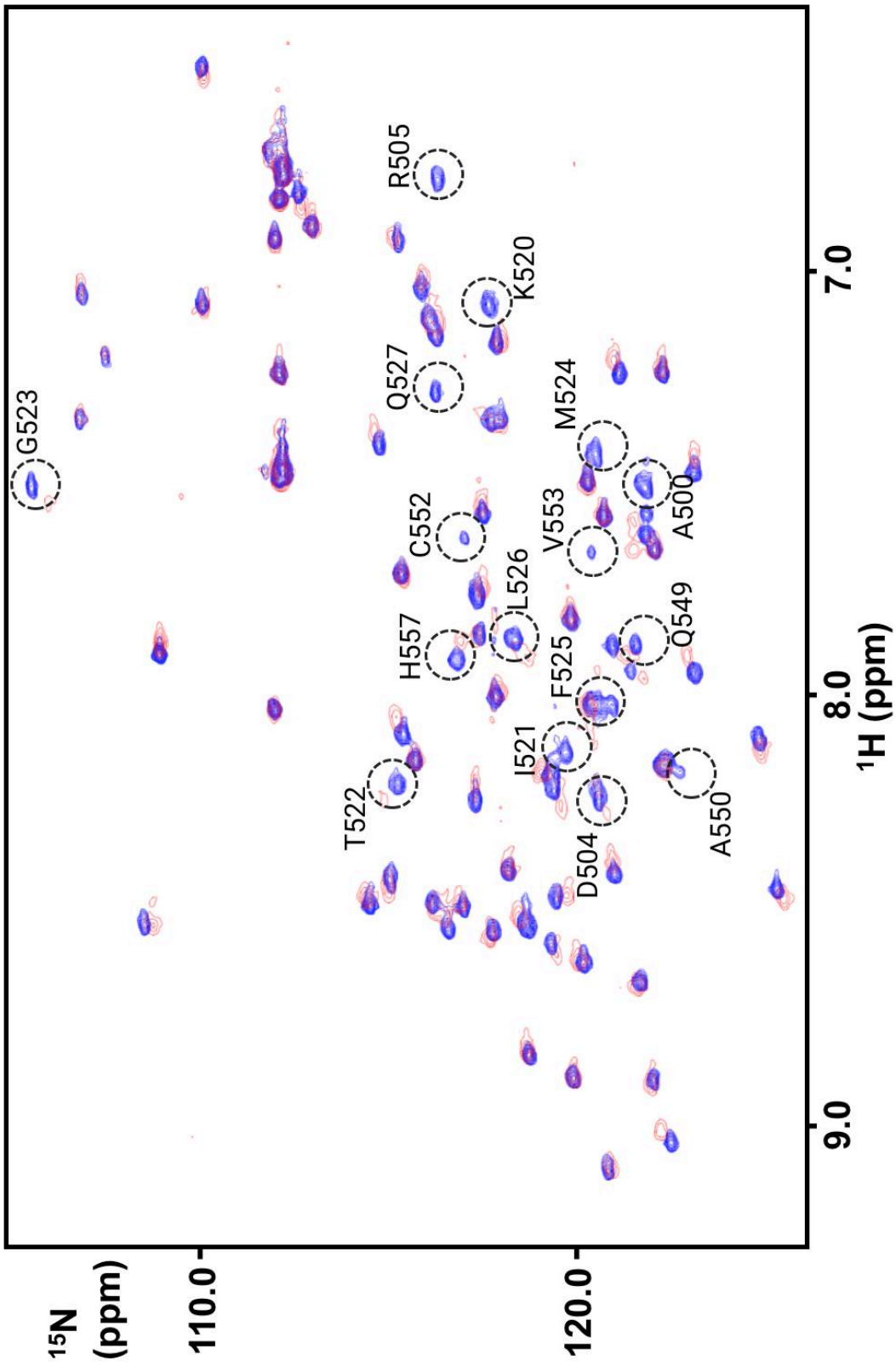
{^{1}H-^{15}N}-HSQC - Isotopically labelled eIF4E4(iv) + unlabelled PABP1(J)



{^{1}H-^{15}N}-HSQC Isotopically labelled PABP1(J) + unlabelled eIF4E4(iv)

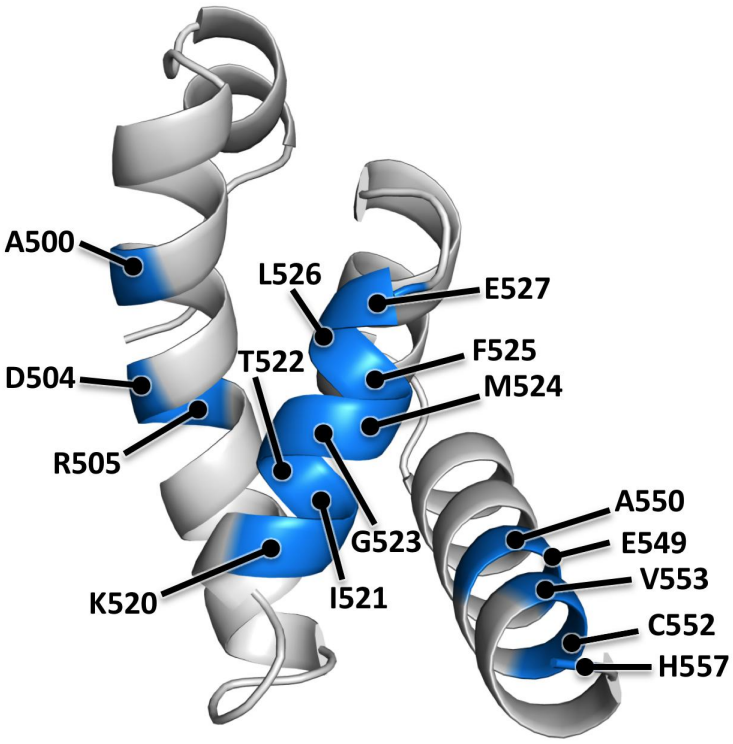


{^{1}H-^{15}N}-HSQC Isotopically labelled PABP1(J) + unlabelled eIF4E4-PAM2 peptide

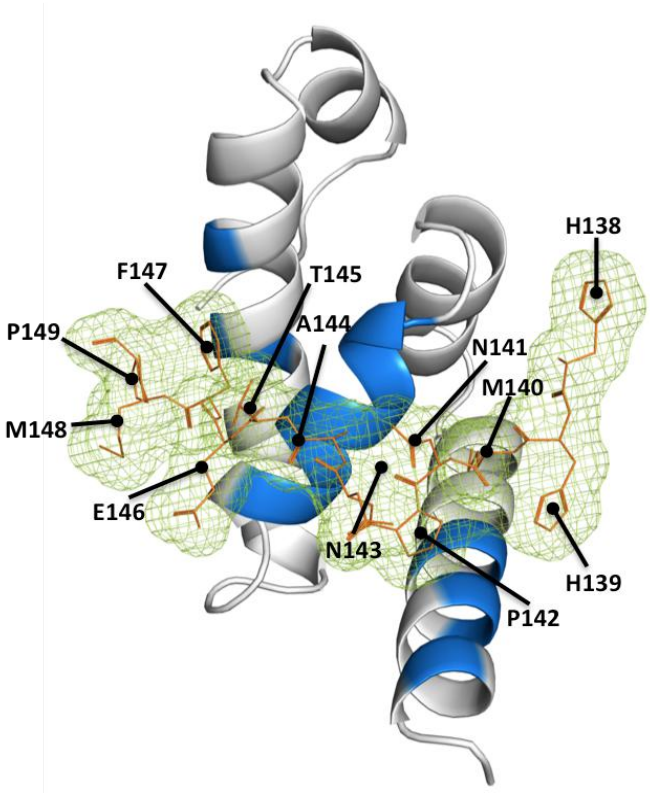


CRYSTALLOGRAPHIC DATA

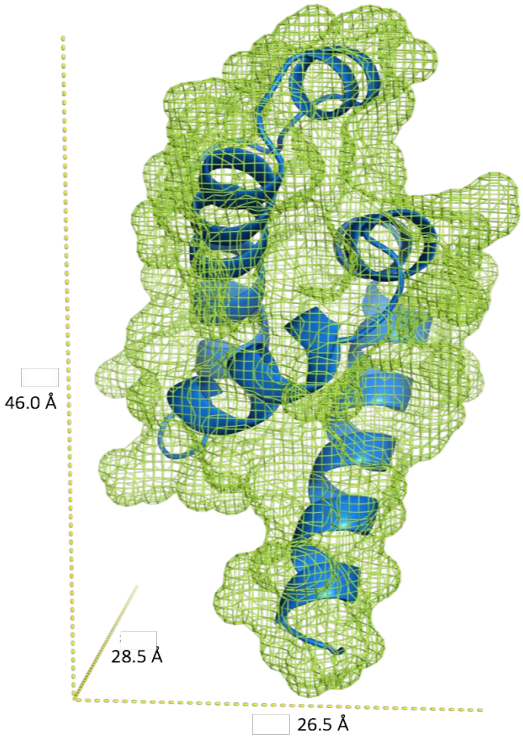
PABP1(J) Structure



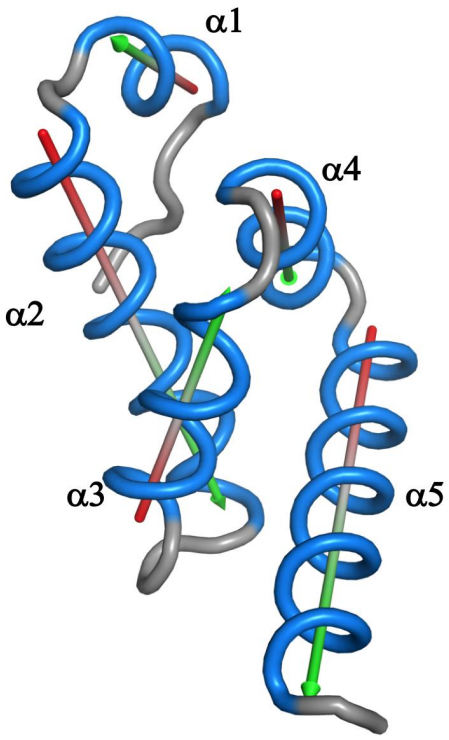
PABP1(J) Structure+ PAM2 peptide



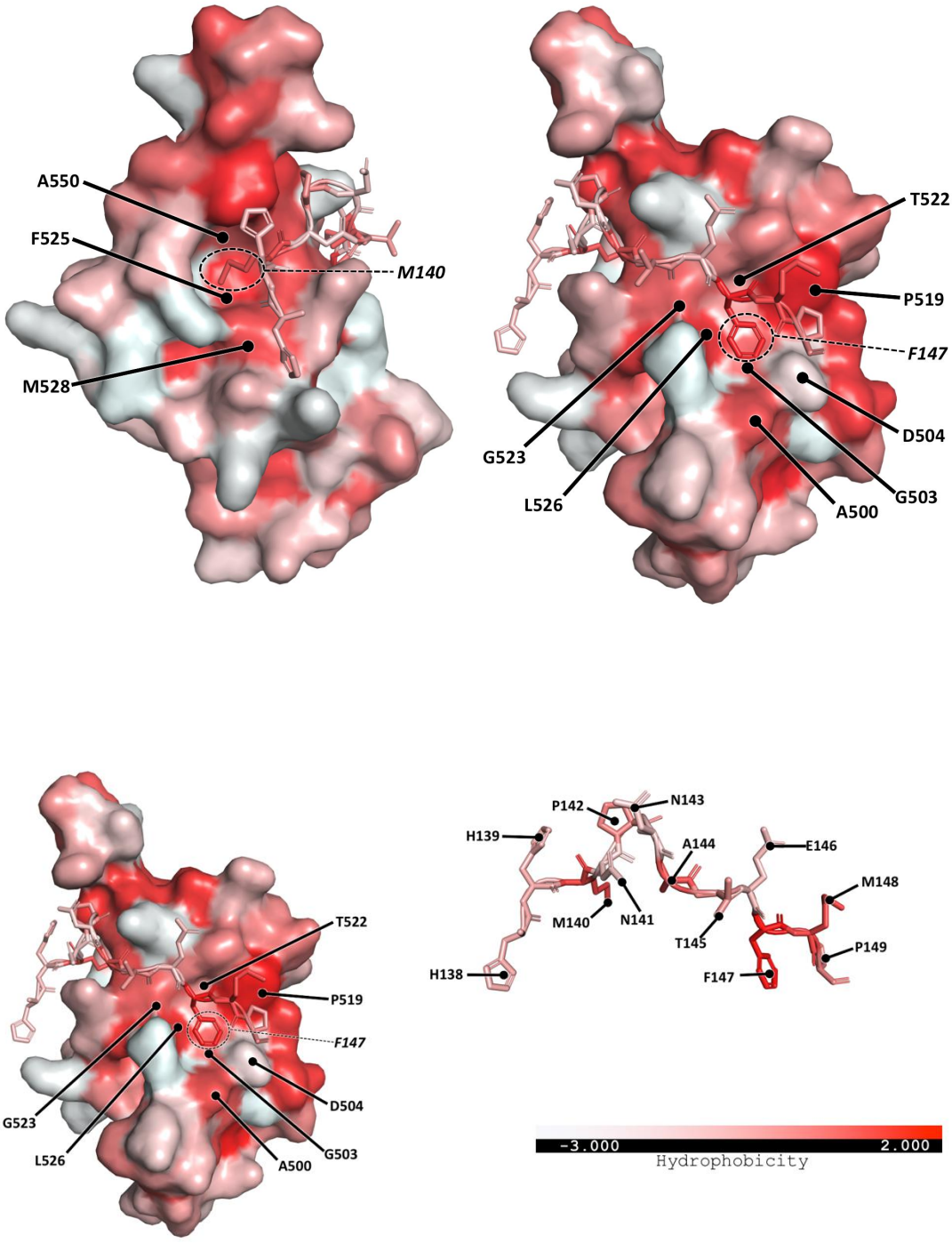
PABP1(J) Dimensions



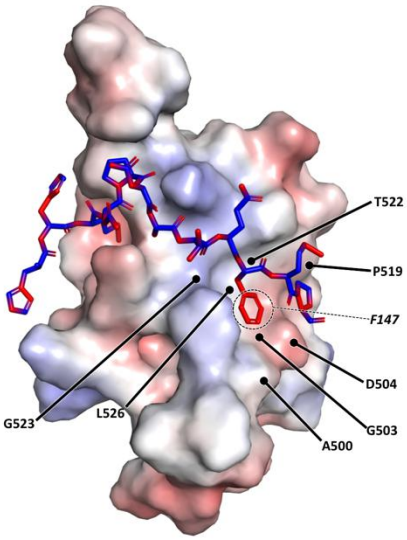
PABP1(J) **α**-helices vectors

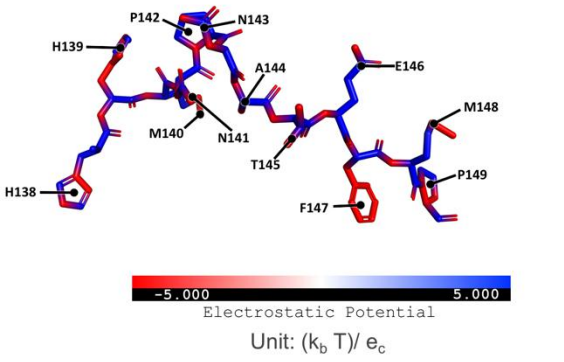


PABP1(J)/eIFE4(iv) hydrophobicity map



PABP1(J)/eIFE4(iv) electrostatic map





Unit: (k_b T)/ e_c

- k_b is Boltzmann's constant: $1.3806504 \times 10^{-23}$ J/K
- T is the temperature of calculation in K
- e_c is the charge of an electron: $1.60217646 \times 10^{-19}$ C

Chapter 10 - Bibliography

AKHOUNDI, M. et al. A Historical Overview of the Classification, Evolution, and Dispersion of *Leishmania* Parasites and Sandflies. **PLoS Neglected Tropical Diseases**, v. 10, n. 3, p. 1–40, 2016.

ALBRECHT, M.; LENGAUER, T. Survey on the PABC recognition motif PAM2. **Biochemical and Biophysical Research Communications**, v. 316, n. 1, p. 129–138, 2004.

ALEXANDROV, A.; DUTTA, K.; PASCAL, S. M. MBP fusion protein with a viral protease cleavage site: One-step cleavage/purification of insoluble proteins. **BioTechniques**, v. 30, n. 6, p. 1194–1198, 2001.

BAKER, N. A. et al. Electrostatics of nanosystems: Application to microtubules and the ribosome. **Proceedings of the National Academy of Sciences**, v. 98, n. 18, p. 10037–10041, 2001.

BARRETT, M. P.; CROFT, S. L. Management of trypanosomiasis and leishmaniasis. **British Medical Bulletin**, v. 104, n. 1, p. 175–196, 2012.

BATES, E. J.; KNUEPFER, E.; SMITH, D. F. Poly(A)-binding protein I of *Leishmania*: functional analysis and localisation in trypanosomatid parasites. **Nucleic acids research**, v. 28, n. 5, p. 1211–1220, 2000.

BAX, A.; IKURA, M. An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the α-carbon of the preceding residue in uniformly ^{15}N/^{13}C enriched proteins. **Journal of Biomolecular NMR**, v. 1, n. 1, p. 99–104, 1991.

CARDOSO, E. et al. Leishmaniasis: History, Evolution of Treatment and the Need for New Drugs. **Current Biotechnology**, v. 3, n. 4, p. 279–288, 2015.

CARRILLO, E. et al. Protein malnutrition impairs the immune response and influences the severity of infection in a hamster model of chronic visceral leishmaniasis. **PLoS ONE**, v. 9, n. 2, 2014.

CHAPPUIS, F. et al. Visceral leishmaniasis: What are the needs for diagnosis, treatment and control? **Nature Reviews Microbiology**, v. 5, n. 11, p. 873–882, 2007.

CLUBB, R. T.; THANABAL, V.; WAGNER, G. A Constant-Time Three-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue ^{1}HN, ^{15}N, and ^{13}C' Chemical Shifts in ^{15}N-^{13}C-Labeled Proteins. **Journal of Magnetic Resonance**, v. 97, p. 213–217, 1992.

COLLIN, S. et al. Conflict and Kala-Azar: Determinants of Adverse Outcomes of Kala-Azar among Patients in Southern Sudan. **Clinical Infectious Diseases**, v. 38, n. 5, p. 612–619, 2004.

COPELAND, N. K.; ARONSON, N. E. Leishmaniasis: Treatment updates and clinical practice guidelines review. **Current Opinion in Infectious Diseases**, v. 28, n. 5, p. 426–437, 2015.

CROFT, S. L.; SUNDAR, S.; FAIRLAMB, A. H. Drug Resistance in Leishmaniasis Drug Resistance in Leishmaniasis. **Clinical Microbiology Reviews**, v. 19, n. 1, p. 111–126, 2006.

DA COSTA LIMA, T. D. et al. Functional characterization of three *Leishmania* poly(A) binding protein homologues with distinct binding properties to RNA and protein partners. **Eukaryotic Cell**, v. 9, n. 10, p. 1484–1494, 2010.

DE MELO NETO, O. P. et al. The unique *Leishmania* EIF4E4 N-terminus is a target for multiple phosphorylation events and participates in critical interactions required for translation initiation. **RNA Biology**, v. 12, n. 11, p. 1209–1221, 2015.

DEVER, T. E.; GREEN, R. Phases of Translation in Eukaryotes. **Cold Spring Harbor Perspectives in Biology Perspect Biology**, v. 4, p. 1–16, 2012.

DHALIA, R. et al. Translation initiation in *Leishmania major*: Characterisation of multiple eIF4F subunit homologues. **Molecular and Biochemical Parasitology**, v. 140, n. 1, p. 23–41, 2005.

DIRO, E. et al. Clinical aspects of paediatric visceral leishmaniasis in North-west Ethiopia. **Tropical Medicine and International Health**, v. 20, n. 1, p. 8–16, 2015.

DOWLATI, Y. Cutaneous leishmaniasis: Clinical aspect. **Clinics in Dermatology**, v. 14, n. 5, p. 425–431, 1996.

EDFELDT, F. N. B.; FOLMER, R. H. A.; BREEZE, A. L. Fragment screening to predict druggability (ligandability) and lead discovery success. **Drug Discovery Today**, v. 16, n. 7–8, p. 284–287, 2011.

EISENBERG, D. et al. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. **Journal of Molecular Biology**, v. 179, n. 1, p. 125–142, 1984.

FIRCZUK, H. et al. An in vivo control map for the eukaryotic mRNA translation machinery. **Molecular Systems Biology**, v. 9, n. 635, p. 1–13, 2013.

FREIRE, E. et al. The four trypanosomatid eIF4E homologues fall into two separate groups, with distinct features in primary sequence and biological properties. **Molecular and Biochemical Parasitology**, v. 176, n. 1, p. 25–36, 2011.

FREIRE, E. et al. *Trypanosoma brucei* translation initiation factor homolog EIF4E6 forms a tripartite cytosolic complex with EIF4G5 and a capping enzyme homolog. **Eukaryotic Cell**, v. 13, n. 7, p. 896–908, 2014. [1]

FREIRE, E. et al. eIF4F-like complexes formed by cap-binding homolog TbEIF4E5 with TbEIF4G1 or TbEIF4G2 are implicated in post-transcriptional regulation in *Trypanosoma brucei*. **RNA**, v. 20, n. 8, p. 1272–1286, 2014. [2]

FREIRE, E. et al. The Role of Cytoplasmic mRNA Cap-Binding Protein Complexes in *Trypanosoma brucei* and Other Trypanosomatids. **Pathogens**, v. 6, n. 4, p. 55, 2017.

GOUTELLE, S. et al. The Hill equation: A review of its capabilities in pharmacological modelling. **Fundamental and Clinical Pharmacology**, v. 22, n. 6, p. 633–648, 2008.

GRZESIEK, S.; BAX, A. Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. **Journal of the American Chemical Society**, v. 114, n. 16, p. 6291–6293, 1992. [1]

GRZESIEK, S.; BAX, A. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. **Journal of Magnetic Resonance**, v. 96, n. 2, p. 432–440, 1992. [2]

HERWALDT, B. L. Leishmaniasis. **Lancet**, v. 354, n. 0140–6736 (Print), p. 1191–1199, 1999.

HINNEBUSCH, A. G.; LORSCH, J. R. The mechanism of eukaryotic translation initiation: New insights and challenges. **Cold Spring Harbor Perspectives in Biology**, v. 4, n. 10, p. 1–25, 2012.

JERABEK-WILLEMSEN, M. et al. MicroScale Thermophoresis: Interaction analysis and beyond. **Journal of Molecular Structure**, v. 1077, p. 101–113, 2014.

KAPP, L. D.; LORSCH, J. R. The Molecular Mechanics of Eukaryotic Translation. **Annual Review of Biochemistry**, v. 73, n. 1, p. 657–704, 2004.

KOZLOV, G.; GEHRING, K. Molecular basis of eRF3 recognition by the MLLE domain of poly(A)-binding protein. **PLoS ONE**, v. 5, n. 4, p. 3–8, 2010.

LASDA, E. L.; BLUMENTHAL, T. Trans-splicing. **Wiley Interdisciplinary Reviews: RNA**, v. 2, n. 3, p. 417–434, 2011.

LASKOWSKI, R. A.; SWINDELLS, M. B. LigPlot+ : Multiple Ligand À Protein Interaction Diagrams for Drug Discovery. **Journal of Chemical Information and Modeling**, v. 51, p. 2778–2786, 2011.

LIANG, X.; HARITAN, A.; ULIEL, S. trans and cis Splicing in Trypanosomatids: Mechanism, Factors, and Regulation. **Eukaryotic cell**, v. 2, n. 5, p. 830–840, 2003.

LINDOSO JAL, COSTA JML, QUEIROZ IT, G. H. Review of the current treatments for leishmaniases. **Research and Reports in Tropical Medicine**, v. 3, n. 12, p. 69–77, 2012.

LIU, B. et al. Fellowship of the rings: The replication of kinetoplast DNA. **Trends in Parasitology**, v. 21, n. 8, p. 363–369, 2005.

MALTEZOU, H. C. Drug resistance in visceral leishmaniasis. **Journal of Biomedicine and Biotechnology**, v. 2010, 2010.

MANSUETO, P. et al. Leishmaniasis in travellers: A literature review. **Travel Medicine and Infectious Disease**, v. 12, n. 6, p. 563–581, 2014.

MCNICOLL, F. et al. Distinct 3'-untranslated region elements regulate stage-specific mRNA accumulation and translation in *Leishmania*. **Journal of Biological Chemistry**, v. 280, n. 42, p. 35238–35246, 2005.

MOERKE, N. J. et al. Small-Molecule Inhibition of the Interaction between the Translation Initiation Factors eIF4E and eIF4G. **Cell**, v. 128, n. 2, p. 257–267, 2007.

MOURA, D. M. N. et al. Two related trypanosomatid eiF4G homologues have functional differences compatible with distinct roles during translation initiation. **RNA Biology**, v. 12, n. 3, p. 305–319, 2015.

NESVIZHSKII, A. I. et al. A statistical model for identifying proteins by tandem mass spectrometry. **Analytical Chemistry**, v. 75, n. 17, p. 4646–4658, 2003.

OSAWA, M. et al. Biological role of the two overlapping poly(A)-binding protein interacting motifs 2 (PAM2) of eukaryotic releasing factor eRF3 in mRNA decay. **RNA**, v. 18, n. 11, p. 1957–1967, 2012.

PACE, D. Leishmaniasis. **Journal of Infection**, v. 69, n. S1, p. S10–S18, 2014.

PAGE, R. et al. NMR screening and crystal quality of bacterially expressed prokaryotic and eukaryotic proteins in a structural genomics pipeline. **Proceedings of the National Academy of Sciences**, v. 102, n. 6, p. 1901–1905, 2005.

PÉREZ, H.; MALAVÉ, I.; ARREDONDO, B. The effects of protein malnutrition on the course of *Leishmania mexicana* infection in C57Bl/6 mice: nutrition and susceptibility to leishmaniasis. **Clinical and experimental immunology**, v. 38, n. 3, p. 453–460, 1979.

PODINOVSKAIA, M.; DESCOTEAUX, A. *Leishmania* and the macrophage: a multifaceted interaction. **Future Microbiology**, v. 10, n. 1, p. 111–129, 2015.

POVELONES, M. L. Beyond replication: division and segregation of mitochondrial DNA in kinetoplastids. **Molecular and biochemical parasitology**, v. 196, n. 1, p. 53–60, 2014.

PREEDY, H. C.; BAILEY, M. S. Leishmaniasis and its relevance to UK Armed Forces. **Journal of the Royal Naval Medical Service**, v. 100, n. 3, p. 238–243, 2014.

PURSE, B. V. et al. How will climate change pathways and mitigation options alter incidence of vector-borne diseases? A framework for leishmaniasis in South and Meso-America. **PLoS ONE**, v. 12, n. 10, p. 1–22, 2017.

READY, P. D. Leishmaniasis emergence in Europe. **Euro Surveillance**, v. 15, n. 10, p. 1–11, 2010.

ROYER, M.; CROWE, M. American cutaneous leishmaniasis: A cluster of 3 cases during military training in Panama. **Archives of Pathology and Laboratory Medicine**, v. 126, n. 4, p. 471–473, 2002.

SCHAFMEISTER, C et al. An All-Hydrocarbon Cross-Linking System for Enhancing the Helicity and Metabolic Stability of Peptides. **Journal of the American Chemical Society**, v. 122, n. 24, p. 5891–5892, 2000.

SIEVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, v. 7, n. 539, 2011.

STEVERDING, D. The history of leishmaniasis. **Parasites and Vectors**, v. 10, n. 1, p. 1–10, 2017.

SUNTER, J.; GULL, K. Shape, form, function and *Leishmania* pathogenicity: from textbook descriptions to biological understanding. **Open Biology**, v. 7, n. 9, p. 170165, 2017.

TAIT, S. et al. Local control of a disorder–order transition in 4E-BP1 underpins regulation of translation via eIF4E. **Proceedings of the National Academy of Sciences**, v. 107, n. 41, p. 17627–17632, 2010.

TIUMAN, T. S. et al. Recent advances in leishmaniasis treatment. **International Journal of Infectious Diseases**, v. 15, n. 8, 2011.

VALESKY, E. M. et al. Cutaneous leishmaniasis: Clinical report of two cases and review of the recent literature. **Journal of the German Society of Dermatology**, v. 5, n. 9, p. 770–772, 2007.

VARGAS-PARADA, L. Kinetoplastids and Their Networks of Interlocked DNA. **Nature Education**, v. 3, n. 9, p. 63, 2010.

VERDINE, G. L.; HILINSKI, G. J. Stapled peptides for intracellular drug targets. 1. ed. Elsevier Inc., 2012. v. 503

VON DER HAAR, T. A quantitative estimation of the global translational activity in logarithmically growing yeast cells. **BMC Systems Biology**, v. 2, p. 87–101, 2008.

WATSON, J. et al. **Molecular biology of the gene**. 7th. ed. [s.l.] Pearson - Cold Spring Harbor Laboratory Press, 2014.

WHO. Control of the Leishmaniases. World Health Organization, Geneva. **Technical Report Series**, v. 949, n. March, p. 22–26, 2010.

WILLIAMSON, M. P. Using chemical shift perturbation to characterise ligand binding. **Progress in Nuclear Magnetic Resonance Spectroscopy**, v. 73, p. 1–16, 2013.

XIE, J.; KOZLOV, G.; GEHRING, K. The “tale” of poly(A) binding protein: The MLLE domain and PAM2-containing proteins. **Biochimica et Biophysica Acta**, v. 1839, n. 11, p. 1062–1068, 2014.

YOFFE, Y. et al. Binding specificities and potential roles of isoforms of eukaryotic initiation factor 4E in *Leishmania*. **Eukaryotic Cell**, v. 5, n. 12, p. 1969–1979, 2006.

ZHANG, H. et al. A Cell-penetrating Helical Peptide as a Potential HIV-1 Inhibitor. **Journal of Molecular Biology**, v. 378, n. 3, p. 565–580, 2008.

ZINOVIEV, A.; SHAPIRA, M. Evolutionary conservation and diversification of the translation initiation apparatus in trypanosomatids. **Comparative and Functional Genomics**, v. 2012, 2012.