Investigations of the molecular interactions of two RNA-binding proteins, LARP4A and LARP4B

Gu, Yifei

Awarding institution:
King's College London

Investigations of the Molecular Interactions of two RNA-Binding Proteins, LARP4A and LARP4B

Yifei Gu

A thesis submitted in fulfilment of the requirements of King's College London,

for the award of Doctor of Philosophy

Randall Centre for Cell & Molecular Biophysics

School of Basic & Medical Biosciences

July 2022

Declaration

I hereby declare that the work presented in this thesis has not been submitted for any other degree or professional qualification, and that it is the result of my own independent work.

Yifei Gu

Full Name

07 July, 2022

Date

Acknowledgements

There are a great number of people I am indebted to throughout my journey in these past four years of my PhD journey. First, my primary supervisor Professor Sasi Conte, I would like to sincerely thank you for being an inspirational mentor and provided the time and an excellent environment to explore scientific research at such an advanced level, it was a privilege to be there. Working in the Conte Lab was truly an amazing experience, your patience and support for me proved to be an important life lesson even outside of the lab. The road to obtaining a PhD can be extremely tough and thankfully with your invaluable guidance and engagement with our work I can say I have truly enjoyed my time. I also owe my thanks to my second supervisor Dr Matthias Krause, and members in my thesis committee Professor Snezhana Oliferenko, Professor James McDonnell and Professor Agi Grigoriadis who dedicated their time and provided many useful suggestions that helped me throughout my work and made the experience truly unforgettable.

I would like to thank all the past and current lab members of the Conte group for making these four years so enjoyable, especially Dr. Isabel Cruz-Gallardo, which provided her expertise and help me get into grips with the first couple of months during my PhD, and always patiently answered all my questions and any doubts I had. Dr. Giancarlo Abis, who not only is an excellent scientist but also a great friend, and other members of the Conte lab, Dr. Neha Agrawal, Dr. Rita Puglisi, Dr. Gian De Nicola, Dr. Tam Bui, Dr. Charles Nichols, Dr. Leonie Kohlhammer, Dr. Jennifer Coleman, Federica Capraro, Sadie Hallett, Florence Mattingly-Peck, Qiongju Qiu, Adrien Le Guennec, James Jarvis, Harshaan Sahota, Jing Sun, Yassmin Elmasri, Edgar Leyva Garcia, Claudio Lopes Colaco, Piotr Kwiatkowski, Pierre Coleman, Javier Lizarrondo Diaz De Cerio and Austin Burroughs.

Being in the Randall Centre was extremely daunting at the start, luckily, I have met many friends and future PhDs that are in the same situation as me, I would like to thank everyone in the Randall for being so friendly and making the overall experience that much better. The Randall is truly like a big family, and I am proud to be working alongside all of you.

I would like to also thank my parents, for unconditionally trusting and believing me, and supporting me the all the way. It truly means a lot to me, and I am grateful to both of you. Finally, my deepest gratitude goes to Reese, I thank her for the physical and emotional support she has provided me throughout these years, and for believing in me, always.

Abstract

RNA-binding proteins (RBPs) bind to RNA targets and control a myriad of cellular functions such as transcription, mRNA localisation, stability, polyadenylation, splicing and decay. Mis-regulation of these processes are known to be prevalent in diseases such as cancer or neurodegeneration.

This project aims to characterise two RNA-binding proteins, namely LARP4A and LARP4B from the La-related RNA-binding protein (LARP) family. LARPs are a group of evolutionary conserved RBPs that regulate RNA metabolism identified by the presence of a conserved region termed the La-module which modulates the interaction to RNA. LARP4A and LARP4B have high amino acid similarity, share the same protein partners, but have different target RNA. Elucidation of the exact binding mechanism and their different RNA preference are two key areas of focus in my research.

LARP4A is known to interact with polyA sequences located in the 3’ ends of mRNAs. LARP4B targets AU-rich stretches, typically found in the 3' untranslated region (UTR) of many mRNAs. Both proteins are known to control key cellular processes such as mRNA stabilisation and translation, they localise to stress granules and have been recently linked to cancer. In addition to their interaction with RNA they are both known to interact with PolyA binding protein (PABPC1) which is a well characterised protein that is involved in the regulation of RNA metabolism, polyA lengthening and the termination of translation. The interaction to PABPC1 involves a short peptide sequence in the N-terminal regions of LARP4A and LARP4B, termed PAM2w. Although not well characterised, evidence of a second interacting region named PABP-binding motif (PBM) also exists downstream of the La-module.

In my project, a range of deletion and point mutants of LARP4A and LARP4B proteins were designed and generated from recombinantly proteins expressed in *Escherichia coli*

to be tested for their properties. Biochemical assays such as Electrophoretic Mobility Shift Assays (EMSA) and Microscale Thermophoresis (MST) were employed to characterize the binding affinity and specificity between different protein constructs and RNA. Furthermore, circular dichroism (CD) was used to analyse any changes of secondary structure from mutant to mutant.

Previous work from our lab shows that LARP4A recognises polyA RNA via a binding mechanism on the N-terminus mediated by disordered regions involving a protein-binding motif PAM2w, rather than the usual RNA-binding La-module. Furthermore, I tested binding to oligoA RNA using PAM2w motif alone and showed that a contiguous region at the N-terminus of LARP4A is required for oligoA RNA binding, and regions only encompassing the PAM2w motif alone are not enough to bind RNA. For LARP4B, to characterise its RNA recognition I investigated the N-terminal domain (NTD) that comprises the La-module and an N-terminal region, by systematically trimming the N-terminal region and have obtained deletion mutants, results show a gradual decrease in binding affinity towards AU-rich RNA the more N-terminal residues removed. Within the N-terminal region there also seems to be an unknown flexible RNA-binding region which is responsible to bind to RNA. The La-module of LARP4B shows increased binding affinity to AU-rich RNA when compared to the extremely weak binding of LARP4A to oligoA RNA. Within the La-module there are 6 conserved residues known to be vital for RNA-recognition. Using sequence alignment, we have generated point mutants in LARP4B that revealed little to no difference in RNA binding using our assays, suggesting that these amino acids play a minor role in RNA-recognition in LARP4B.

Table of contents

List of figures

Figure *Page*

List of tables

Table *Page*

Chapter 1. Introduction

1.1 Overview

The central dogma of molecular biology conjured by Francis Crick (Crick, 1970) states that DNA gives rise to mRNA via transcription, which in turn gives rise to protein via translation. This indicates a straightforward pathway, but while this holds true, decades of recent research have shown that there are many layers of regulation that goes on within transcription and translation (Sonenberg and Hinnebusch, 2009; Casamassimi and Ciccodicola, 2019), and other key players such as non-coding RNA (ncRNA) (Palazzo and Lee, 2015) have emerged for their key biological roles. Our lab is interested in RNA-binding proteins that control the fate and function of several mRNAs that may be involved in critical pathways of diseases (Gebauer *et al.*, 2020). A major goal in RNA biology is the characterisation of proteins that play a role in RNA regulation. I present novel experimental data regarding two known RNA-binding proteins, La-related protein 4A (LARP4A) and La-related protein 4B (LARP4B), and their RNA-binding partners.

1.2 RNA-Binding Proteins (RBPs)

RNA-binding Proteins are important players in regulating gene expression, translation and can interact with precursor mRNAs (pre-mRNAs) which are synthesized by RNA polymerase II (RNAPII) to form complexes called ribonucleoprotein complexes (RNPs) (Hentze *et al.*, 2018). RBPs can change the fate and function of the target RNA, they are involved in all steps of RNA biogenesis including post-transcriptional control, pre-mRNA splicing, polyadenylation, mRNA translation which can all affect the stability and localization of the target RNA (Re *et al.*, 2014). Improper RNA maintenance equals

improper protein expression and this will lead to diseases such as cancer and neurodegenerative disorders, hundreds of RBPs are found to be mutated and their levels are altered during these diseases (Lukong *et al.*, 2008; Zhou *et al.*, 2014; Wolozin and Apicco, 2015). Despite being of such importance, very few RBPs have been studied systematically (Glisovic *et al.*, 2008), but recently the importance of RBPs has been acknowledged and more and more of them are getting dissected and studied by researchers.

RBPs can be localized in both the nucleus and the cytoplasm, and most interact specifically to their targets. They usually contain structural motifs and RNA binding domains, the most common one being RNA Recognition Motif (RRM), however there are also other motifs such as the hnRNP K-Homology domain (KH), the double-stranded RNA-binding domain (dsRBD), Zinc Fingers (ZnFs), S1 domains, the Pumilio/FBF (PUF or Pum-HD) domain and the Piwi/Argonaute/Zwille (PAZ) domain (Lunde, Moore and Varani, 2007) that bind and interact with their specific targets.

These protein-RNA interactions can be quite specific and can be characterised via methods such as in solution nuclear magnetic resonance (NMR) and crystallography. It is revealed that residues in these RNA-binding domains interact with bases and backbones of their RNA targets (Černý and Hobza, 2007) of non-covalent nature, such as hydrogen bonding or stacking interactions, mainly mediated by Van der Waals forces between aromatic residues and the purine/pyrimidine rings of the RNA (Černý and Hobza, 2007). Often, the residues that interact with RNA are highly conserved throughout evolution (Kenan, Query and Keene, 1991; Alfano *et al.*, 2004), which further underscores their importance for RNA-binding.

1.3 RNA recognition motif

The RNA recognition motif (RRM) is one of the most common RNA-binding domains mainly found in eukaryotes. It was first identified in the late 1980s when it was demonstrated that mRNA precursors (pre-mRNA) and heterogeneous nuclear RNAs (hnRNAs) are always found in complex with proteins (Dreyfuss, Swanson and Piñol-Roma, 1988). The RRM is found in all kingdoms of life and is often found as multiple copies within a single protein and/or in tandem with other RBDs. The canonical RRM (see Figure 1) is around 80-90 amino acids long, and is arranged into a four-stranded, anti-parallel β-sheet that is flanked by two α-helices (Nagai, 1996; Cléry, Blatter and F. H. T. Allain, 2008). This arrangement is usually seen as a α/β 'sandwich' orientation, and usually the RNA-recognition often occurs on the β-sheet surface, due to the conserved ribonucleoprotein motifs 1 and 2 (RNP-1 and RNP-2) (Maris, Dominguez and Allain, 2005). To form contact with the target RNA, one or up to four of the β-strands may be used for interaction, and additional elements such as exposed loops, additional β-strands, or often extensions from the N- and/or C-termini can aid the interaction to RNA (Cléry, Blatter and F. H.-T. Allain, 2008; Daubner, Cléry and Allain, 2013). The RRM is an extremely diverse motif in both structure and function, and further studies are required to unveil and understand its structural and functional role in different proteins.

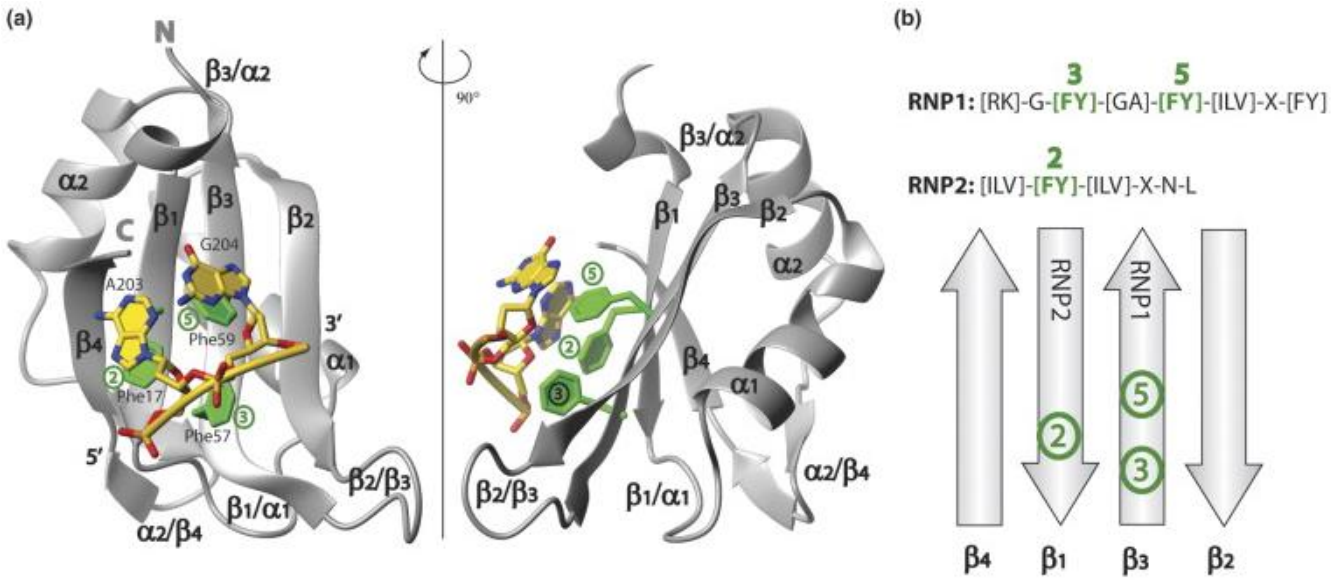


Figure 1. RRM β-sheet interaction with RNA. *(a) Structure of hnRNP A1 RRM2 in complex with single stranded telomeric DNA as a model of single stranded nucleic acid binding. (b) Scheme of the four-stranded β-sheet with the place of main conserved RNP1 and RNP2 aromatic residues indicated in green. RNP1 and RNP2 consensus sequences of RRMs are shown (X is for any amino acid). (Figure from Cléry et al. 2008)*

1.4 La-related Proteins (LARPs)

The name La comes from the protein which was first discovered as an autoantigen in immune disorders (Alspaugh, Talal and Tan, 1976), now known to be ubiquitous in eukaryotic cells. La is known to bind a common terminal motif UUU-3'-OH of transcripts transcribed by RNA polymerase III, shown to act as an RNA chaperone and has protective properties such as the prevention of RNA misfolding (Maraia *et al.*, 2017). The La-related Protein superfamily consists of LARP3 (La), LARP1, LARP4A, LARP4B, LARP6 and LARP7 (see Figure 2). It is a family of RBPs that has many different non-redundant cellular functions, they all bind to different RNA targets (Bousquet-Antonelli and Deragon, 2009) and this gives them a range of functions including roles in mRNA translation and mRNA stability.

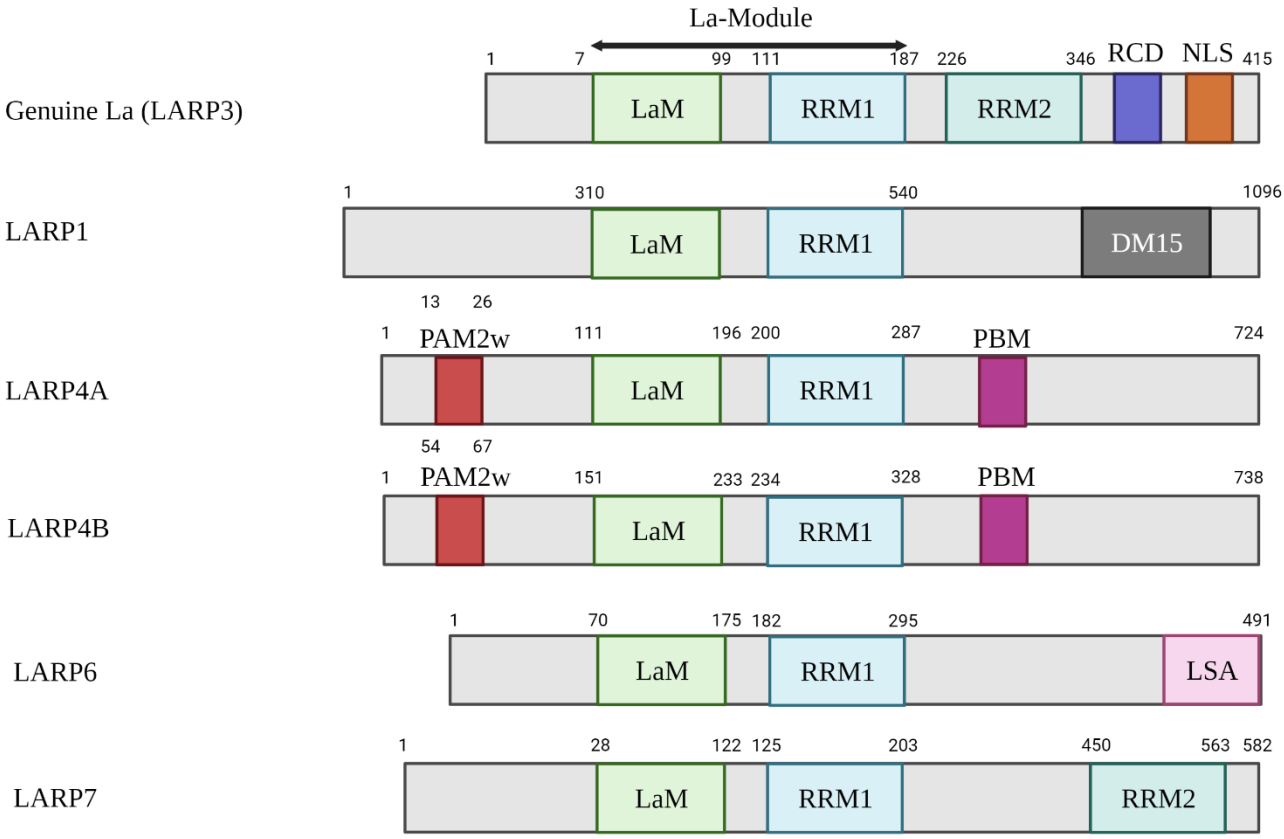


Figure 2. Schematic representation for the family of LARP proteins. *All LARPs share a bipartite La-module, composed of a La Motif (LaM) and the RNA Recognition Motif (RRM). In addition, the LARPs have additional domains that distinguish them between their subfamilies and may or may not contribute to different functions. RCD: RNA chaperone domain, NLS: nuclear localisation signal, DM15: DM15-repeat containing region, PAM2w: an atypical Poly(A) Binding Protein Interaction Motif-2, PBM: Poly(A) Binding Protein Interaction Protein Motif, LSA: La and S1 associated motif. (Image created using BioRender, and the domain organisation is deduced from literature and the database from Uniprot).*

LARPs are conserved throughout the eukaryote evolution, and the common motif between all the LARPs is the La-module, consisting of the conserved signature La motif (LaM) and an RNA Recognition Motif (RRM) first discovered in La (Alfano *et al.*, 2004; Maraia *et al.*, 2017). The La-module has a high sequence conservation in the RNA binding domain (see Figure 3), and it is the RNA binding platform. Some LARPs, for example La and LARP7, have additional RRMs that enable them to have extra specificity towards their target. The LaM and RRMs are individual structural domains, the LaM has around 70 amino acids and is a modified winged-helix motif that contains three α-helices that allow the formation of a hydrophobic cavity with RNA binding capabilities (Alfano *et al.*, 2004; Dong *et al.*, 2004). The RRM, around 80 amino acids, has slight differences between each LARP, for example LARP7 has three β-strands whereas La has four

(Maraia *et al.*, 2017), which could potentially be a reason for target specificity: the LaM and RRM work together to interact with their RNA targets.

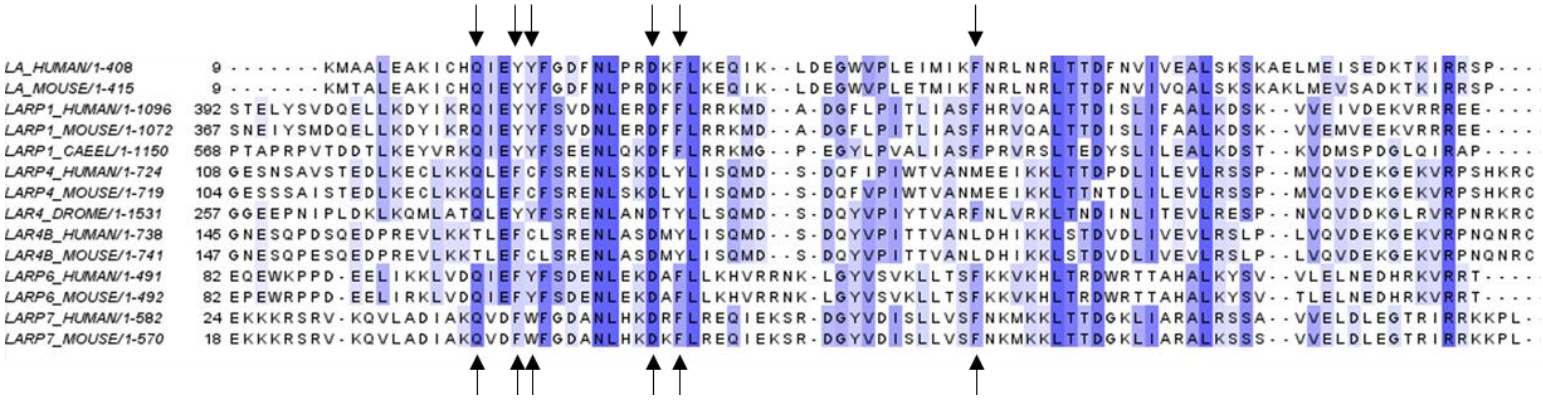


Figure 3. Sequence alignment of the LaM and their N-terminal regions in La-related proteins, *for each family, the sequence of the human LaM was aligned with proteins of different species using Clustal Omega in Uniprot portal (http://www.uniprot.org/align/). The alignments were edited and analysed with Jalview software. Residues were coloured in blue according to the extent of conservation. The arrows indicate the conserved residues.*

The LaM is originally discovered in La, which allowed the formation of a hydrophobic cavity that provides La, LARP7 and LARP6 with the RNA-binding capabilities (Alfano *et al.*, 2004; Martino *et al.*, 2015a; Uchikawa *et al.*, 2015) (see Figure 4). In this cavity, specific and non-specific contacts are formed between La and UUU-3'-OH RNA ligand and in particular mediated by 6 key conserved amino acid residues present in the La motif: Q20, Y23, Y24, D33, F35 and F55. These residues are extremely conserved, as shown in Figure 3, a study using the LaM of 234 proteins from 78 species belonging to the different LARP1, 4, 6, 7, and genuine La families show that position 20 (hLa numbering) was occupied by Q in 96.6%, positions 23 and 24 by F/Y/W in 97.9 and 99.9% respectively, position 33 by D in 99.7%, position 35 by F/Y in 96.2% and position 55 by F in 97.6% (Merret *et al.*, 2013). Using biophysical and mutagenesis techniques, this hydrophobic pocket was shown to be important in LARP7, having an importance in interaction to its target RNA which is the 7SK RNA UUU-3'OH (Uchikawa *et al.*, 2015), and in LARP6 (Martino *et al.*, 2015a), where the LaM pocket is vital for recognition with collagen 5'UTR stem-loop RNA. However, they are not absolutely conserved for LARP4s and LARP6 as shown in Figure 3, due to the fact that there is an evolutionary reorganization

in these two families (gene duplication event that happened very early in the vertebrate lineage), which also led to the acquisition of a PAM2 motif that is not present in the other LARPs (Merret et al., 2013).

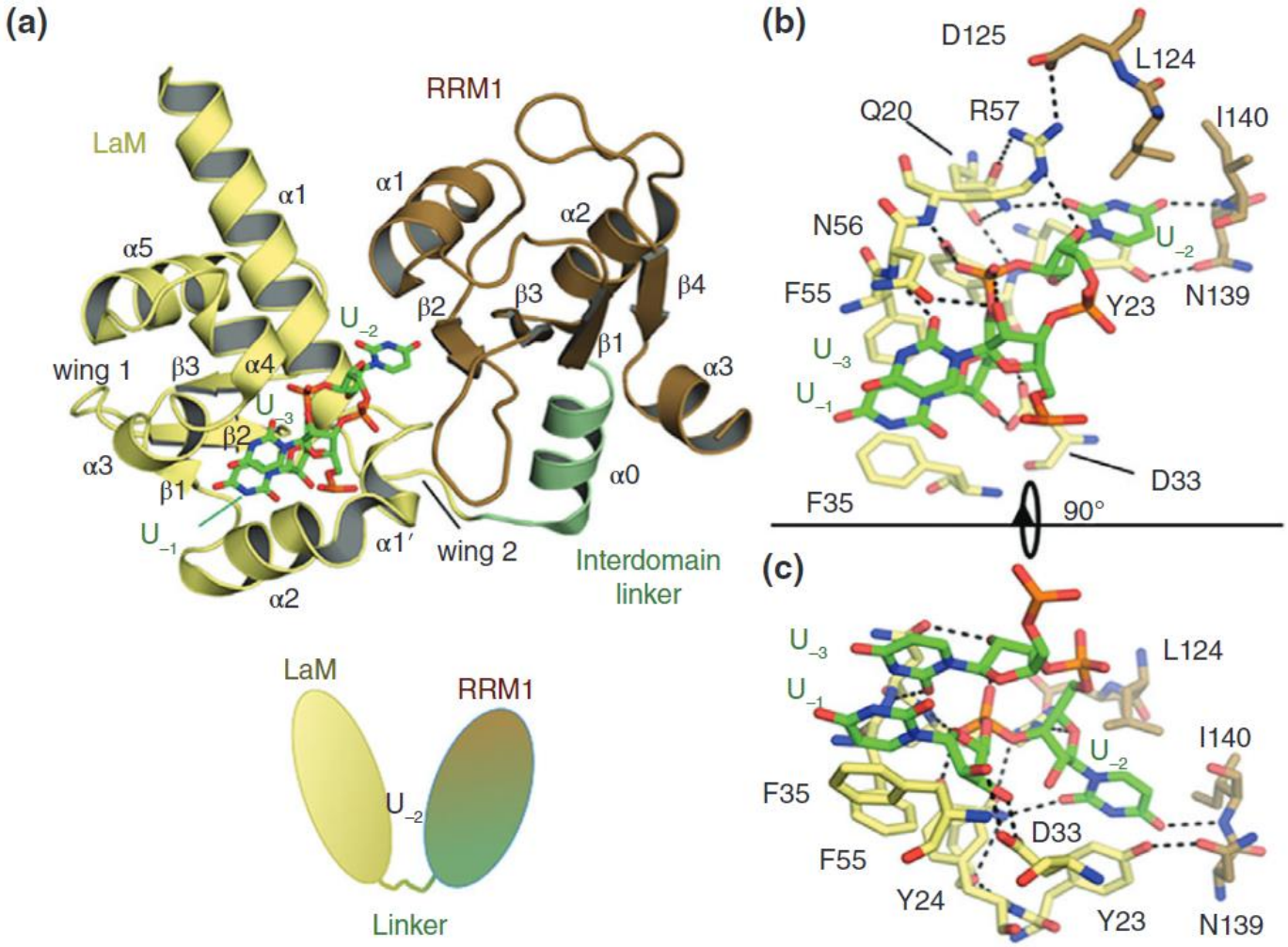


Figure 4. The La-module of human La protein in complex with UUU 3'OH RNA. *a) The La-module of La (PDB 1YTY) is composed by the LaM (yellow), an interdomain linker (pale green) and the RRM1 (brown). Cartoon under the structure shows schematic V-shaped model with U-2 in the cleft between the LaM and RRM1. b) Close-up view of La-RNA interaction showing the three 3' terminal bases. Selected side chains are shown as sticks; dashed lines indicate hydrogen bonds. c) Close-up view of the interaction, rotated by 90° (Figure from Maraia et al. 2017)*

1.5 LARP4 family

LARP4A and LARP4B are both cytoplasmic RBPs that have 40% amino acid identity and 74% in their La-modules (Maraia *et al.*, 2017) and have strong links to cancer (Stavraka and Blagden, 2015). They are the most divergent of the LARPs as they are

known to lack two aromatic amino acids in their LaM which is crucial for terminal oligo uridine binding, and even lack an invariant aspartic acid residue (D33 in La), which was shown to specifically make contact with the 2nd last U, vital for other LARPs to interact with the target RNA (Teplova *et al.*, 2006). LARP4A knockdown in cancer cells is found to promote cell migration (Bai *et al.*, 2011) and over-expression reduces cell elongation and increases cell circularity (Seetharaman *et al.*, 2016). Meanwhile LARP4B is found to be a tumour suppressor in human glioma cells, overexpression of LARP4B in glioma cell lines strongly inhibited proliferation by inducing mitotic arrest and apoptosis (Koso *et al.*, 2016). They both have two Poly A binding protein (PABP) binding modules (PolyA binding protein interaction motif-2 – PAM2w and PolyA binding protein interaction protein motif - PBM) that flank the La-module, see Figure 5. PABP has roles in regulation of polyadenylation, translation initiation and the protection of poly(A) tails of cellular mRNAs from nuclease degradation (Kühn and Wahle, 2004). LARP4s also interact with the Receptor for activated C kinase 1 (RACK1), which is a 40S ribosome- and mRNA-associated protein that participates in eukaryotic translation and ribosome quality control (Yang *et al.*, 2011).

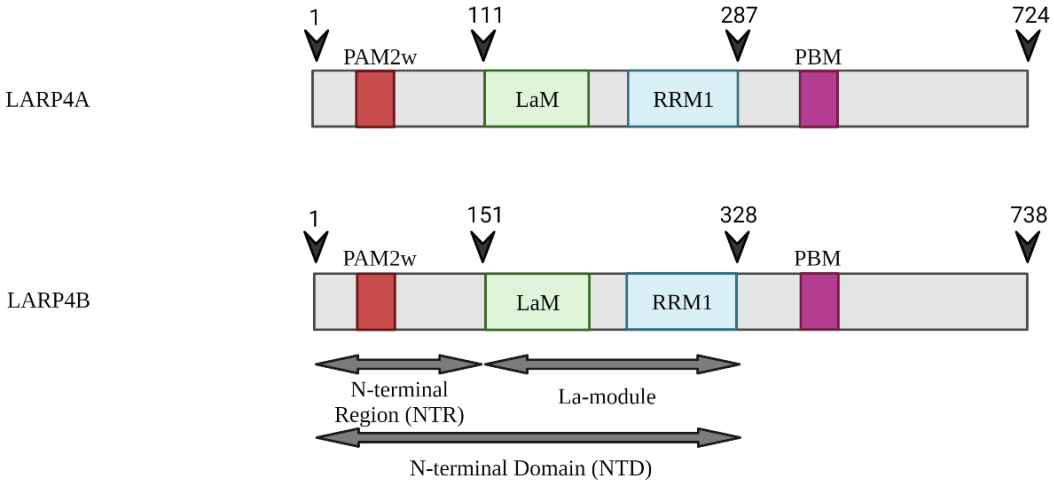


Figure 5. LARP4A and LARP4B domain organisation. *In LARP4A, the NTD spans residues 1-287, and the NTR residues 1-111. LARP4A La-module, encompassing amino acids 111-287, is composed of two domains: the LaM (111-196) and the RRM1 (200-287). In LARP4B, the NTD spans residues 1-328, and the NTR residues 1-151. LARP4B La-module, encompassing amino acids 151-328, is composed of two domains: the LaM (151-233) and the RRM1 (234-328).*

Both proteins have roles in protein synthesis regulation and mRNA stabilization and share the same binding partners (PABP and RACK1), however their target RNA is different. LARP4A is known to bind to and recognizes polyA sequences present in mRNAs (Yang *et al.*, 2011), whereas LARP4B recognizes AU-rich motifs in mRNAs (Küspert *et al.*, 2015). What makes LARP4A and LARP4B select different targets even though they are very similar in primary structure remain an open question. It is hypothesized (and shown by preliminary experiments) that the N-terminal regions of LARP4A have great effect towards RNA binding and selectivity, whether this holds true for LARP4B is still to be tested.

The interaction of LARP4A and LARP4B with Poly A binding protein (PABP) is crucial for the function of these proteins (Maraia *et al.*, 2017; Mattijssen *et al.*, 2017). Cytoplasmic PABPs (PABPC) of which the best studied is PABPC1 contain four RRM domains focusing on oligoA interaction (Goss and Kleiman, 2013). The C-terminal part of PABPC1 contains an unstructured, proline-rich region of low conservation (Kühn and Pieler, 1996; Patel and Bag, 2006) and a highly conserved domain termed MLLE (Mademoiselle), characterized by its invariant KITG (Lysine, Isoleucine, Threonine, Glycine) MLLE signature motif (Kozlov *et al.*, 2010). Many proteins interacting with PABPC1-MLLE have been identified, including several translation factors, such as eIF4G, Paip1, Paip2 and eRF3. These proteins contact PABP via a common motif of 12–15 amino acid residues, known as PABP interacting motif 2 (PAM2).

The PAM2 motif in LARP4s do not contain the typical sequence (LXXXAXX**F**XP) found in other PAM-2 containing proteins, instead of the variant phenylalanine – underlined, LARP4s possess a tryptophan (Merret *et al.*, 2013). The PAM2 motif in LARP4s is therefore known as a PAM2w motif, seems to be conserved throughout LARP4s. It is shown that despite it being a tryptophan residue, it still functions to bind PABP via the MLLE domain which is the main protein-interacting domain (Yang *et al.*,

2011; Merret *et al.*, 2013), hence the PAM2w motif is crucial for the function of LARP4s towards PABP binding (Yang *et al.*, 2011). A crystal structure of the LARP4B PAM2w/MLLE complex is shown in Figure 6, within the crystal structure, a single LARP4B-PAM2w complex contacts a neighbouring, symmetry-related MLLE entity, providing an extended interface that creates a deep pocket into which the peptide binds, running over or in between the surfaces of both symmetry-related MLLE molecules (Grimm et al., 2020).

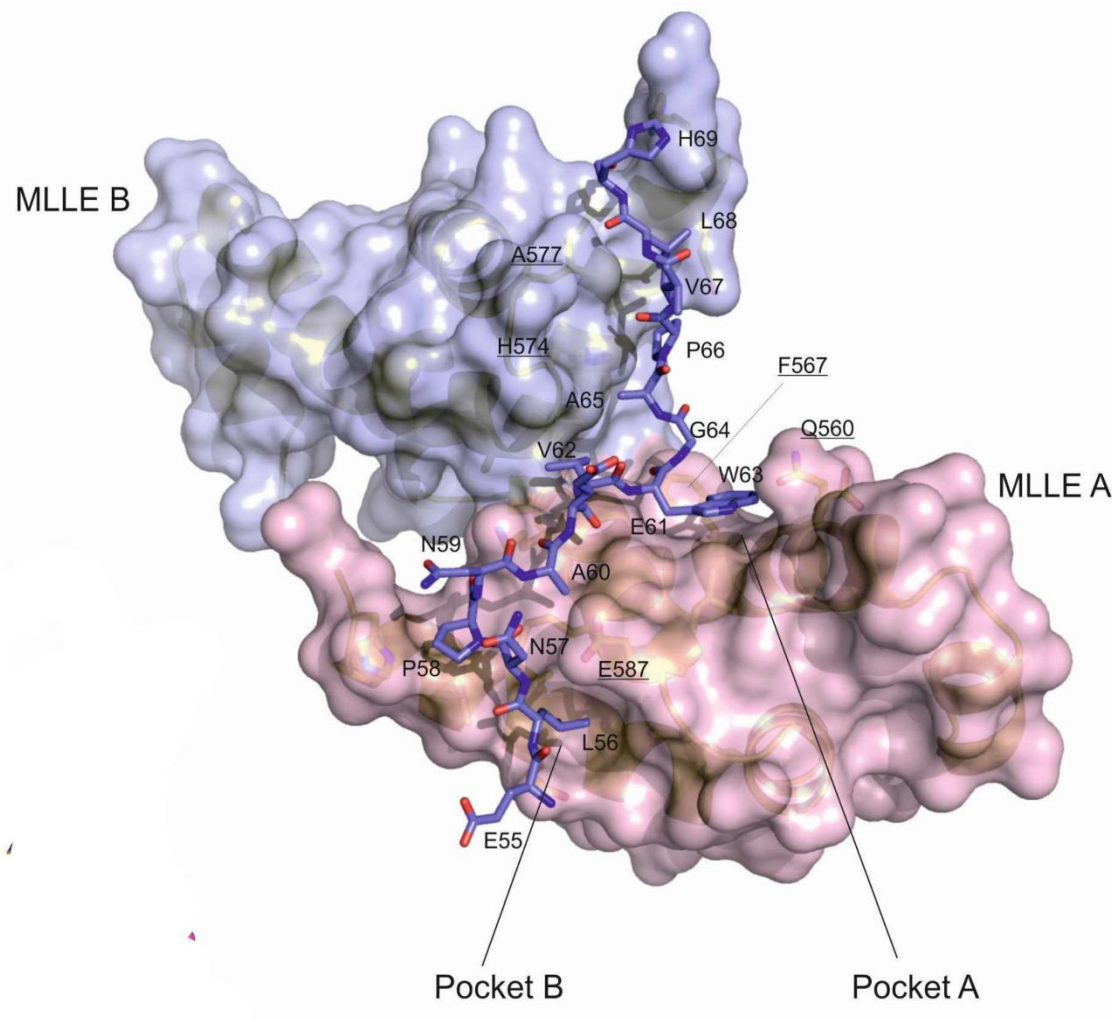


Figure 6. Surface representation of MLLE molecule A (pink) and B (blue) with the bound LARP4B–PAM2w peptide, *shown as a stick model. MLLE residue names are underlined and the residue from LARP4B are not (Figure from Grimm et al. 2020)*

1.6 LARP4A

LARP4A, or previously known as LARP4, is one member of the LARP4 family, they are predominantly localised in the cytoplasm during steady state, where they associate with mRNAs and participate in translation regulation. LARP4A has been reported to bind to polyA RNA, interact with PABP MLLE domain via the PAM2w motif, promote mRNA stability through binding to polyA tails and protecting them from deadenylase activity (Yang *et al.*, 2011; Mattijssen *et al.*, 2021). LARP4A plays a role in poly(A) lengthening of mRNAs and associates with poly(A) sequences of at least 15 nucleotides (Yang *et al.*, 2011; Mattijssen *et al.*, 2017). Yeast two-hybrid analysis and reciprocal immunoprecipitations (IPs) from HeLa cells revealed that LARP4A interacts with RACK1, a 40S ribosome and mRNA-associated protein. LARP4 cosediments with 40S ribosome subunits and polyribosomes, and its knockdown decreases translation, suggesting that LARP4A is a functional component of a fraction of translating mRNPs. (Yang *et al.*, 2011). Over 140 mutations in LARP4A have been reported in multiple cancers. LARP4A appears to regulate cancer cell migration and invasion, a study using prostate cancer cells exhibited an elongated phenotype consistent with increased migration upon LARP4A knockdown (Bai *et al.*, 2011). LARP4A knockdown also increased cell migration and invasion of other cancer cells such as breast cancer, whereas overexpression decreased elongation (Seetharaman et al., 2016).

LARP4A contains a La-module which folds into two domains that lack a rigid orientation, the structure is deduced via NMR (Figure 7) and it is shown that the LaM of LARP4A has the same winged-helix domain that was previously described in La, and the RRM1 of LARP4A adopts a canonical RRM-like fold that consists of four antiparallel β-strands which pack together, forming a centrally located platform which is flanked on one side by two α-helices (Cléry, Blatter and F. H. T. Allain, 2008; Cruz-Gallardo *et al.*, 2019). Figure 7 shows the structure of the La-module of LARP4A determined using standard

heteronuclear multidimensional NMR techniques (Cruz-Gallardo *et al.*, 2019). This shows that the La-module of LARP4A contains two distinct domains that are analogous to other LARPs, with a linker connecting the two domains.

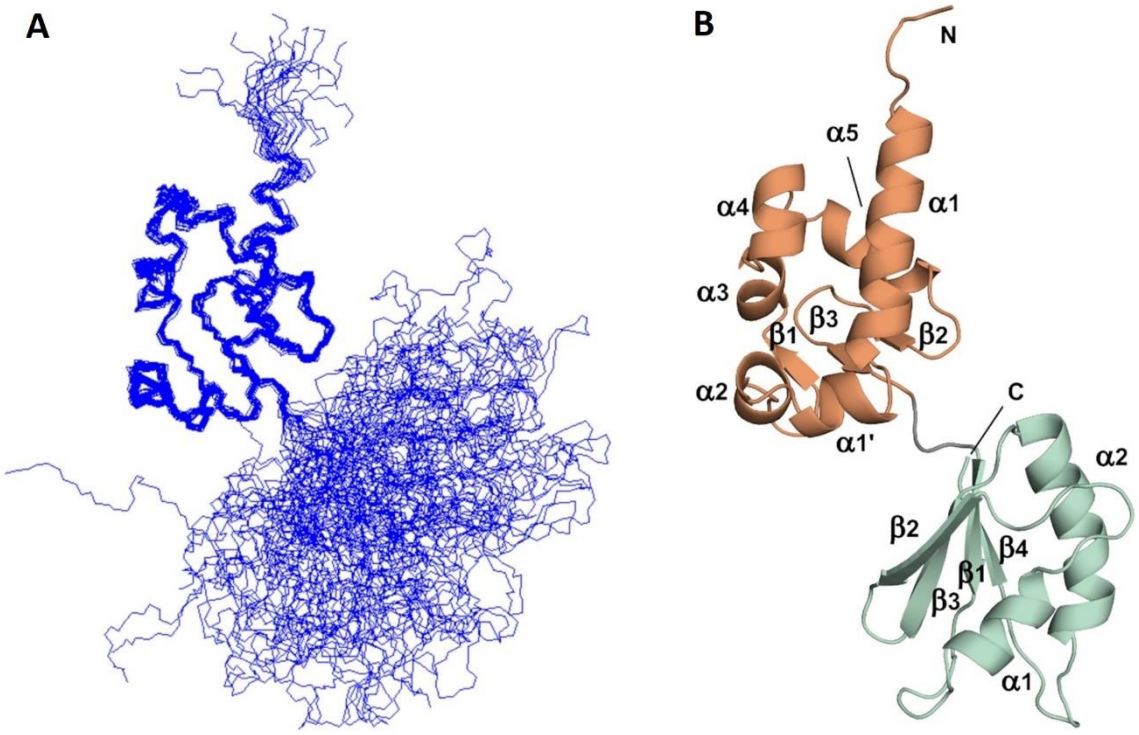


Figure 7. The structure of LARP4A La-module by NMR. *A. The entire LARP4A La-module structure solved using standard heteronuclear multidimensional NMR techniques, this shows the non-fixed relative orientation of the two domains in solution. B. A cartoon representation of LARP4A La-module, where the LaM is shown in orange, and the RRM shown in green, the α-helices and β-sheets are labelled. (Figure from Cruz-Gallardo et al., 2019)*

The linker in LARP4A is another divergent trait of LARPs that was previously unveiled (Martino *et al.*, 2015b), the linker is shown to be only a three amino acid stretch (HKR). This is specific and invariant for LARP4A and invertebrate LARP4 proteins (Merret *et al.*, 2013) but not present in any other LARP, nor in LARP4B members where it diverges into Q(N/S)R. The linker also presents itself as an extended conformation, making the whole protein seen elongated rather than a V-shaped configuration when compared to La (see Figure 4a), this is also a unique characteristic seen in LARP4A that could potentially give rise to target specificity.

Surprisingly, the La-module of LARP4A appears to lack high RNA-binding capability in isolation (Cruz-Gallardo et al., 2019), whether this is linked to its different extended tandem LaM/RRM configuration, its dynamic ensemble, sequence composition, electrostatic properties or a combination of all, remains to be established. LARP4A was thus investigated using electrophoretic mobility shift assays (EMSA) previously in the Conte lab (Cruz-Gallardo *et al.*, 2019). Using a range of truncation mutants spanning the N-terminus, it was realised that the N-terminal region (NTR) contains the principal determinants for oligoA recognition. Further investigations of the N-terminus have highlighted an important role for the variant PAM2w motif that serves as the binding platform for the PABPC1-MLLE domain but unexpectedly revealed to have also roles in poly(A) recognition, thereby sitting at the crossroads of protein–protein and protein–RNA interactions (Yang *et al.*, 2011; Cruz-Gallardo *et al.*, 2019). Further EMSA experiments using the PAM2w double mutant L15AW22A which is known to abrogate the protein-protein interaction (Yang *et al.*, 2011) has shown a 20-fold decrease in affinity in oligoA RNA-binding when compared to the wild type NTD (Cruz-Gallardo *et al.*, 2019). This experiment shows the importance of the PAM2w motif in LARP4A as an RNA binder when it was previously only known for protein-protein interactions. Therefore, the PAM2w motif not only is important in binding PABPC1, but also plays a key role in oligoA binding.

1.7 LARP4B

LARP4B, (or in some papers known as LARP5) is a paralog of LARP4A, it acts as a positive factor of translation and interacts with the translation machinery via similar protein-protein interactions as LARP4A (PABPC1, RACK1) (Angenstein *et al.*, 2002; Bayfield, Yang and Maraia, 2010; Schäffler *et al.*, 2010). LARP4B is shown to co-

sediment with polysomes in polysomal fractionation experiments. Immunofluorescence analysis of hemagglutinin antigen (HA)-tagged LARP4B revealed that the protein re-localizes to stress granules upon arsenite treatment (Schäffler *et al.*, 2010), and *in vivo* experiments that have been carried out have shown that LARP4B positively influences the translation of a spectrum of cellular mRNAs, which led to the conclusion of LARP4B acts as a general translation factor (Schäffler *et al.*, 2010).

In terms of disease, LARP4B is not as well studied when compared to LARP4A, however it has been found to act both as an oncogene and a tumour suppressor, depending on the type of cancer. In a study by Koso et al., it was demonstrated that over-expression of LARP4B suppresses tumour proliferation and promotes mitotic arrest and apoptosis in glioma cell lines partially dependent on the La-module suggesting that interaction with RNA and its other interacting proteins is important (Koso *et al.*, 2016). LARP4B over-expression causes mRNA levels of the mitotic regulatory gene CDKN1A and the Bcl2 family member BAX to be upregulated (Koso *et al.*, 2016). Interestingly, these factors were also described as LARP4B targets in HEK293 cells (Küspert *et al.*, 2015). This indicates that LARP4B could potentially control their stability and/or translation. As an oncogene, lentiviral suppression of LARP4B in an acute myeloid leukaemia mouse model reduced leukemic stem cells, attenuated their self-renewal capacity and caused them to undergo a cell cycle arrest thus prolonging the survival of acute myeloid leukemia mice (Zhang et al., 2015).

Much less is known about how LARP4B recognizes RNA, *In vivo* RNA targets of LARP4B have been identified using photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) analysis (Küspert *et al.*, 2015), a biochemical method used for identifying the binding sites of cellular RBPs. As LARP4B contains the La-module it was considered a bona fide mRNA-binding protein but previous studies had detected only interactions with PABPC1 and RACK1 (Angenstein

et al., 2002; Schäffler *et al.*, 2010). The PAR-CLIP experiment identified several mRNA targets that can be crosslinked to LARP4B, the distribution of the LARP4B-binding sites along the mRNAs revealed a preferential binding of LARP4B towards AU-rich sequences within the 3'UTR (Küspert *et al.*, 2015). In order to validate this result, *in vitro* EMSA binding assays were carried out with an RNA sequence of the CKB (Creatine Kinase B) 3′ UTR, this sequence was chosen because it contains an AU-rich sequence and is among the top hits from the PAR-CLIP (Küspert *et al.*, 2015). ^{32}P-labeled RNA was incubated with increasing amounts of recombinant LARP4B and analysed by native gel electrophoresis, LARP4B bound to the CKB sequence with a dissociation constant of ∼200 nM. Mutants that lack the U-stretch in CKB (CKB-del) showed no binding, therefore suggesting that binding was strictly dependent on the uridines within the CKB. Marginal binding was observed to a sequence found in the 3'UTR of a nontarget (TOP3B) that lacks the classic AU-rich element. (Küspert et al., 2015). Therefore, using a combination of mutagenesis studies and *in vitro* binding assays, it is clear that CKB is a target of LARP4B.

Altogether, LARP4B plays a central role in many cellular processes. However the mechanistic details of how LARP4B interacts with its partners, including RNA substrates, are poorly understood.

1.8 Aims of this study

Despite LARP4A and LARP4B having many vital roles in translation regulation and mRNA homeostasis, the mechanism of how LARP4s bind their corresponding protein and RNA targets is not well understood. LARP4A has been studied in the lab for many years and its mechanism to bind to RNA has just been unveiled, it is interesting to see how LARP4B, a paralog of LARP4A, interacts with its corresponding targets.

The main strategy of this study is by the generation of several deletion and point mutants in *Escherichia coli*, to understand the molecular basis of RNA target specificity in LARP4B, using LARP4A as a positive control, to understand the contribution of the N-terminal region and the La-module for RNA binding in LARP4B. The protein-protein interaction roles of LARP4B will be investigated by looking at the binding of LARP4B with its protein partner PABPC1, and the role of the variant PAM2w motif in LARP4B will be investigated using pull-down assays.

Chapter 2. Materials and Methods

2.1 Overview

Reagents for buffers were purchased through *Merck*, *Fisher Scientific* and *VWR International* unless stated differently otherwise. For the transfer of liquid volumes from 0.5 to 1000 µL, pipettes from either *Gilson* or *Eppendorf Research* were used, using pipette tips from *Starlab Group*. For volumes ranging from 1 to 25 mL, an automatic pipette boy from *SLS Laboratories* was used alongside serological pipettes purchased from *Alpha Laboratories Limited*. For measuring large volumes over 25 mL, either plastic or glass measuring cylinders from *Fisher Scientific* were used. One-time use universal tubes and falcon tubes were purchased from either *VWR International* or *Fisher Scientific*. All the primers were ordered directly from *Merck.*

2.2 Molecular Biology

2.2.1 Conventional cloning

Conventional cloning or traditional cloning is the 'classic' cloning method. It is an essential tool to generate the recombinant protein in a bacterial expression system and relies on 1. preparation of a vector to receive an insert DNA by digesting both with restriction endonucleases, 2. ligation of the insert to the vector by DNA ligase, 3. transformation of competent *E. coli* cells and 4. selection to identify positive clones that has successfully incorporated the DNA of interest, Figure 8) (Celie, Parret and Perrakis, 2016).

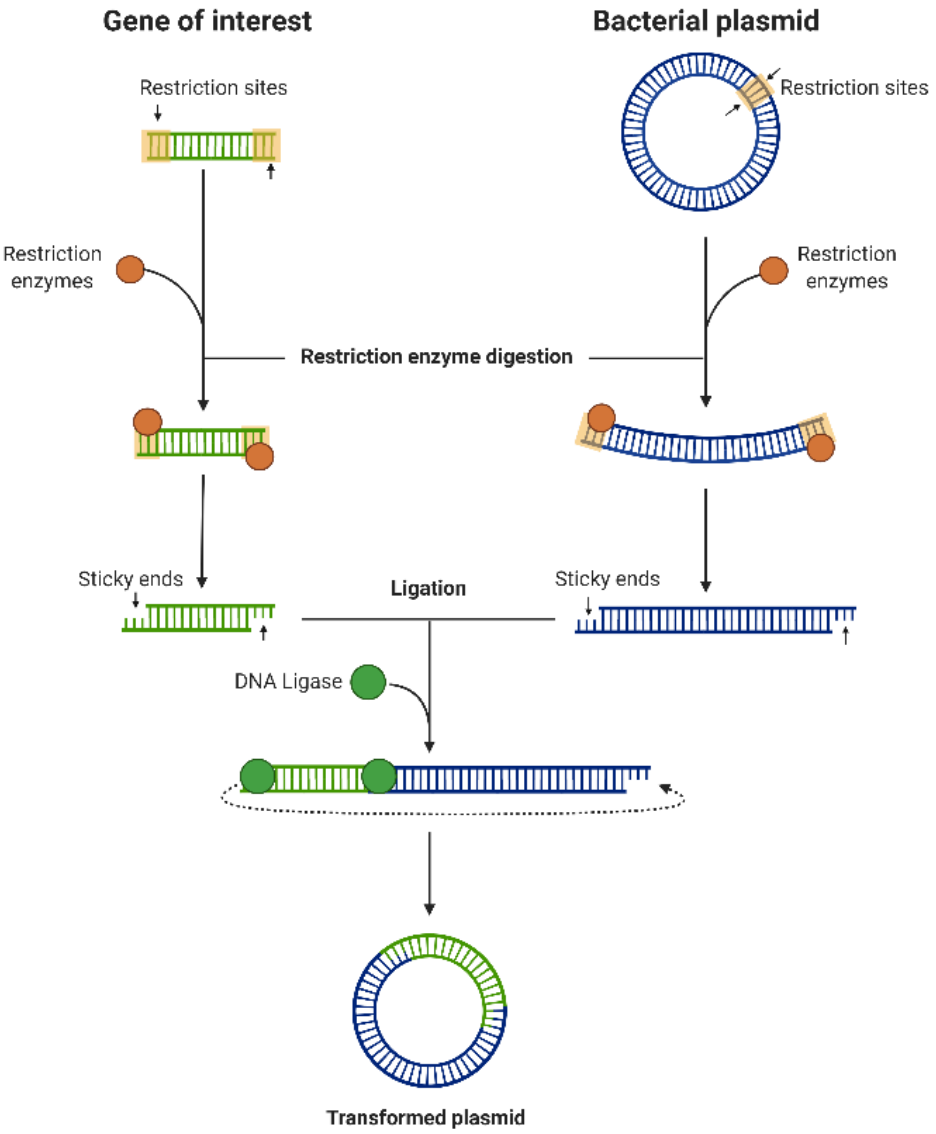


Figure 8. An overview of conventional cloning, *in which the gene of interest is obtained through restriction enzyme cleavage, and the receiving plasmid is linearized by the same restriction enzymes. DNA ligase joins the DNA of interest into the vector, resulting in a transformed plasmid (Image created using BioRender).*

Forward and reverse primers were designed using *Primer3plus* online web-tool (https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). Their likelihood to form secondary structures are minimalized by choosing the primer that has the least chances to form hairpins, self and homodimers using the *OligoAnalyser* tool from Integrated DNA Technologies. The PCR reaction to generate the insert sequence was done with KOD Hot Start DNA Polymerase from *Merck*, starting with initial activation

of the polymerase at 95°C for 2 minutes, followed by 30 cycles of denaturation at 95°C for 20 seconds, annealing for the appropriate temperature (primer specific) for 10 seconds and extension at 70°C for 10 seconds. At the end a final step of extension at 70°C for 2 minutes was applied (Figure 9). The PCR products were run on a 1% agarose gel, then purified further to remove the primers, nucleotides, enzymes and other impurities from the DNA samples using MinElute PCR Purification Kit by *Qiagen*.

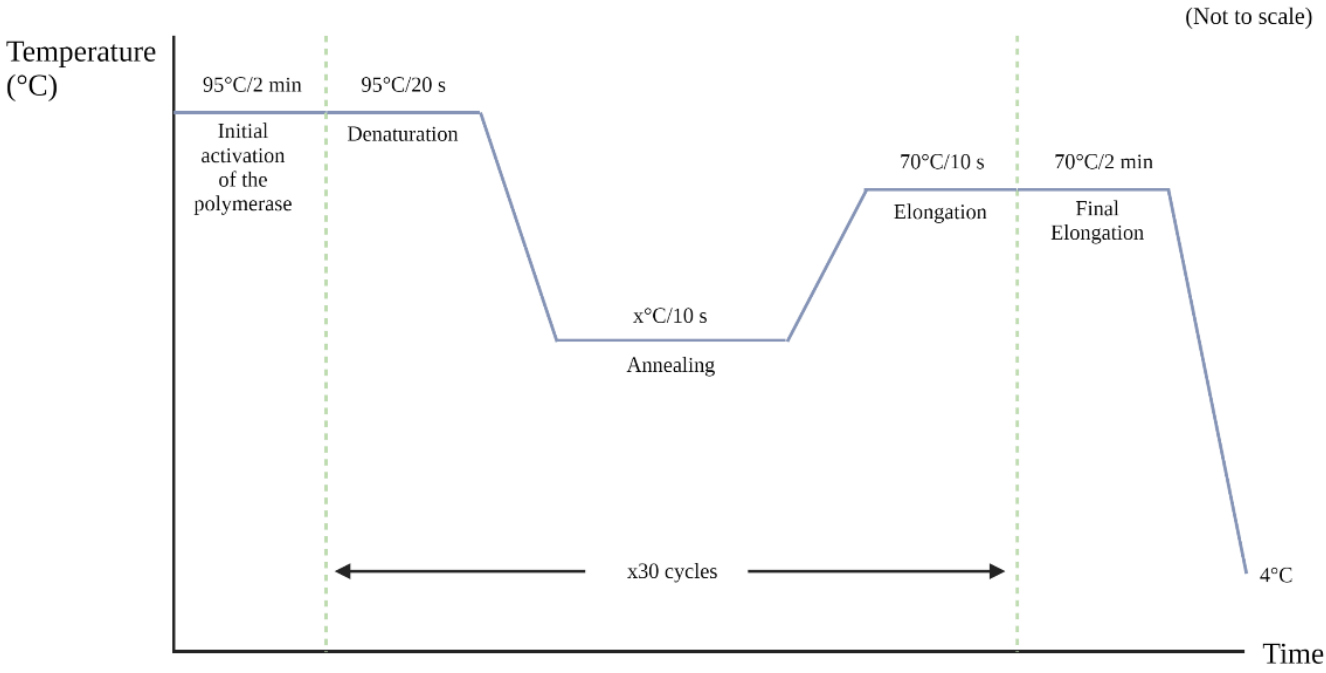


Figure 9. PCR cycles for insert generation. *It begins with the initial activation of the polymerase at 95°C for 2 minutes, followed by 30 cycles of denaturation at 95°C for 20 seconds, and annealing for the appropriate temperature (primer specific) for 10 seconds and extension at 70°C for 10 seconds, at the end a final step of extension at 70°C for 2 minutes is carried out. (Image created using BioRender).*

Both the insert sequence and the target vector were digested with two restriction enzymes that flank the sequence of the insert and are present in the Multiple Cloning Site (MCS) of the vector, creating two distinct restriction sites on both the insert and the target vector which allow the sequence to be inserted with directionality. The PCR products were

digested with restriction enzymes at 37°C for 3 hours, and later the inserts were purified with MinElute PCR Purification Kit. The vectors were additionally treated with Antarctic Phosphatase from *New England Biolabs* to prevent plasmid re-circularization by dephosphorylating the ends of the vector, before been purified with the MinElute PCR Purification Kit. The different restriction enzymes, the conditions used for the constructs and the primers used for insert generation are shown in table 1.

Primer	**Sequence**	**Tm**
LARP4A 1-50 Fw	5'-TTCCGGATCCATGTTGCTTTTCGTGGAGCAG-3' Blue: spacer Red: BamH1 Restriction site Green: Annealing sequence	56.2°C
LARP4A 1-50 Rv	5'-GAAGAACTCGAGTCATGATGTAGCTGCTATTTCATGC-3' Blue: spacer Red: Xho1 Restriction site Green: Annealing sequence	53.2°C
LARP4A 1-79 Fw	5'-TTCCGGATCCATGTTGCTTTTCGTGGAGCAG-3' Blue: spacer Red: BamH1 Restriction site Green: Annealing sequence	56.2°C
LARP4A 1-79 Rv	5'-GAAGAACTCGAGTCATGTGGTTTCACAAGATGAAGAA-3' Blue: spacer Red: XhoI Restriction site	52.4°C

	Green: Annealing sequence	
LARP4B NTD 1-328 Fw	5'-TTCCCCATGGGCAGCAGCCATCACCATC-3' Blue: spacer Red: NcoI Restriction site Green: Annealing sequence	56.9°C
LARP4B NTD 1-328 Rv	5'-GAAGAAGCGGCCGCTCAGTCCAGGGGTCTAAATCCATT-3' Blue: spacer Red: NotI Restriction site Green: Annealing sequence	54.2°C

Table 1. Primers used for the amplification of the constructs made via conventional cloning. *Note the annealing temperature chosen is always the lower Tm between the two primers. The colour-code highlights the different functions of the sequence parts.*

The plasmid used for LARP4A 1-50/1-79 are pET28 SUMO and the plasmid used for LARP4B NTD 1-328 was pRSF Duet1 (see table 4 for additional details, page 47). The ligation reaction was performed using T4 DNA ligase from Promega, in a 16°C overnight reaction, with a 7:1 insert to vector molar ratio. The products were used to transform chemically competent *E. coli* DH5α cells overnight, 50 µL cells were thawed on ice for 5 minutes and the product of the ligation was added to the competent cells, which were placed on ice for 30 minutes. The cells were then heat-shocked for 30 seconds at 42°C to allow the DNA to enter the cells and cooled for 5 minutes on ice (Froger and Hall, 2007). 1mL of LB media was added to the cells which were left at 37°C for 1 hour. 100 µl of

cells were plated on an antibiotic selection plate and left for 16-20 hours at 37°C. A colony PCR using REDTaq® DNA Polymerase from *Merck* was then performed according to the manufacturer's protocol with the original primers to verify correct insert acquisition and positive samples containing the correct size of insert were purified using Monarch® Plasmid Miniprep Kit according to the manufacturer's protocol from *New England Biolabs* and sent for sequencing to *Eurofins Genomics.*

2.2.2 Site-directed mutagenesis

LARP4B mutants (see table 2) were generated using the Q5® Site-directed mutagenesis kit from New England Biolabs which uses the Q5® Hot Start High-Fidelity DNA Polymerase with custom mutagenic primers designed by NEBaseChanger® to create insertions, deletions and substitutions. It is a simple method to create specific, targeted changes in double stranded plasmid DNA (Edelheit, Hanukoglu and Hanukoglu, 2009).

The vector was amplified by PCR following the following protocol: initial denaturation of 98°C for 30 seconds, followed by 25 Cycles of 98°C for 10 seconds, annealing at the appropriate temperature (primer specific) for 20 seconds, extension at 72°C for 30 seconds/kb and a final extension at 72°C for 2 minutes (Figure 10). After the amplification, the mixture was treated with an enzyme mix consisting of a kinase, a ligase and DpnI to allow the plasmid to circularize and removing the template DNA. After a transformation in NEB® 5-alpha Competent E. coli cells provided with the kit, the DNA was purified using miniprep and the DNA were sent for sequencing for verification of mutagenesis.

Primer	Sequence	Tm	Ta
LARP4B La-module 151-328 Fw	5’-ACCCCTGGACTAGAGCCTGTATG-3'	61°C	62°C
LARP4B La-module 151-328 Rv	5'-CTAAATCCATTCTTTGGCAAAAATG-3'	61°C	
LARP4B NTD T163A Fw	5'-ACTTAAAAAAGCATTGGAATTCTGCTTATCTAG-3'	62°C	64°C
LARP4B NTD T163A Rv	5'-ACTTCTCGGGGGTCTTCC-3'	66°C	
LARP4B NTD F166A Fw	5'-AACATTGGAAGCCTGCTTATCTAGG-3'	56°C	57°C
LARP4B NTD F166A Rv	5'-TTTTTAAGTACTTCTCGGG-3'	56°C	
LARP4B NTD C167A Fw	5'-ATTGGAATTCGCCTTATCTAGGGAGAAC-3'	56°C	57°C
LARP4B NTD C167A Rv	5'-GTTTTTTTAAGTACTTCTCGG-3'	57°C	
LARP4B NTD D176A Fw	5'-CCTTGCTAGTGCCATGTATCTTATATC-3'	60°C	59°C
LARP4B NTD D176A Rv	5'-TTCTCCCTAGATAAGCAG-3'	57°C	
LARP4B NTD Y178A Fw	5'-TAGTGACATGGCTCTTATATCACAGATGGATAG-3'	59°C	60°C
LARP4B NTD Y178A Rv	5'-GCAAGGTTCTCCCTAGATAAG-3'	62°C	
LARP4B NTD L197A Fw	5'-GGTGGCTAACGCCGACCACATCAAG-3'	62°C	59°C
LARP4B NTD L197A Rv	5'-GTTGTGATTGGCACATAC-3'	58°C	
LARP4B La-module T163A Fw	5'-ACTTAAAAAAGCATTGGAATTCTGCTTATCTAG-3'	62°C	64°C
LARP4B La-module T163A Rv	5'-ACTTCTCGGGGGTCTTCC-3'	66°C	
LARP4B La-module F166A Fw	5'-AACATTGGAAGCCTGCTTATCTAGG-3'	56°C	57°C
LARP4B La-module F166A Rv	5'-TTTTTAAGTACTTCTCGGG-3'	56°C	
LARP4B La-module C167A Fw	5'-ATTGGAATTCGCCTTATCTAGGGAGAAC-3'	56°C	57°C
LARP4B La-module C167A Rv	5'-GTTTTTTTAAGTACTTCTCGG-3'	57°C	
LARP4B La-module D176A Fw	5'-CCTTGCTAGTGCCATGTATCTTATATC-3'	60°C	59°C
LARP4B La-module D176A Rv	5'-TTCTCCCTAGATAAGCAG-3'	57°C	
LARP4B La-module Y178A Fw	5'-TAGTGACATGGCTCTTATATCACAGATGGATAG-3'	59°C	60°C
LARP4B La-module Y178A Rv	5'-GCAAGGTTCTCCCTAGATAAG-3'	62°C	
LARP4B La-module L197A Fw	5'-GGTGGCTAACGCCGACCACATCAAG-3'	62°C	59°C
LARP4B La-module L197A Rv	5'-GTTGTGATTGGCACATAC-3'	58°C	

Table 2. Primers used for Q5® site-directed mutagenesis. *The forward and reverse primers are listed with their respectively melting and annealing temperature.*

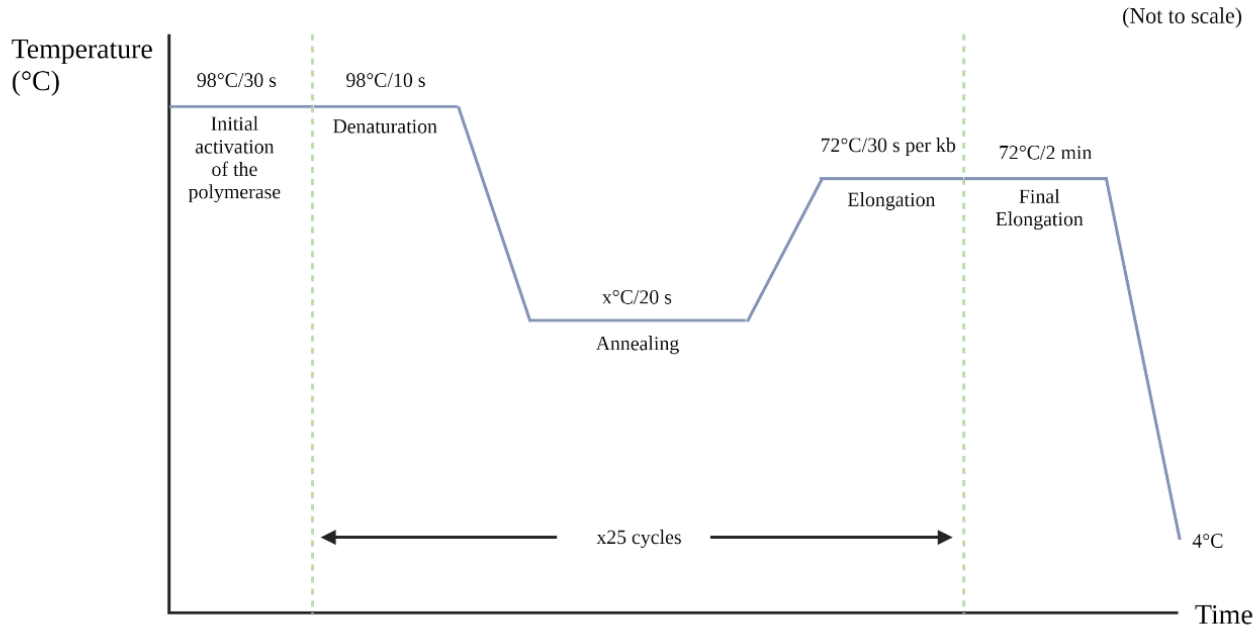


Figure 10. PCR cycles for Q5® site-directed mutagenesis. *Starting with the initial activation of the polymerase at 98°C for 30 seconds, followed by 25 cycles of denaturation at 98°C for 10 seconds, annealing for the appropriate temperature (primer specific) for 20 seconds and extension at 72°C for 30 seconds per DNA kilobase, at the end a final step of extension at 72°C for 2 minutes (Image created using BioRender).*

2.3 Bacterial Expression

2.3.1 Generation of chemical competent cells

The protocol to obtain chemical competent cells was adapted from Nakata et. al. (Nakata, Tang and Yokoyama, 1997). A 50 µL aliquot of previously purchased *E. coli* cells (see Table 3) was allowed to thaw on ice for 5 minutes, then plated on a Luria broth (LB)/agar plate without antibiotics, incubated overnight in an incubator at 37°C. The next day a single colony from said plate was inoculated in 50 mL of sterile LB in a 300 mL flask and the cells were allowed to grow, shaking at 50-60 RPM at 37°C, overnight. The

following day, the overnight culture was diluted 1:100 in 100 mL of fresh sterile LB in a 500 mL flask at 37°C, 300 RPM until the Optical Density at 600 nm (OD_{600nm}) of the cells reached 0.4. The cells were split into two pre-chilled Falcon tubes and cooled on ice for 10 minutes, then centrifuged at 2400xg for 10 minutes at 4°C. The supernatant was discarded, and cells were re-suspended in 10 mL of 100 mM $CaCl_2$ that was previously ice-cold and filtered, and then left on ice for 30 minutes. The cells were once again centrifuged at 2400xg for 10 minutes at 4°C, the supernatant was discarded and then cells were re-suspended in 2 mL of 100 mM $CaCl_2$ and 15% of sterile glycerol. Cells were flash-frozen in liquid nitrogen and stored in aliquots of 50 µL in -80°C.

***E. coli* strain**	**Genotype**	**Features**	**Supplier**
XL10-Gold® Ultracompetent Cells	Tet^R Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F´ proAB lacIqZΔM15 Tn10 (Tet^R) Amy Cam^R]	XL10-Gold ultracompetent cells were created for transformation of large DNA molecules with high efficiency. These cells exhibit the Hte phenotype, which increases the transformation efficiency of ligated and large DNA molecules. XL10-Gold ultracompetent cells are ideal for constructing plasmid DNA libraries because they decrease size bias and produce larger, more complex plasmid libraries.	Agilent
NEB® 5-alpha Competent E. coli	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	NEB 5-alpha Competent E. coli is a derivative of the popular DH5α. It is T1 phage resistant and endA deficient for high-quality plasmid preparations.	New England Biolabs
Rosetta™ 2(DE3) Competent Cells	F^- ompT $hsdS_B(r_B^- m_B^-)$ gal dcm (DE3) pRARE2 (Cam^R)	Rosetta 2 host strains are BL21 derivatives designed to enhance the expression of eukaryotic proteins that contain codons rarely used in E. coli. These strains supply tRNAs for 7 rare codons (AGA, AGG, AUA, CUA, GGA, CCC, and CGG) on a compatible chloramphenicol-resistant	Novagen

		plasmid. The tRNA genes are driven by their native promoters. DE3 indicates that the host is a lysogen of λDE3, and therefore carries a chromosomal copy of the T7 RNA polymerase gene under control of the lacUV5 promoter. Such strains are suitable for production of protein from target genes cloned in pET vectors by induction with IPTG.	

Table 3. List of chemically competent cells used in the project. *In this research, XL10-Gold® Ultracompetent Cells were used to amplify DNA, NEB® 5-alpha Competent E. coli was used to amplify DNA while utilising the Q5® Site-directed mutagenesis kit and Rosetta™ 2(DE3) Competent Cells were used for the E.coli protein expression and purification.*

2.3.2 Bacterial Transformation

Bacteria cells take up foreign DNA in a process called transformation, which was first described by Frederick Griffith (Fred Griffith, 1928). It is a key step of cloning where our plasmid of choice gets transferred within the bacteria, selected by an antibiotic resistance gene and colonies of the transformed bacteria can be used to make large amount of protein (in our case, LARPs) for study. Chemically competent NEB® 5-alpha *E. coli* cells are generally cells used for cloning work and amplification; Rosetta™ 2(DE3) which are made chemically competent, are used for protein expression because they enhance the expression of eukaryotic proteins by providing the rare codons that are not frequently used in *E. coli.*

50 µL of cells were thawed on ice for 5 minutes and 50-100 ng/µl of DNA was added to the cells, which were placed on ice for 30 minutes. The cells were then heat-shocked for 30 seconds at 42°C to allow the DNA to enter the cells and cooled for 5 minutes on ice (Froger and Hall, 2007). 1 mL of LB media was added to the cells which were left at 37°C for 1 hour. 100 µl of cells were plated on an antibiotic selection plate and left for 16-20 hours at 37°C.

2.3.3 Expression tests

Expression tests are used to test which conditions (temperature, incubation time, IPTG concentration) are suitable for protein expression. One colony from the transformed Rosetta™ 2(DE3) cells was inoculated in 15 mL of LB preculture media containing the appropriate antibiotics and grown at 37ºC, shaking at 180 RPM for 16-20 hours. 1 mL of the pre-culture is diluted into 15 mL of fresh LB media, left to grow until the cultures reach OD_{600nm} of 0.4-0.6, and induced by adding 1 mM of Isopropyl β-D-1-thiogalactopyranoside (IPTG). A sample before IPTG induction was collected. IPTG mimics allolactose which is a metabolite of lactose, which when present induces protein expression where the gene is under the control of the lac operator. The proteins for this study starts getting expressed when there is an addition of 1 mM IPTG, since the genes of interest are under control of the lac operator. Samples are taken out for analysis at 3-hour and 16-hour time points following induction and kept at either 37ºC or 18ºC after induction. These samples were centrifuged for 2 minutes at 13100xg, the supernatant was removed, and 2% SDS was added in the exact volume calculated by the equation:

$$V_{SDS\ 2\%\ (ml)} = \frac{OD_{600nm}}{10}$$

Another set of samples were subject to sonication to test for the solubility of the protein. The cells were resuspended in a volume of Lysis buffer containing 50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole, 5% glycerol, 2 mM PMSF, 1 tablet of cOmplete™, EDTA-free Protease Inhibitor Cocktail (*Merck*) and 0.01 g of lysozyme in a volume of:

$$V_{LysisBuffer\ (ml)} = \frac{OD_{600nm}}{10}$$

Cells were incubated on ice for 20 minutes, and then sonicated using the SoniPrep 150 sonicator (*MSE*) for 12% amplitude for a pulse sequence of 5 seconds on, 15 seconds off for a total sonication time of 25 seconds. The samples were spun down at 13100xg for 10 minutes at 4ºC. The supernatants which correspond to the soluble protein fractions, were transferred to a new Eppendorf tube, SDS loading dye 5x (0.25% Bromophenol Blue, 0.5 M DTT, 50% glycerol, 10% SDS, 0.25 M Tris-HCl, pH 6.8) was added to the sample and they were heated for 5 minutes at 95ºC.

The remaining pellets obtained were the insoluble fractions, they were resuspended in a volume of 2% SDS equal to:

$$V_{SDS\ 2\%\ (ml)} = \frac{OD_{600nm}}{10}$$

5x SDS loading dye was added to all samples, heated for 5 minutes at 95ºC, and analysed on SDS-PAGE.

2.3.4 Large scale expressions

The proteins were expressed in *E. coli* Rosetta™ 2(DE3) cells grown in LB medium. One colony of the transformed cells was inoculated in a 50 mL LB preculture with appropriate antibiotics (see Table 4, page 47) and grown at 37°C, shaking at 180 RPM for 16-20 hours. The precultures were diluted in 1-2 litre sterile LB media so that the starting

OD_{600nm} was 0.05-0.1, the cultures were induced with 1mM IPTG when they have reached an OD_{600nm} of 0.6-0.8 (for some constructs even an OD_{600nm} of 1 was used, see table 4 for a detailed list of constructs) and left growing at 18°C or 37°C, shaking at 180 RPM for 16-20 hours. 1 mL samples were taken before IPTG induction and at harvest to test for protein expression. The cultures were centrifuged at 6238xg for 40 minutes to separate the pellets and stored at -20°C.

2.4 Protein purification

2.4.1 Nickel Immobilized Metal Affinity Chromatography (IMAC)

The protein of interest was purified from the bacterial large-scale growths, they were resuspended in 2 mL/gram (of cell pellet) in lysis buffer (50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, 5% glycerol, 2 mM PMSF, 0.01 g of lysozyme (*Sigma*), a tablet of cOmplete™ protease inhibitor cocktail (*Roche*) per 50 mL, pH 8.0) and left on ice for 20-30 minutes to allow for the lysozyme to induce bacterial lysis (Pushkaran *et al.*, 2015). The cells were sonicated for 30% amplitude for a pulse sequence of 5 seconds on, 15 seconds off for a total sonication time of 3 minutes using a Vibra-Cell™ sonicator (*Sonics*). The soluble fractions were separated by centrifugation at 11855xg for 40 minutes at 4°C. The supernatant was filtered using 0.45 µm filters (*Millipore*).

All the proteins for this study have been cloned with a 6xHistidine tag (His-tag) engineered on the N-terminus of the constructs. The His-tag has a selective affinity to immobilized Ni^{2+} ions that are present in an appropriate Nickel Immobilized Metal Affinity Chromatography (IMAC) resin. Electron donor groups on the histidine imidazole ring forms coordination bonds with Ni^{2+}, thus allowing other bacterial proteins to be washed off the resin and only the protein of interest is retained (Bornhorst and Falke,

2010). Figure 11 shows the general principle of Nickel IMAC. For my system I used 5 mL HisTrap™ FF column (*GE healthcare*) pre-equilibrated using Nickel binding buffer (50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, 5% glycerol, pH 8.0) and the ÄKTA system (UPC-900/P-900/Frac-950/CU-950 – *Amersham Bioscience*).

The protein was eluted with a gradient of 10 mM to 500 mM imidazole in 40 column volumes, using a steep increasing gradient (0-100%) with the addition of Nickel elution buffer (50 mM Tris-HCl, 300 mM NaCl, 500 mM imidazole, 5% glycerol, pH 8.0), imidazole competes for the binding of the metal ions present on the resin, hence the His-tagged protein is eluted, and collected in 1.5 mL fractions in a 96-well collector. The fractions containing proteins and the flowthrough were analysed on SDS-PAGE, and the fractions containing high concentration of protein, identified by the molecular weight on the gel are collected and pooled together for on overnight dialysis using a 6-8 kDa cut-off dialysis tube with 5 L Nickel dialysis buffer (50 mM Tris-HCl, 200 mM KCl, 0.2mM EDTA and 1mM DTT, pH 7.25) to remove any imidazole and to get the protein ready for further purification steps. TEV protease, SUMO protease or a 3C Pre-Scission protease were added with the protein depending on the construct to cleave off the His-tag, SUMO-tag or the GST-tag respectively (see table 4 for a list of the constructs used in this study).

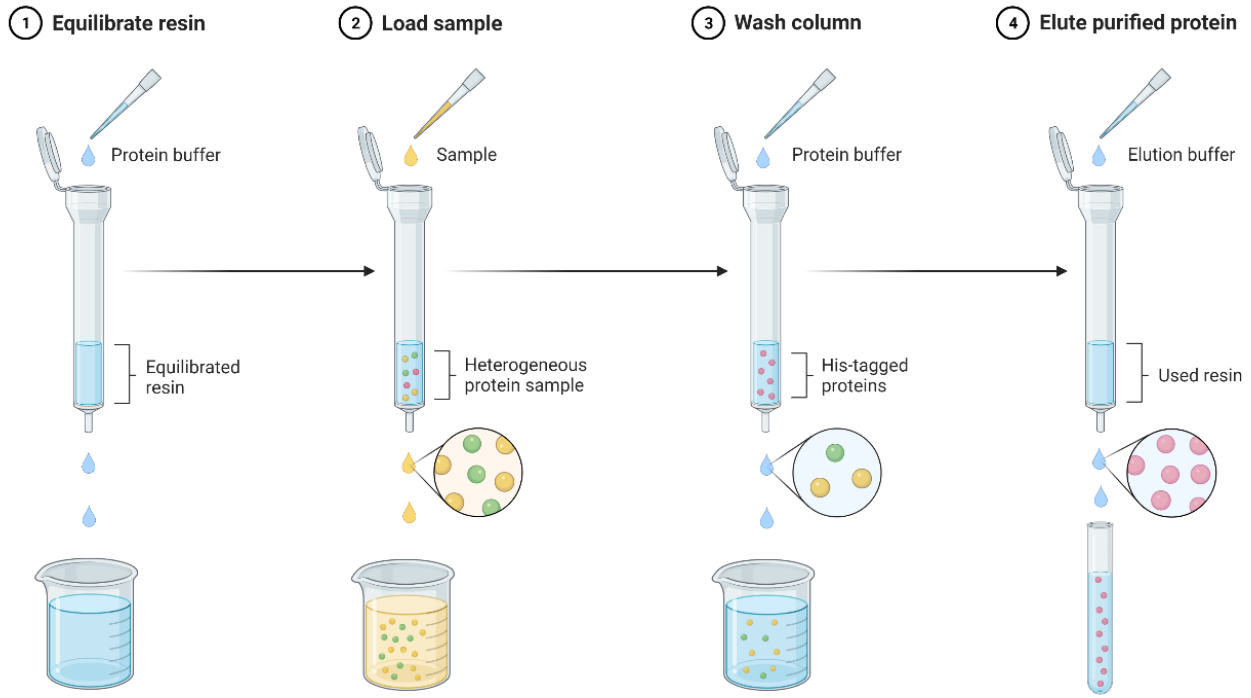


Figure 11. The workflow of Nickel Immobilized Metal Affinity Chromatography. *The resin is equilibrated in the buffer suitable for the protein, then with the introduction of the His-tagged protein, they bind to the Ni^{2+} ions in the resin, and any non-specific interactions are washed out and the protein eluted with a high concentration of imidazole which competes for the binding of the His-tagged proteins with the Ni^{2+} ions, we can obtain purified protein in the end (Image created using BioRender).*

Construct	Tag	Vector	Antibiotic Resistance	Induction OD	Temperature after induction
LARP4A 1-50	N-term His-SUMO	pET28_SUMO	Kanamycin	0.6-0.8	18°C
LARP4A 1-79	N-term His-SUMO	pET28_SUMO	Kanamycin	0.6-0.8	18°C
LARP4A NTD	N-term His	pRSFDuet-1	Kanamycin	0.75	18°C
LARP4A La module	N-term His	pETDuet-1	Ampicillin	0.90	18/37°C
LARP4A NTR	N-term His	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4A L15AW22A	N-term His	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4B NTD 1-328	N-term His-GST	pRSFDuet-1	Kanamycin	0.90	18°C
LARP4B NTD 1-339	N-term His-GST	pRSFDuet-1	Kanamycin	0.90	18°C
LARP4B La module	N-term His-TEV	pRSFDuet-1	Kanamycin	0.90	18°C
LARP4B NTR	N-term His-GST	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LAR4PB L56AW63A	N-term His-GST	pRSFDuet-1	Kanamycin	0.75	18°C
LAR4PB W63F	N-term His-GST	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4B 40-328	N-term His-GST	pRSFDuet-1	Kanamycin	0.80	18°C
LARP4B 71-328	N-term His-GST	pRSFDuet-1	Kanamycin	0.80	18°C
LARP4B 95-328	N-term His-GST	pRSFDuet-1	Kanamycin	0.80	18°C
LARP4B NTD T163A	N-term His-GST	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4B NTD F166A	N-term His-GST	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4B NTD C167A	N-term His-GST	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4B NTD D176A	N-term His-GST	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4B NTD Y178A	N-term His-GST	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4B NTD L197A	N-term His-GST	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4B La module T163A	N-term His-GST	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4B La module F166A	N-term His-GST	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4B La module C167A	N-term His-GST	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4B La module D176A	N-term His-GST	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4B La module Y178A	N-term His-GST	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4B La module L197A	N-term His-GST	pRSFDuet-1	Kanamycin	0.6-0.8	18°C
LARP4B LaM	N-term His-TEV	pRSFDuet-1	Kanamycin	1.00	18/37°C
LARP4B RRM	N-term His-TEV	pRSFDuet-1	Kanamycin	1.00	18/37°C
LARP4A FL	N-term His-SUMO	pET28_SUMO	Kanamycin	0.90	18/37°C
LARP4B FL	N-term His-SUMO	pET28_SUMO	Kanamycin	0.90	18/37°C
His SUMO MLLE	N-term His-SUMO	pET28_SUMO	Kanamycin	0.6-0.8	18°C
SUMO alone	N-term His	pET28_SUMO	Kanamycin	0.6-0.8	18°C

Table 4. List of constructs used in this study. *Details of each construct, including the tag, vector, antibiotic resistance, induction OD_{600nm} and temperature after the induction are listed.*

2.4.2 Separation and removal of tag

Following dialysis, the mixture was then loaded on a second Nickel IMAC resin pre-packed into a gravity column using 5 mLs of Ni-NTA affinity resin (*Generon*), pre-equilibrated with the Nickel binding buffer (50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, 5% glycerol, pH 8.0) to separate the protein from the cleaved tags, the protease and non-digested products. All the proteases used have been engineered to have a His-tag on the N-terminus and therefore be bound to the resin.

The flowthrough of the column contains the cleaved protein that does not have an His-tag and was collected. The column was then washed with 10 ml of Nickel binding buffer (50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, 5% glycerol, pH 8.0), and then eluted with 15 column volumes of Nickel elution buffer (50 mM Tris-HCl, 300 mM NaCl, 500 mM imidazole, 5% glycerol, pH 8.0) which should enable elution of the cleaved tags, the protease and non-digested products. Samples from all fractions were taken and run onto an SDS-PAGE gel to verify the contents.

2.4.3 Ion Exchange purification

Ion exchange chromatography (IEX) is a chromatographic separation method that exploits the surface net charge of the protein. At any pH, proteins will have a net charge that is governed by the amino acid residues that are capable of deprotonation and protonation, acidic residues such as Glutamic acid and Aspartic acid contribute negative charges when deprotonated and basic residues such as Arginine, Histidine and Lysine contribute positive charges when protonated. IEX is split into cation exchange and anion exchange chromatography. We decided on which strategy to use based on the overall charge of the protein of interest (calculated using a theoretical isoelectric point), which

can vary vastly between different proteins. Since I am investigating protein-RNA interactions in this project, it is important to separate the nucleic acids that are often still present following IMAC purification. Since nucleic acids are negatively charged, IEX is a great solution to separate the unwanted nucleic acids from our recombinant protein (Jungbauer and Hahn, 2009).

The HiTrap™ Diethylaminoethanol (DEAE) Sepharose FF column and the HiTrap™ Heparin HP affinity columns, contain DEAE and Heparin (negatively charged) respectively. DEAE is an ion-exchange resin coated with positively charged counter-ions (Corradini, Cavazza and Bignardi, 2012), that was used based on the increase of negative charge upon binding of the proteins, the resin will bind to anything that is negatively charged (including nucleic acids) and leave the positively charged entities in the flowthrough. Heparins are negatively charged polydispersed linear polysaccharides which can bind a wide range of biomolecules, it is not only an affinity ligand but also an ion exchanger with high charge density and distribution (Xiong, Zhang and He, 2008). Both the IEX columns are first equilibrated with IEX binding buffer (50 mM Tris-HCl, 100 mM KCl, 0.2 mM EDTA, 1 mM DTT, 10% glycerol, pH 7.25). The protein was eluted with a salt gradient of 100 mM to 1 M KCl in 40 column volumes, using a steep increasing gradient (0-50%) with the addition of IEX elution buffer (50 mM Tris-HCl, 2 M KCl, 0.2 mM EDTA, 1 mM DTT, 10% glycerol, pH 7.25). The charged salt ions compete with bound proteins for the charged resin functional groups. In general, proteins with few charged groups will elute at low salt concentrations, whereas proteins with many charged groups will have greater retention times and elute at high salt concentrations (Acikara, 2013). You can also have a protein with many charged groups with even distribution of positive and negative charges, the way the elution happens is usually

determined by the theoretical isoelectric point and by trial and error on the actual IEX column.

The net charge of all the proteins and proteins mutants has been estimated using the ProtParam tool in *ExPASy* (Gasteiger *et al.*, 2005). Depending on the protein or mutant purified, either DEAE or heparin column was used (discussed in Chapter 3). The protein is either collected in the flowthrough of DEAE or elution of heparin. All fractions were analysed on SDS-PAGE and fractions containing the protein of interest were collected, pooled and dialysed overnight using 5 L of final protein buffer (20 mM Tris-HCl, 100 mM KCl, 0.2 mM EDTA and 1 mM DTT, pH 7.25).

2.5 Protein concentration and measurements

After purifications the proteins are spin concentrated using 30 kDa, 10 kDa or 5 kDa ultrafiltration centrifugal concentrators from *Sartorius* spinning at 2000xg for 10 minutes until the desired protein concentration is reached, making sure that there is no concentration gradient by resuspending the sample in between runs. After the protein has reached the required concentration, around 100-200 µM, they are flash frozen with liquid nitrogen and stored in -80°C.

Protein concentration was measured by a ND-100 NanoDrop spectrophotometer (*ThermoScientific*) using the Beer-Lambert law (Swinehart, 1962):

$$c = \frac{A_{280nm}}{\varepsilon\lambda}$$

Where c is the protein concentration, A_{280nm} is the absorbance at 280 nm, ε is the extinction coefficient calculated by ProtParam tool in ExPASy and λ is the path length (0.1 cm).

2.6 Sodium Dodecyl Sulphate Poly-acrylamide Gel Electrophoresis (SDS-PAGE)

To visualise the content of the protein mixture at any stage, SDS-PAGE is used to separate the proteins by their molecular sizes, with smaller ones travelling further along the gel and the larger ones at the top of the gel. Proteins contains an overall positive or negative charge, by denaturing the protein using SDS, it denatures the protein and provides a uniform negative charge, it is then possible for them to move and migrate to the positive electrode (Shapiro, Viñuela and v. Maizel Jr., 1967). 10%, 12% or 14% home-made acrylamide gels were prepared with Accugel 29:1 30% Acrylamide Bisacrylamide solution (*National Diagnostics*). Lower acrylamide-percentage gels were used for larger proteins of interest since they resolve better in the higher molecular weight and vice versa. After the addition of 1% Ammonium Persulfate and 0.01% of TEMED the gel is allowed to polymerise for 1 hour topped with a comb allowing 10 wells or 15 wells. Once fully polymerized, the gels are assembled in the SDS-PAGE Mini-PROTEAN® Tetra Cell system (Bio-Rad) and topped with Tris-Glycine SDS running buffer (25 mM Tris-HCl, 192 mM Glycine, 0.1% SDS). SDS loading dye 5x (0.25% Bromophenol Blue, 0.5 M DTT, 50% glycerol, 10% SDS, 0.25 M Tris-HCl, pH 6.8) was added to the sample and they were heated for 5 minutes at 95ºC. Gels were run at 190 V for 45 minutes then stained with Coomassie Brilliant Blue© staining solution (1% Coomassie Brilliant Blue© (*Bio-Rad*), 45% methanol, 45% H2O, 10% acetic acid) for 1 hour to visualise the protein bands, de-stained using the de-staining solution (45% methanol, 45% H2O, 10% acetic acid) for 1 hour and imaged with Molecular Imager® Gel Doc XR System (*Bio-Rad*).

2.7 Assays and Analysis

2.7.1 Electrophoretic Mobility Shift Assay (EMSA) - Fluorescence

Electrophoretic Mobility Shift Assay (EMSA) or sometimes called Gel-Shift assays is a commonly used technique to detect protein complexes with nucleic acids. It is based on the observation that the electrophoretic mobility of a protein-nucleic acid complex is usually lower with respect to the free nucleic acid, so one can clearly distinguish the bound complex from the free nucleic acid, in our case the RNA. To study the protein-RNA interaction, EMSAs are performed to determine whether a protein is capable of binding to an RNA sequence. This is done by having the protein serially diluted and adding RNA to the samples. The range of protein concentrations used are reported below. If there is binding, there will be a formation of a complex which in turn causes a shifted band to appear due to the increase in size (Hellman and Fried, 2007) (Figure 12).

PolyA20 RNA (5'-AAAAAAAAAAAAAAAAAAAA-3') and AU-rich/CKB RNA (5'-UGGUGAGUUUAUUUUUUUGA-3') labelled at the 5' end with 5-Carboxyfluorescein, purchased from *IBA Gmbh* were used as the ligand for the LARP4A and LARP4B respectively at a final working concentration at 10 nM. The experiments were conducted with the first point of protein concentration ranging from 100-200 µM then serial (1:1) dilutions were done to obtain the binding affinity. The reaction was done in a EMSA buffer of 20 mM Tris pH 8, 100 mM KCl, 5% glycerol, 1 mM DTT, 0.1 mg/ml BSA. 0.01 mg/ml tRNA mix from *E. coli* was added in some experiments as the competitor to assess the RNA-binding specificity of LARP4A and LARP4B proteins. 2 µl of 30% Ficoll was added to each 12 µl reaction and each reaction was loaded on a 9% native polyacrylamide gel pre-run 1 hour at 4°C, 100 V in 0.5x Tris-borate-EDTA buffer. The gel was run 1 hour at 4°C, 125 V and visualized by ChemiDoc MP imager (*Bio-rad*) using the Epi Blue-Light module for excitation and the 530/28 nm filter was used to visualise

the gels. The band intensities were determined using ImageLab (*Bio-Rad*) and the dissociation constants (K_D) were determined by fitting the data into GraphPad Prism to a sigmoidal binding curve using the Hill equation: $Y=B_{max}*X\text{^}h/(K_D\text{^}h + X\text{^}h)$, Where X is the concentration of ligand, Y is the specific binding, B_{max} is the maximum binding, h is the Hill coefficient, K_D is the dissociation constant.

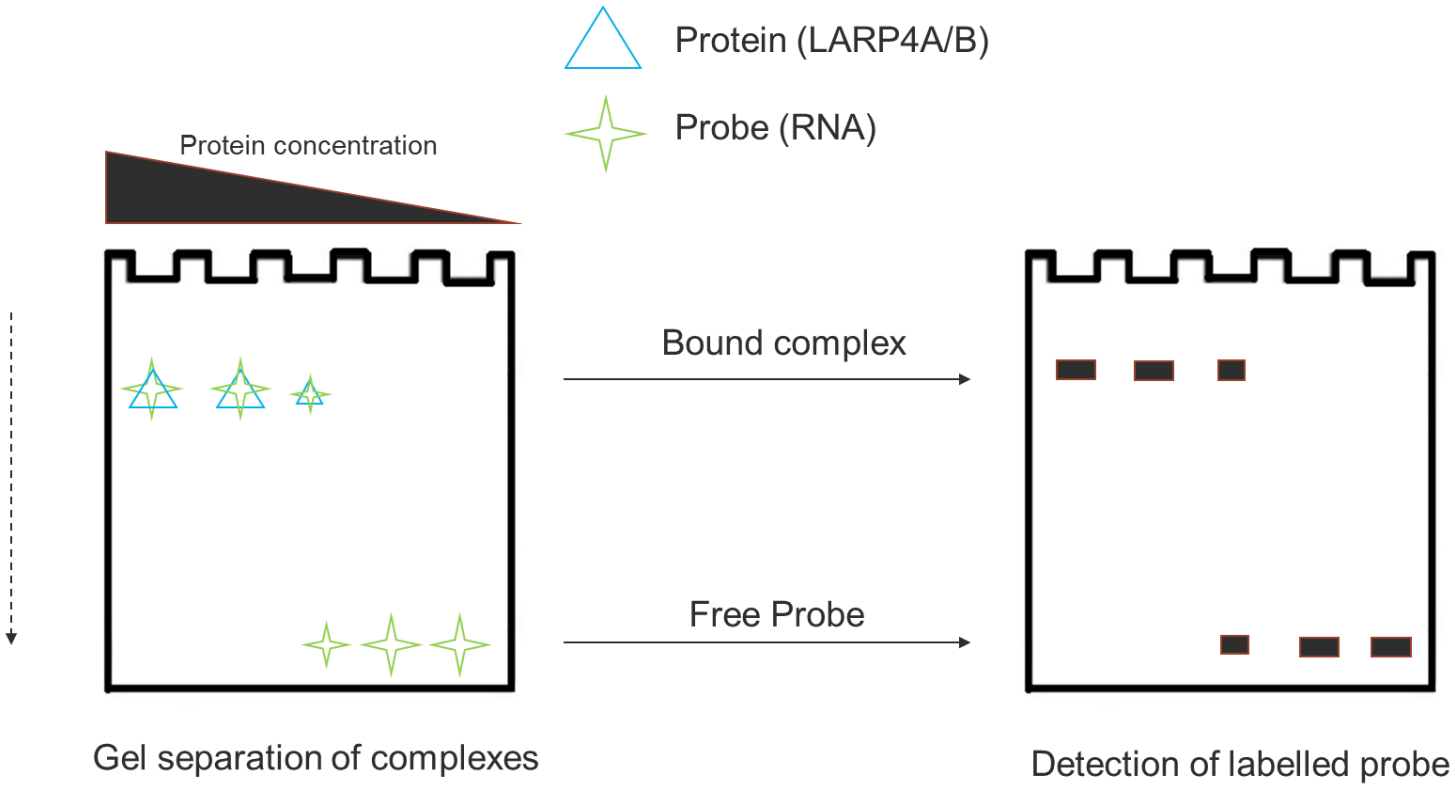


Figure 12. A diagram showing the principles of Electrophoretic Mobility Shift Assay (EMSA). *EMSA is a rapid and sensitive method to detect protein–nucleic acid interactions. It is based on the observation that the electrophoretic mobility of a protein–nucleic acid complex is typically lower than that of the free nucleic acid. Upon detection of the labelled RNA using either fluorescence or radioactivity, protein-RNA complexes are retarded in the gel, therefore appear higher than the free RNA, which appear lower than the complexes.*

2.7.2 Electrophoretic Mobility Shift Assay (EMSA) – Radioactive

For radiolabelled RNA, PolyA15 RNA (5'-AAAAAAAAAAAAAAA-3') and AU-rich/CKB RNA (5'-UGGUGAGUUUAUUUUUUUGA-3') were purchased from *IBA Gmbh*, labelled at the 5' end with γ-^{32}P ATP using T4 polynucleotide kinase. The kinase reaction contained 1 µL of RNA at 50 µM, 1 µL ATP at 50 µM, 3 µL of γ-^{32}P ATP at 10

mCi/µL, 2.5 µL of T4 reaction buffer, 1 µL of T4 polynucleotide kinase and topped up to 25 µL using nuclease free water. The reaction was left at 37°C for 30 minutes, the unincorporated nucleotides were removed on G-25 spin columns designed for the removal of unincorporated nucleotides from end-labelled oligonucleotides, the sample was passed through the column and centrifuged for 2 minutes at 735xg. The final working concentration of the RNA is 2 nM.

EMSA experiments were conducted with a starting protein concentration ranging from 10-200 µM then serial (1:1) dilutions or 1:2 dilutions (see table 5) was done to obtain the binding affinity.

Protein	**Target RNA**	**Starting Concentration (µM)**	**Type of dilution**
LARP4A 1-50	oligoA15	200	1:1
LARP4A 1-79	oligoA15	200	1:1
LARP4B NTD	AU-rich/CKB	10	1:1
LARP4B NTD	oligoA20	200	1:1
LARP4B NTD	oligoU20	200	1:1
LARP4B NTD	oligoC20	200	1:1
LARP4B NTR	AU-rich/CKB	50	1:1
LARP4B La-module	AU-rich/CKB	50	1:1
LARP4B 40-328	AU-rich/CKB	10	1:2
LARP4B 71-328	AU-rich/CKB	50	1:1
LARP4B 95-328	AU-rich/CKB	50	1:1

LARP4B LaM	AU-rich/CKB	200	1:1
LARP4B RRM	AU-rich/CKB	200	1:1
LARP4B L56AW63A	AU-rich/CKB	10	1:1
LARP4B W63F	AU-rich/CKB	10	1:1

Table 5. List of proteins used for radioactivity EMSAs in this study. *The target RNA, starting protein concentration (µM), and the type of dilution are listed.*

The EMSA reaction was done in a buffer of 20 mM Tris pH 7.25, 200 mM KCl, 5% glycerol, 1 mM DTT, 0.1 mg/ml BSA with or without 0.01 mg/ml tRNA from *E. coli* as the competitor to assess the RNA-binding specificity of LARP4A and LARP4B proteins. 2 µl of 30% Ficoll was added to each reaction and each reaction was loaded on a 9% native polyacrylamide gel pre-run 1 hour at 4°C, 100 V in 0.5x Tris-borate-EDTA buffer. The gel was run 1 hour at 4°C, 125 V and then the gels were dried using 3MM chromatography paper using a Model 583 Gel Dryer (*Bio-Rad*) for 1 hour, then exposed to a phosphor-imaging plate overnight. The plate was visualised by the phosphor-imager Typhoon Trio and band intensities were quantified using ImageQuant TL software and the dissociation constants (K_D) were determined by fitting the data into GraphPad Prism to a sigmoidal binding curve using the Hill equation shown previously: $Y=B_{max}*X\text{^}h/(K_D\text{^}h + X\text{^}h)$.

2.7.3 Microscale Thermophoresis (MST)

Microscale Thermophoresis (MST) is an effective technique to determine the binding affinity of biomolecules by detecting a temperature-induced change in the fluorescence of a target as a function of the concentration of a non-fluorescent ligand. The basis of

MST depends on the thermophoresis of a molecule, which is the directed movement of particles during a microscopic temperature gradient induced by an infrared laser, and a temperature-induced intensity change of the fluorescent probe, which can be influenced by binding events. Depending on factors such as size, charge or conformation, they will give out different signals in an unbound or a bound state which is all measured in the machine (Jerabek-Willemsen *et al.*, 2014).

For my studies, the MST experiments were performed at settings of 50% LED power and 20% MST power on a Monolith NT.115 (*Nanotemper Technologies*) at 25°C with standard capillaries. The reaction buffer was 20 mM Tris pH 7.25, 100 mM KCl, 0.2 mM EDTA, 1 mM DTT AND 0.05% Tween-20 in a total volume of 20 µL. PolyA20 RNA (5'-AAAAAAAAAAAAAAAAAAAA-3') and AU-rich/CKB RNA (5'-UGGUGAGUUUAUUUUUUUGA-3') labelled at the 5' end with 5-Carboxyfluorescein was used as the ligand for the LARP4A and LARP4B respectively at a final working concentration at 25 nM. The experiments were conducted with a starting protein concentration ranging from 100-200 µM then serial (1:1) dilutions was done (see Table 6).

Protein	**Target RNA**	**Starting Concentration (µM)**	**Type of dilution**
LARP4B NTD	AU-rich/CKB	25	1:1
LARP4B NTR	AU-rich/CKB	50	1:1
LARP4B La-module	AU-rich/CKB	200	1:1

LARP4B NTD T163A	AU-rich/CKB	50	1:1
LARP4B NTD F166A	AU-rich/CKB	50	1:1
LARP4B NTD C167A	AU-rich/CKB	50	1:1
LARP4B NTD D176A	AU-rich/CKB	50	1:1
LARP4B NTD Y178A	AU-rich/CKB	50	1:1
LARP4B NTD L197A	AU-rich/CKB	50	1:1
LARP4B La-module T163A	AU-rich/CKB	200	1:1
LARP4B La-module F166A	AU-rich/CKB	200	1:1
LARP4B La-module C167A	AU-rich/CKB	200	1:1
LARP4B La-module D176A	AU-rich/CKB	200	1:1
LARP4B La-module Y178A	AU-rich/CKB	200	1:1
LARP4B La-module L197A	AU-rich/CKB	200	1:1

Table 6. List of protein used for MST experiments in this study. *The target RNA, starting protein concentration (µM), and the type of dilution are listed.*

Raw data were analysed from the Monolith NT.115 software, and the dissociation constants (K_D) were determined by fitting the data into GraphPad Prism to a sigmoidal binding curve using the Hill equation.

2.7.4 Circular Dichroism (CD)

Circular Dichroism is a biophysical method for analysing the secondary structure, folding and the binding properties of proteins. Proteins are optically active macromolecules that exhibit unequal absorption of left (A_L) and right (A_R) circularly polarized light (Greenfield, 2007). When asymmetric molecules interact with light, they might absorb left and right circularly polarized light to different extents, and it can be expressed as follows:

$$CD = AL - AR$$

CD signal of proteins is usually separated into either far UV or near UV. Far UV or backbone CD with wavelengths ranging from 190 to 250 nm originates from backbone and peptide bonds signals of the protein, this spectrum provides info on secondary structure content (Rodger and Marshall, 2021). Different secondary structures have distinctive CD trace characteristics (Micsonai *et al.*, 2015). Far UV CD spectra can be a good indicator of whether secondary structure is changing when comparing between different mutants or upon ligand binding. The near UV region is generally from 250 to 300 nm and is an important region that contains signals from the aromatic amino acids (tryptophan, tyrosine and phenylalanine) within the protein and could reflect tertiary structure properties of proteins (Rodger and Marshall, 2021). The near UV region also includes signals from disulphide bonds.

In this study proteins subjected to CD (LARP4A 1-50, LARP4A 1-79, LARP4B NTD, LARP4B NTD T163A, F166A, C167A, D176A, Y178A, L197A, LARP4B La-module, LARP4B La-module T163A, F166A, C167A, D176A, Y178A and L197A) were dialysed in CD buffer (20 mM Tris pH 7.25, 100 mM KCl and 1 mM DTT) with a final concentration of 0.2 mg/ml. The experiments were run with the help of Dr. Tam Bui on the Applied Photophysics Chirascan plus spectrometer (*Leatherhead, UK*), using Suprasil rectangular cells of 10- and 5-mm path lengths (*Starna Scientific Ltd*) in the region of 400 to 190 nm under constant nitrogen flush. The experiments were run at 25°C, with 2 nm spectral bandwidth, 1 nm data step-size, either 1 or 1.5 s instrument time points.

Prior to measuring the CD spectra, UV spectra from regions between 400 to 230 nm were obtained to determine accurately protein concentration. The CD spectra were taken from 260 to 190 nm, with the raw data in millidegrees, and converted into mean residue ellipticity [θ] ($deg \cdot cm^2 \cdot dmol^{-1}$). The data was post-processed by Savitsky-Golay smoothing method (Luo, Ying and Bai, 2005) with a convolution width factor of 4.

Chapter 3. Cloning, expression and purification of recombinant proteins used in this study

3.1. Overview

In this study, several recombinant proteins were generated from *E. coli*. These included LARP4A and LARP4B full length and N-terminal domain (NTD) deletion mutants. As the study focused principally on the NTD and its interactions with RNA and protein partners, several mutants of the NTD were designed, cloned and prepared, Figure 13 shows a complete schematic of all mutants. LARP4A and LARP4B share a similar architecture, they both contain regions that are known to interact with protein partners such as Poly A binding Protein (PABPC1) which controls critical aspects of mRNA metabolism (Mangus, Evans and Jacobson, 2003; Thompson and Gilbert, 2017; Nicholson and Pasquinelli, 2019), and a La-module consisting of a La motif (LaM) and an RNA-recognition motif (RRM). However, they are still known to interact with different RNA targets, LARP4A was tested for interaction with several homopolymers (poly A, U, C, G) using electrophoretic mobility shift assay (EMSA) and isothermal titration calorimetry (ITC) and showed specificity towards poly A RNA of ≥15 nt (Yang *et al.*, 2011); whereas LARP4B was shown to bind to AU-rich sequences by PAR-CLIP analysis (Küspert *et al.*, 2015). Hence the aim of the study is to produce several constructs of LARP4A and LARP4B and test the extent of RNA-binding and pinpoint the exact RNA binding locus of LARP4B, using LARP4A as a positive control. *E. coli* was chosen as expression system since we can obtain the protein relatively fast, it is inexpensive, high-density cultures are easily obtained (Lozano Terol *et al.*, 2021) and this system has been used in the Conte Laboratory already to make LARP4A and LARP4B proteins.

Figure 13. A list of constructs purified in this study. *The affinity tags and the protease cleavage sites are coloured in green, followed by the construct. His: 6xHistidine tag, SUMO: SUMO tag, TEV: TEV protease cleavage site, GST: GST-tag, 3C: 3C protease cleavage site, PAM2w: Poly(A) Binding Protein Interaction Motif-2, PBM: Poly(A) Binding Protein Interaction Protein Motif, LaM: La motif, RRM1: RNA recognition motif. (Image created using BioRender)*

3.2. Cloning

3.2.1 Cloning of LARP4A mutants spanning residues 1-50 and 1-79

In the investigations of LARP4A interactions with oligoA RNA, several deletion mutants of LARP4A involving the full N-terminal Domain encompassing residues 1-287, and deletion mutants 24-287, 50-287 and 79-287 were employed in EMSA (Cruz-Gallardo *et al.*, 2019). The work already conducted by a postdoctoral researcher in the Conte lab, Isabel Cruz Gallardo, showed that the region between residues 1 and 24 was important for oligoA RNA binding, with the PAM2w mutant L15AW22A mutant having over 20x less affinity to oligoA RNA when compared to the wild-type NTD (Cruz-Gallardo *et al.*, 2019). When I joined the lab, I contributed to this research in investigating the importance of the region encompassing residues 1-24 to RNA binding.

Given that the PAM2w region (residues 13-26) is imperative in RNA binding as shown by previous experiments, the question was whether the short fragment spanning residues 1-24 and encompassing the PAM2w would be sufficient for oligoA RNA binding. I cloned 2 deletion mutants, spanning residues 1-50 and 1-79. These are longer than just the first 24 residues, as it was hypothesised that the longer fragments may be more stable for *E.coli* expression.

Traditional cloning was performed to achieve LARP4A 1-50 and 1-79 into an empty pET28 SUMO vector using full length LARP4A in pCMV2 FLAG as a template. The pET28 SUMO vector was chosen since it includes a SUMO tag which promotes folding and stability of the target protein, and large quantities of the protein could be produced (Müller *et al.*, 2001). The insert sequence was generated using PCR and primers (Table 1) that were chosen to amplify residues 1-50 and 1-79. (See Figure 14)

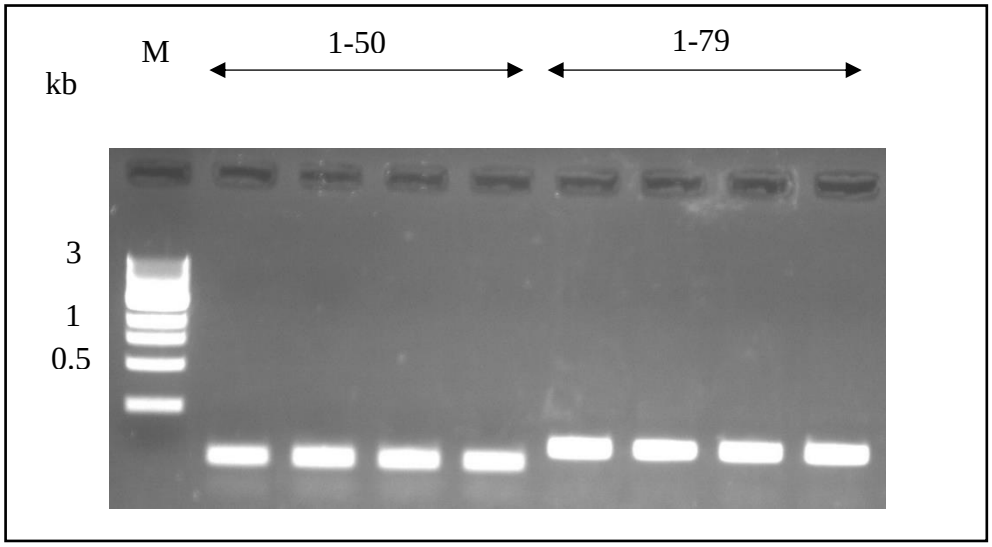


Figure 14. PCR reaction of LARP4A 1-50 and 1-79. *Insert generation of LARP4A 1-50 (Theoretical size 150bp) and 1-79 (Theoretical size 237bp). M: Marker*

After the PCR reaction generating the insert sequence, restriction enzymes XhoI/BamHI-HF were used to generate restriction enzyme sites on both the insert and the vector pET28 SUMO, ligated using T4 DNA ligase and transformed in competent cells. There were a high number of colonies after ligation, so using the original primers, five random colonies were chosen to undergo colony PCR reactions to check if the final vector had contained the insert (See Figure 15).

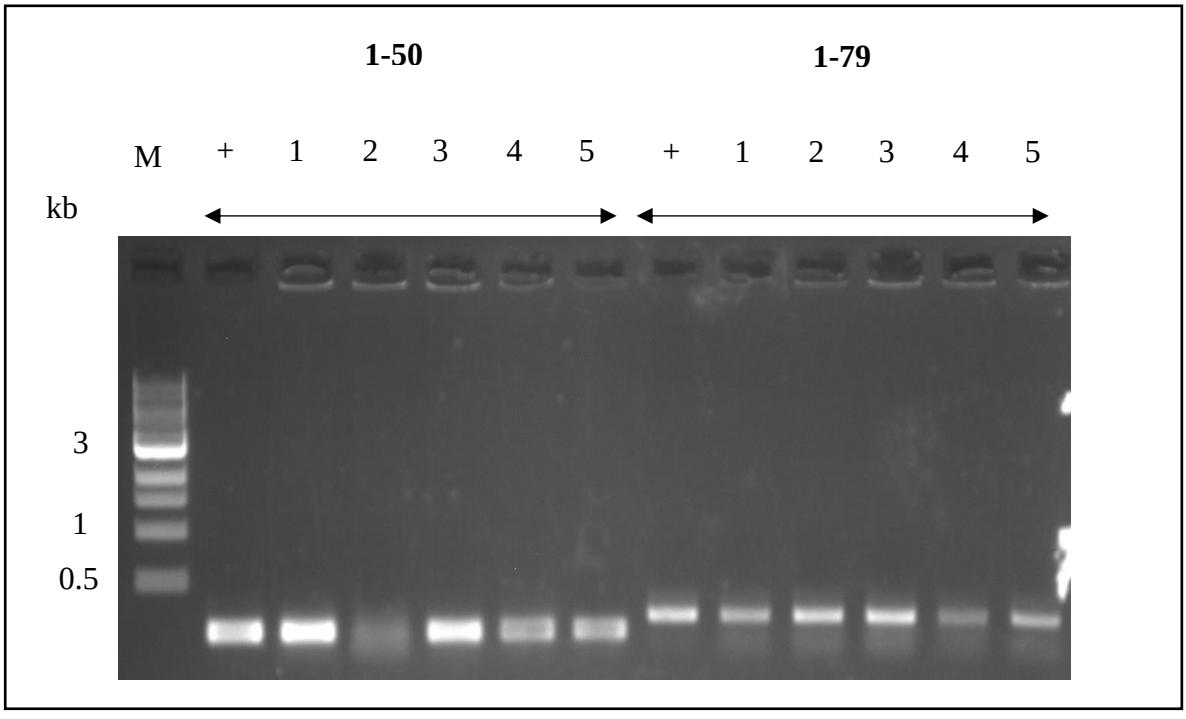


Figure 15. Colony PCR reactions for both LARP4A 1-50 (Theoretical size 150bp) and 1-79 (Theoretical size 237bp). *The first lane for each construct is a positive control using the original vector.*

The positive control for both 1-50 and 1-79 shows the PCR reaction with the plasmid pCMV2 FLAG that contains the full-length LARP4A sequence, using the same forward and reverse primers. From the positive control we can deduce the correct size of the insert. PCR 1, 3, 4 and 5 for LARP4A 1-50 showed positive results, while PCR 2 shows a smeared band which might be caused by the addition of an excess amount of DNA (Roux, 2009). For 1-79 all reactions seemed to be positive, they were all represented as single bands that match the size of the positive control. Therefore colony 3 for LARP4A 1-50 and colony 4 for LARP4A 1-79 were chosen and their DNA verified by sequencing, these were used subsequently for protein expression and purification.

3.2.2 Shortening LARP4B NTD 1-339 to 1-328

To test and study the RNA-binding capabilities and interactions of LARP4B with its targets, the constructs of LARP4B NTD and La-module were originally cloned by Emily Baldwin and Dr. Isabel Cruz Gallardo to end at residue 339, based on a first prediction of RRM domain boundaries. However, these mutants suffered degradation during purification (IMAC and subsequent steps). Most of the constructs ending in residue 339 degraded at the C-terminus to a shorter version as seen in the SDS-PAGE gel, it is a C-terminus degradation since as they bind to IMAC they must include an intact His tag and therefore an intact N-terminus. By sequence alignment of the N-terminus of LARP4A and 4B using Clustal Omega in UniProt portal (Figure 16) (http://www.uniprot.org/align/), LARP4B was chosen to be 'shortened' to residue 328, which should correspond to residue 287 in human LARP4A NTD for which there were no degradation problems (Cruz-Gallardo *et al.*, 2019) The constructs of LARP4B NTD

was therefore designed to terminate at residue 328 instead of 339 with the purpose of minimizing the degradation problem.

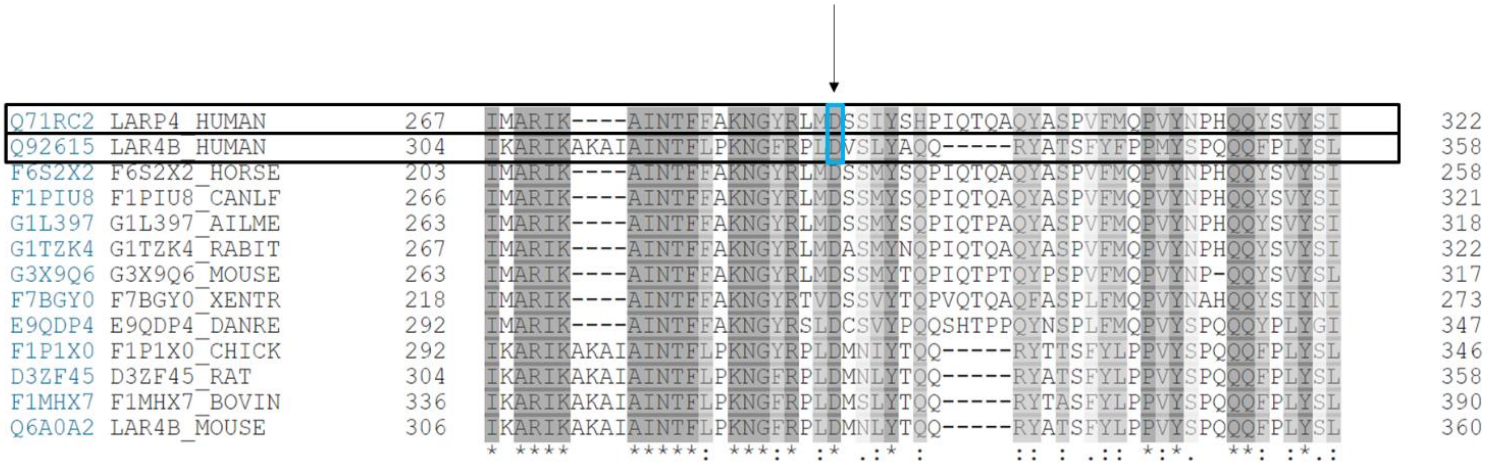


***Figure 16. Sequence alignment of LARP4A and 4B in different organisms**, with focus on the area where the NTD terminates in, residue 328 in LARP4B is highlighted by the arrow.*

Traditional cloning was performed to obtain His and GST tagged LARP4B 1-328 into an empty pRSF Duet-1 vector using LARP4B NTD 1-339 as a template. The insert sequence was generated using PCR and primers (Table 1) that were chosen to amplify residues containing the His-tag, GST-tag and LARP4B 1-328 (See Figure 17)

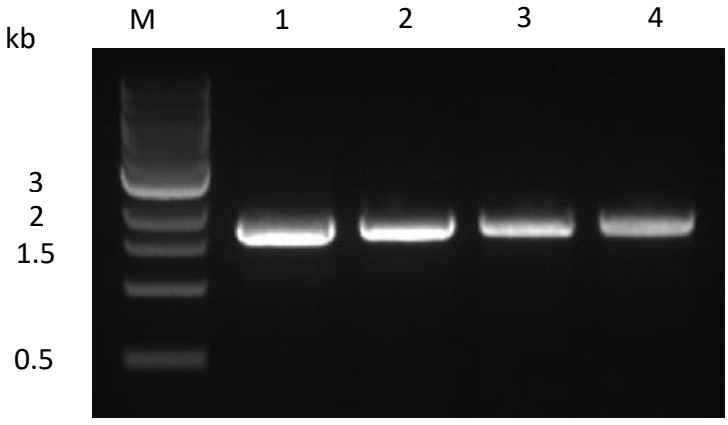


***Figure 17. PCR showing the Insert generation of LARP4B 1-328.** A total of four PCR reactions were done for the construct, (Theoretical size 1.7kb) M:Marker*

The four PCR reactions generated inserts that were all the correct sizes as verified in the agarose gel. They were pooled, and restriction enzyme digest was done on both the target vector pRSF Duet-1 and the pooled inserts and were ligated and transformed into competent cells. There were a high number of colonies after ligation, so using the original primers, three random colonies were chosen to undergo colony PCR reactions to check if the final vector contained the insert (See Figure 18). A positive control was included using the original vector subjected to PCR using the same forward and reverse primers. Colonies 1 and 2 were positive, but colony 3 appears to be slightly lower than the positive control. Colony 2 was selected for sequencing, revealing that it contains the His-tag, GST-tag and LARP4B ending 1-328 instead of 339.

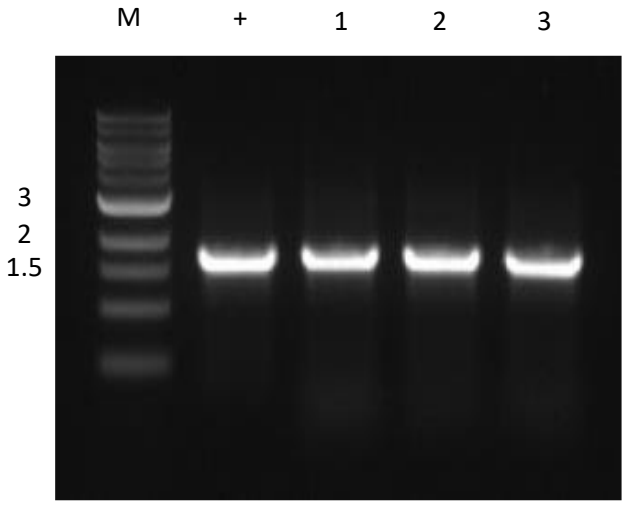


Figure 18. Colony PCR reactions for LARP4B 1-328 *(Theoretical size 1.7kb). The first lane is a positive control using the original vector. M:Marker*

3.2.3 Constructs obtained using mutagenesis

In this study, several alanine point mutants were generated using the commercial Q5 site-directed mutagenesis kit which is routinely used to point mutate and thus identify the contribution of a specific residue to the activity of the protein of interest. Alanine has a

non-bulky and relatively inert methyl side-chain that can be used to eliminate the side-chain interactions while not affecting the conformation of the main chain so that the native protein structure can be preserved (Morrison and Weiss, 2001).

The plasmids were sent to sequencing by *Eurofins genomics* to confirm the desired mutation. In addition to point mutation several deletion mutants were also achieved using this method (refer to Table 2 for the primers used for each mutant). The mutants generated using this strategy were LARP4B La-module 1-328, LARP4B L56AW63A 1-328, LARP4B 40-328, LARP4B 71-328, LARP4B 95-328, LARP4B 119-328, LARP4B T163A, F166A, C167A, D176A, Y178A and L197A in the context of both NTD and the La-module.

3.3 LARP4A constructs – Purification

3.3.1 LARP4A 1-50 and 1-79

After the LARP4A 1-50 and 1-79 mutants had been cloned, the protein constructs were expressed in *E. coli* Rosetta II cells grown in LB medium with 18°C overnight IPTG induction. Figure 19 shows the summary of LARP4A 1-50 expression and purification.

LARP4A 1-50 in detail:

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKV
SDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLD
MEDNDIIEAHREQIGGSMLLFVEQVASKGTGLNPNAKVWQEIAPGNTDATPVT
HGTESSWHEIAATS

(His-Tag) (SUMO-Tag) (Construct)

Molecular Weight (Da): 19433.62

Extinction Coefficient (M^{-1} cm^{-1}): 12490

Theoretical pI: 5.94

Molecular Weight post-cleavage (Da): 5292.87

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 11000

Theoretical pI post-cleavage: 4.92

Figure 19A shows the expression of LARP4A 1-50. The protein that appears overexpressed in the after-induction lane runs with a larger MW than expected on the SDS PAGE, given that the theoretical size is 19 kDa, whereas an overexpression band of ~25 kDa is seen. RNA binding proteins are known to run slightly higher than their theoretical sizes because of their positive charges, also migration of the protein on SDS-PAGE can be influenced by the charge and mass of the molecule. It could very well be that the protein is folded in such a way that it does not migrate as you would expect (Al-Tubuly, 2000).

LARP4A 1-50 was purified using the ÄKTA system, using Immobilized Metal Affinity Chromatography (IMAC) to separate the native proteins of the host cell from our protein of study. This is achieved by a HisTrap FF column, which contains Ni^{2+} ions that interact with the histidine residues in the His-tag and is eluted by a gradient of imidazole. The elute is collected in a 96-well collector and the peaks of the chromatogram (Figure 19D) were collected and different samples were run on an SDS-PAGE gel (Figure 19C). The protein of interest is collected and pooled together for further purification.

Comparing the input with the flowthrough (Figure 19C) we can see that around 25 kDa there is a band that is present in the input which correspond to LARP4A 1-50 protein and

it is much fainter in the flowthrough fraction, indicating that the column has bound most of the protein. During elution there is an *E. coli* contaminant that copurify around 70 kDa, it is a Hsp70 molecular chaperone (DnaK) that tend to be seen in most Nickel purifications (Rial and Ceccarelli, 2002) and would be ideally eliminated in further purifications such as ionic exchange. But since LARP4A 1-50 appears strongest in fractions 4-11, which also contains the contaminant, it could not be avoided since they are present in the same fractions. It is important to note that there is some degradation appearing around 20 kDa during this stage which could correspond to the SUMO tag alone, from previous purifications involving SUMO it also seems to appear. The sample was dialysed overnight with Nickel dialysis buffer, and then further purified using a HiTrap DEAE Sepharose FF column (Figure 19E, F). The DEAE column is an anion exchanger, used to remove any nucleic acids that are bound to the protein since they have a different charge than our protein. The resin will bind to anything that is negatively charged (more details in methods) and the peak of the elution fraction as shown in Figure 19E is ran on an SDS-PAGE gel in Figure 19F. Lanes 4 and 5 remain blank since nucleic acids are not stained in an SDS-PAGE gel; however, using a nano-drop these elution fractions have a high signal at 260 nm (data not shown) which indicates they contain nucleic acids. The protein was collected in the flowthrough and spin concentrated after dialysis in the final protein buffer of 20 mM Tris pH 7.25, 100 mM KCl, 0.2 mM EDTA, 1 mM DTT, to 100-200 μM. The final concentrated protein is shown in Figure 19G, The problem of the SUMO-tag degradation persists, even after the DEAE column, and since the SUMO tag and LARP4A 1-50 are very close in size, Size Exclusion Chromatography (SEC) which separates proteins with their molecular weight could not be a viable strategy to separate the protein from the spontaneously cleaved SUMO tag.

However, the SUMO tag has the benefit of producing large amounts of soluble protein (Müller *et al.*, 2001), but it should be removed because it might alter the protein's

behaviour in EMSAs and affect the Circular Dichroism signature. The tag was removed after the DEAE purification by the addition of SUMO Protease, also known as Ulp, which is a recombinant fragment of ULP1 (Ubl-specific protease 1) from *Saccharomyces cerevisiae*. It is highly specific for the SUMO protein fusion, recognizing the tertiary structure of SUMO rather than an amino acid sequence, and ULP1 cleaves after the C-terminal glycine at the end of the SUMO sequence. (Hickey, Wilson and Hochstrasser, 2012). The protease was added in the ratio of 0.05mg Protease:0.5mg protein and incubated for 5 hours in 4°C. After digestion, IMAC was used to separate the His-SUMO tag from the cleaved protein. The results of the cleavage are shown in Figure 19H, where a single band is present around 13kDa, assigned to LARP4A 1-50 protein.

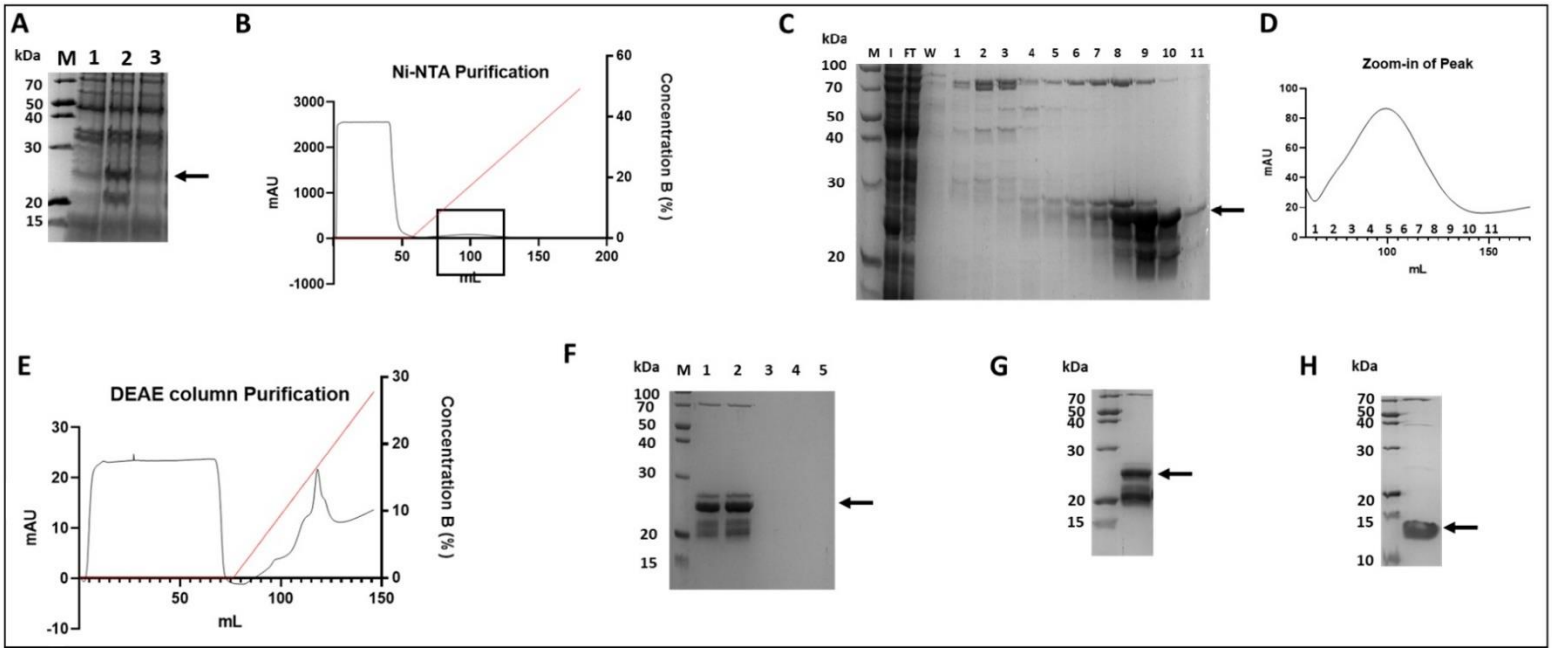


Figure 19. LARP4A 1-50 expression and purification. *[A]: Expression of LARP4A 1-50 protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane F: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification. [E]: Chromatogram of DEAE column purification. [F]: DEAE column purification, lane M: protein ladder; lane 1: DEAE input; lane 2: DEAE flowthrough; lane 3-5: DEAE elute peak [G]: Final protein after concentration and pre-SUMO cleavage [H]: Final protein after SUMO protease cleavage (note that the protein appears larger than the theoretical MW).*

LARP4A 1-79 is purified using the same methodology and procedure as LARP4A 1-50, and the Nickel column purification is shown below in Figure 20. They have similar profiles in terms of the number of contaminants and degradation. In the final gel (Figure

20H) the contaminant at 70 kDa still persists at a low level and the protein shows to be a single band at 20 kDa.

LARP4A 1-79 in detail:

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSS
EIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHR
EQIGGSMLLFVEQVASKGTGLNPNAKVWQEIAPGNTDATPVTHGTESSWHEIAATSGA
HPEGNAELSEDICKEYEVMYSSSCETT

(His-Tag) (SUMO-Tag) (Construct)

Molecular Weight (Da): 22595.99

Extinction Coefficient (M^{-1} cm^{-1}): 15470

Theoretical pI: 5.50

Molecular Weight post-cleavage (Da): 8455.24

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 13980

Theoretical pI post-cleavage: 4.39

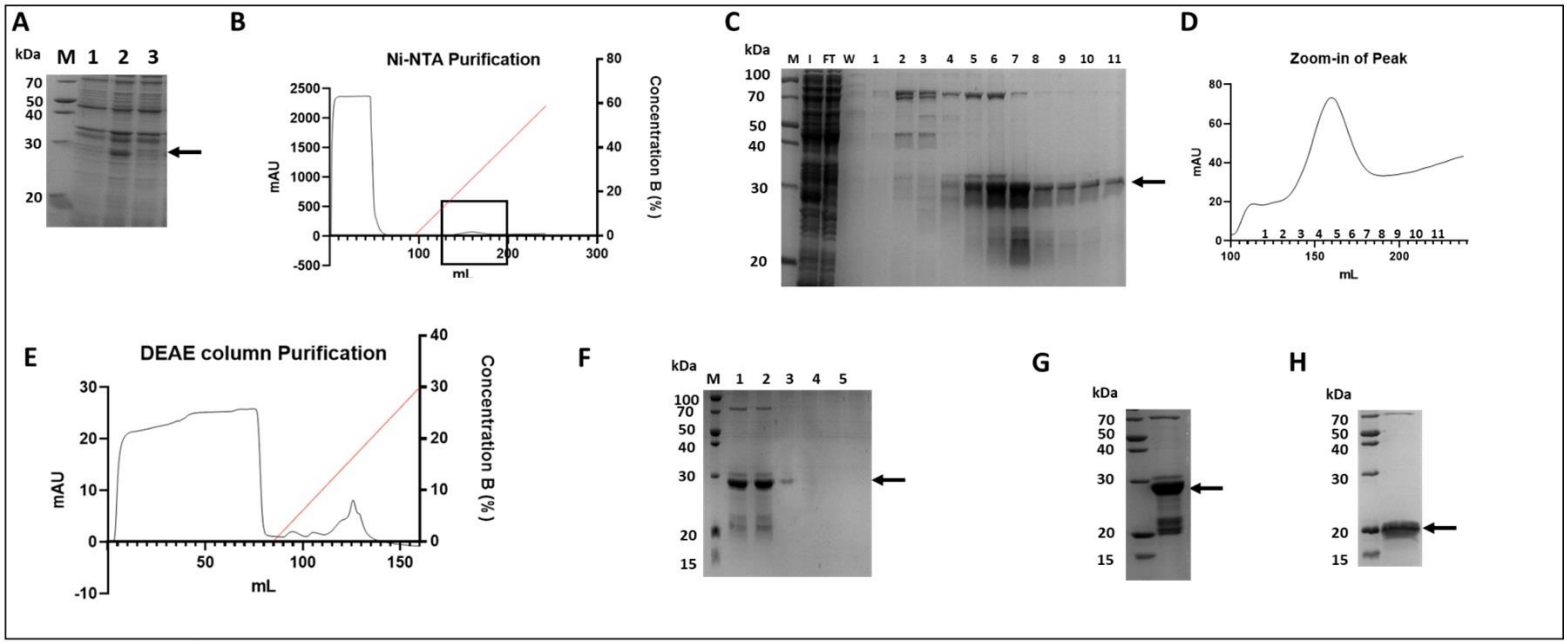


Figure 20. LARP4A 1-79 expression and purification. *[A]: Expression of LARP4A 1-79 protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane F: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification. [E]: Chromatogram of DEAE column purification. [F]: DEAE column purification, lane M: protein ladder; lane 1: DEAE input; lane 2: DEAE flowthrough; lane 3-5: DEAE elute peak. [G]: Final protein [H]: Final protein after SUMO protease cleavage*

3.3.2 LARP4A NTD, NTR and La-module

The expression and purification of LARP4A NTD, (Figure 21) NTR (Figure 22) and La-module (Figure 23) was performed to act as controls for the EMSA experiment, these constructs were tested to see if they behaved in the same way as described in Cruz-Gallardo et al., 2019.

LARP4A NTD in detail:

MGSSHHHHHHSQDPENLYFQSMLLFVEQVASKGTGLNPNAKVWQEIAPGNTDATPVT
HGTESSWHEIAATSGAHPEGNAELSEDICKEYEVMYSSSCETTRNTTGIEESTDGMILGP
EDLSYQIYDVSGESNSAVSTEDLKECLKKQLEFCFSRENLSKDLYLISQMDSDQFIPIWT
VANMEEIKKLTTDPDLILEVLRSSPMVQVDEKGEKVRPSHKRCIVILREIPETTPIEEVKG
LFKSENCPKVISCEFAHNSNWYITFQSDTDAQQAFKYLREEVKTFQGKPIMARIKAINTF
FAKNGYRLMD

(His-Tag) (TEV-protease cleavage site) (Construct)

Molecular Weight (Da): 34753.99

Extinction Coefficient (M^{-1} cm^{-1}): 35410

Theoretical pI: 4.98

Molecular Weight post-cleavage (Da): 32346.47

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 33920

Theoretical pI post-cleavage: 4.78

Figure 21A shows the overexpression of the recombinant protein LARP4A NTD. In the 'before induction' lane there is no visible overexpression, but after IPTG induction the recombinant protein shows up in both the insoluble and soluble lane, the band appears just under 40 kDa which matches the theoretical molecular weight which is 34 kDa. After Ni-NTA purification, the peak of the chromatogram is shown in Figure 21B and 21D, which shows the UV absorption in the 280 nm. Fractions were run on SDS-PAGE to follow the progress of the Ni-NTA purification. LARP4A NTD contains a small His-Tag with a TEV protease cleavage site at a size of 2.4 KDa, and it is cleaved after the Ni-NTA purification during dialysis overnight with the addition of TEV protease as described in the Methods. Figure 21F lane 2 shows the cleaved protein at 40 kDa and the TEV protease at 30 kDa, the His-tagged protease and non-cleaved protein are collected in the elution of the second Ni-NTA and the flowthrough is collected for further purification in the DEAE column, removing the nucleic acids.

After another round of purification using the DEAE column, nucleic acids contaminants were separated from the protein. The protein was spin concentrated and flash frozen in liquid nitrogen until needed. The final cleaved protein is visible just below 40 kDa, as shown in Figure 21G.

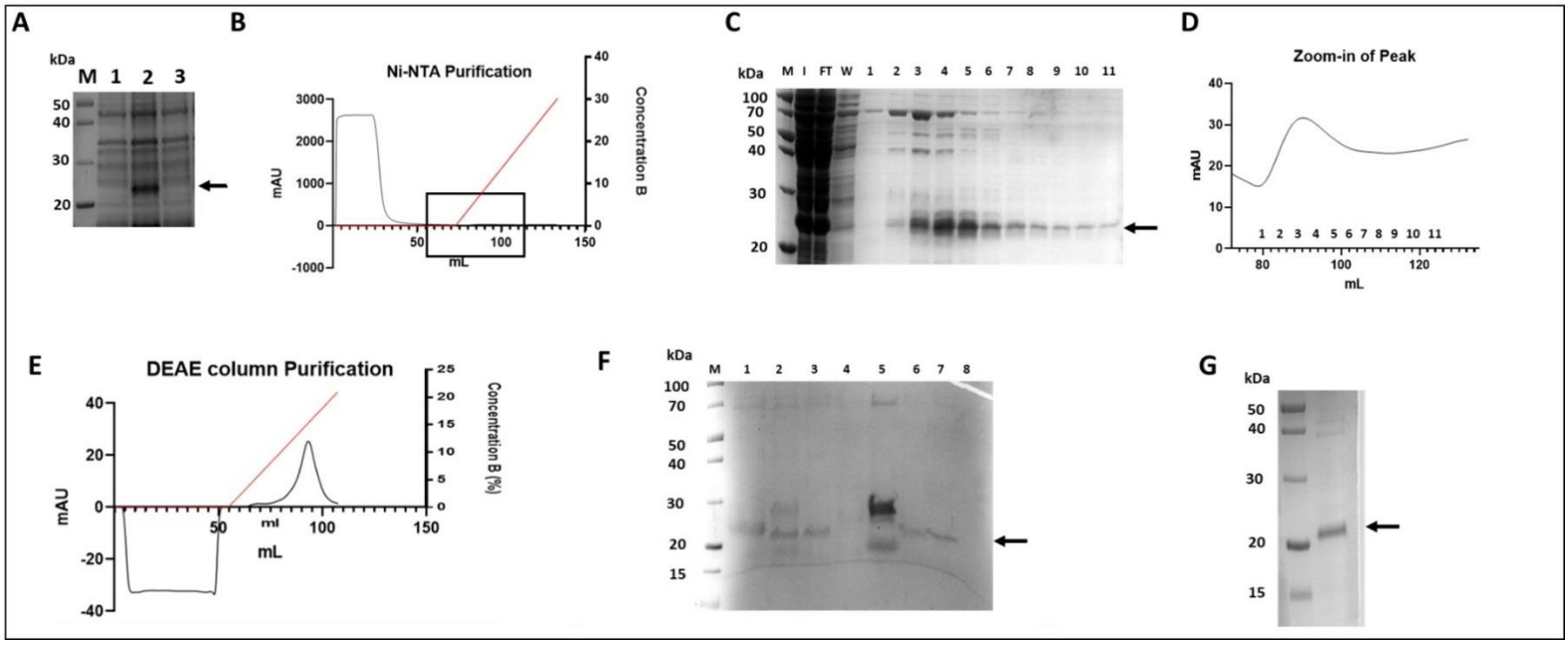


Figure 21. LARP4A NTD expression and purification. *[A]: Expression of LARP4A NTD protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of DEAE column purification. [F]: TEV protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before TEV cleavage; lane 2: after TEV cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: second Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8: DEAE wash; lane 9-12: elute fractions. [G]: Final protein*

LARP4A NTR in detail:

MGSSHHHHHHSQDPENLYFQSMLLFVEQVASKGTGLNPNAKVWQEIAPGNTDATPVT
HGTESSWHEIAATSGAHPEGNAELSEDICKEYEVMYSSSCETTRNTTGIEESTDGMILGP
EDLSYQIYDVSGESN

(His-Tag) (TEV-protease cleavage site) (Construct)

Molecular Weight (Da): 14336.45

Extinction Coefficient (M^{-1} cm^{-1}): 18450

Theoretical pI: 4.54

Molecular Weight post-cleavage (Da): 16015.00

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 12490

Theoretical pI post-cleavage: 4.05

For LARP4A NTR, the purification workflow is the same as LARP4A NTD, the pI of NTR after the protease cleavage is 4.05, which is similar to NTD, at 4.98, hence DEAE chromatography is chosen as a viable strategy to obtain the protein in the flowthrough during the IEX step. The purified protein is shown in Figure 22G, the protein seems to be pure as only one dominant band is present, representing LARP4B NTR shown just above 20 kDa.

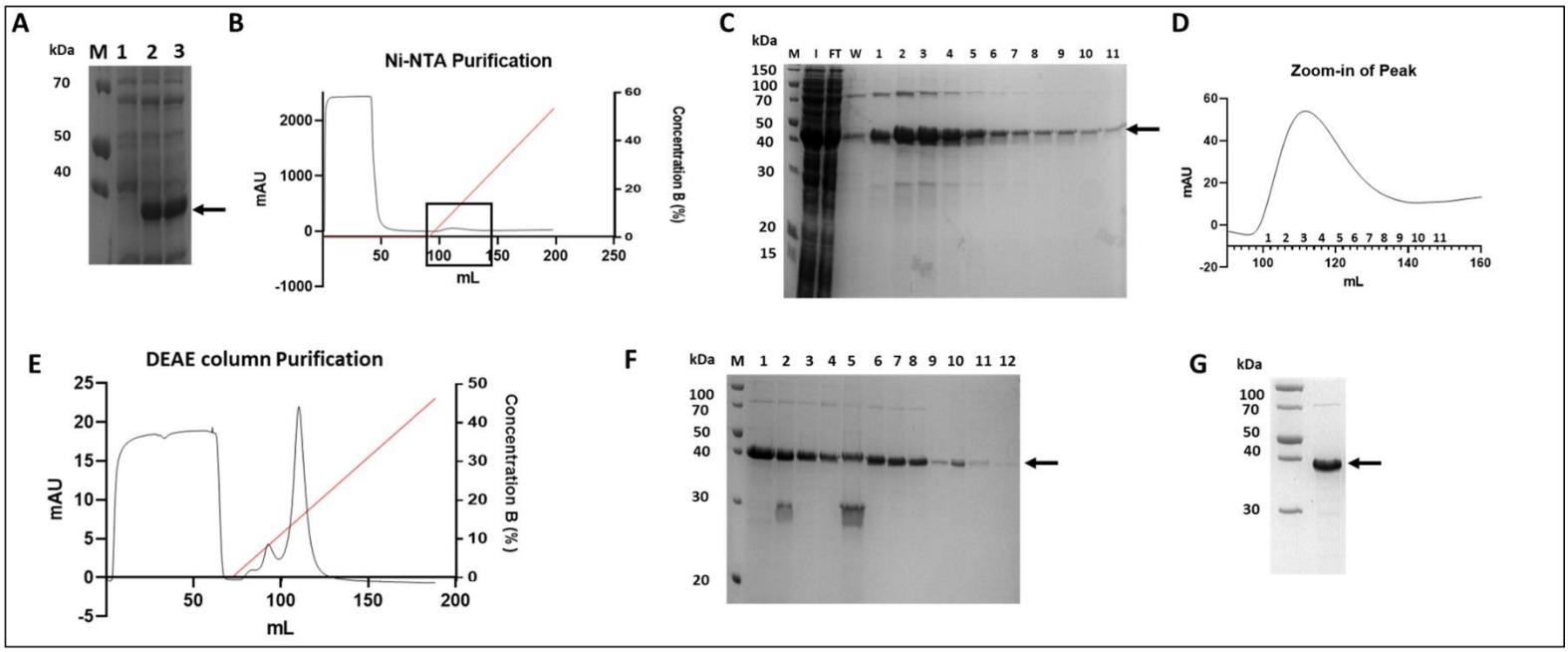


Figure 22. LARP4A NTR expression and purification. *[A]: Expression of LARP4A NTR protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of DEAE column purification. [F]: TEV protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before TEV cleavage; lane 2: after TEV cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: second Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8: elute peak. [G]: Final protein*

LARP4A La-module purification in Figure 23 shows similar purification workflow with a small difference. The protein undergoes Nickel IMAC, and then is dialysed with TEV protease for tag removal just like NTD and NTR, but since the pI of the La-module protein is 5.83 (estimated with *Protparam*), (compared to the pI of NTD: 4.78 and NTR: 4.05), we discovered that a Heparin column is required in order to separate the negatively charged nucleic acids in the flowthrough of the column whereas the protein of interest is captured from the elution (Figure 23F, lanes 9-12). The La-module contains less negatively charged amino acids compared to NTD or NTR, therefore will bind better to the negatively charged heparin column. It is noted that if the pI is >5, a heparin based IEX approach is more suitable for the purification. After pooling the elution fractions of the Heparin column containing LARP4A La-module, they are dialysed overnight with the final dialysis buffer, the protein was spin concentrated and flash frozen in liquid nitrogen until needed. The protein seems to be pure as only one dominant band is present, representing LARP4A La-module shown just below 20 kDa (Figure 23G).

LARP4A La-module in detail:

MGSSHHHHHHSQDPENLYFQSNSAVSTEDLKECLKKQLEFCFSRENLSKDLYLISQMD
SDQFIPIWTVANMEEIKKLTTDPDLILEVLRSSPMVQVDEKGEKVRPSHKRCIVILREIPE
TTPIEEVKGLFKSENCPKVISCEFAHNSNWYITFQSDTDAQQAFKYLREEVKTFQGKPIM
ARIKAINTFFAKNGYRLMD

(His-Tag) (TEV-protease cleavage site) (Construct)

Molecular Weight (Da): 22957.19

Extinction Coefficient (M^{-1} cm^{-1}): 18450

Theoretical pI: 6.07

Molecular Weight post-cleavage (Da): 20549.66

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 16960

Theoretical pI post-cleavage: 5.83

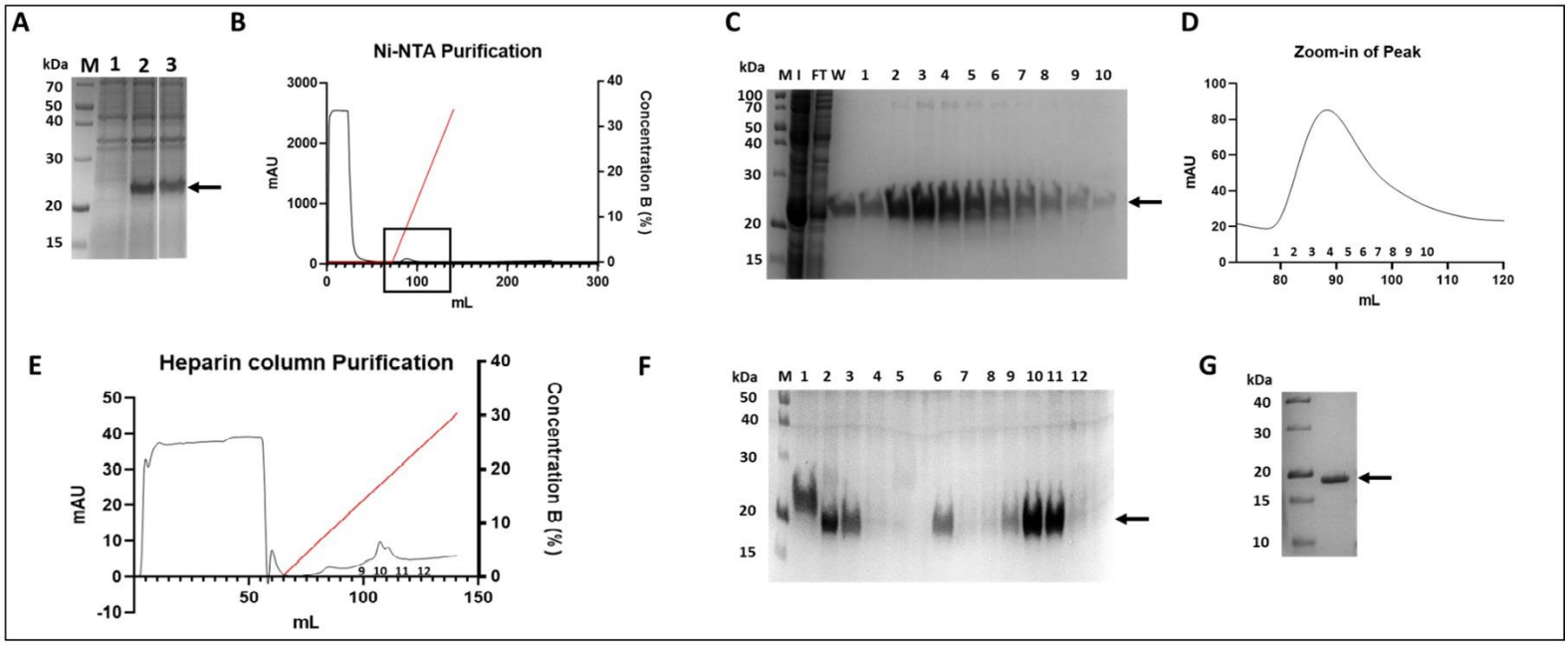


Figure 23. LARP4A La-module expression and purification. *[A]: Expression of LARP4A La-module protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-10: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of Heparin column purification. [F]: TEV protease cleavage and Heparin column purification, lane M: protein ladder; lane 1: before TEV cleavage; lane 2: after TEV cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: second Ni-NTA elute; lane 6: Heparin column input; lane 7: Heparin flowthrough; lane 8: wash; lane 9-12: elute fractions. [G]: Final protein*

3.3.3 LARP4A L15AW22A

The double mutant L15AW22A contains the mutation of 2 key amino acids in the PAM2w domain of LARP4A, which are known to abrogate the interaction between LARP4A PAM2w motif and the MLLE domain of PABPC1 (Yang *et al.*, 2011) by disrupting the key hydrophobic interactions. It is also known to disrupt the binding of LARP4A and oligoA RNA (Cruz-Gallardo *et al.*, 2019). This construct is purified to act as a positive control in pulldown assays where the interaction with its known protein partner PABPC1-MLLE is compared with the wild-type LARP4A NTD. The purification of L15AW22A is shown in Figure 24.

LARP4A L15AW22A in detail:

MGSSHHHHHHSQDPENLYFQSMLLFVEQVASKGTGANPNAKVAQEIAPGNTDATPVT
HGTESSWHEIAATSGAHPEGNAELSEDICKEYEVMYSSSCETTRNTTGIEESTDGMILGP
EDLSYQIYDVSGESNSAVSTEDLKECLKKQLEFCFSRENLSKDLYLISQMDSDQFIPIWT
VANMEEIKKLTTDPDLILEVLRSSPMVQVDEKGEKVRPSHKRCIVILREIPETTPIEEVKG
LFKSENCPKVISCEFAHNSNWYITFQSDTDAQQAFKYLREEVKTFQGKPIMARIKAINTF
FAKNGYRLMD

(His-Tag) (TEV-protease cleavage site) (Construct)

Molecular Weight (Da): 34596.78

Extinction Coefficient (M^{-1} cm^{-1}): 29910

Theoretical pI: 4.98

Molecular Weight post-cleavage (Da): 32189.25

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 28420

Theoretical pI post-cleavage: 4.78

For L15AW22A, the purification workflow is the same as LARP4A NTD and NTR, the pI of L15AW22A after the protease cleavage is 4.78, hence DEAE chromatography is chosen as a viable strategy to obtain the protein in the flowthrough during the IEX step. The purified protein is shown in Figure 24G, the protein seems to be pure as only one dominant band is present, representing LARP4B L15AW22A shown just below 40 kDa.

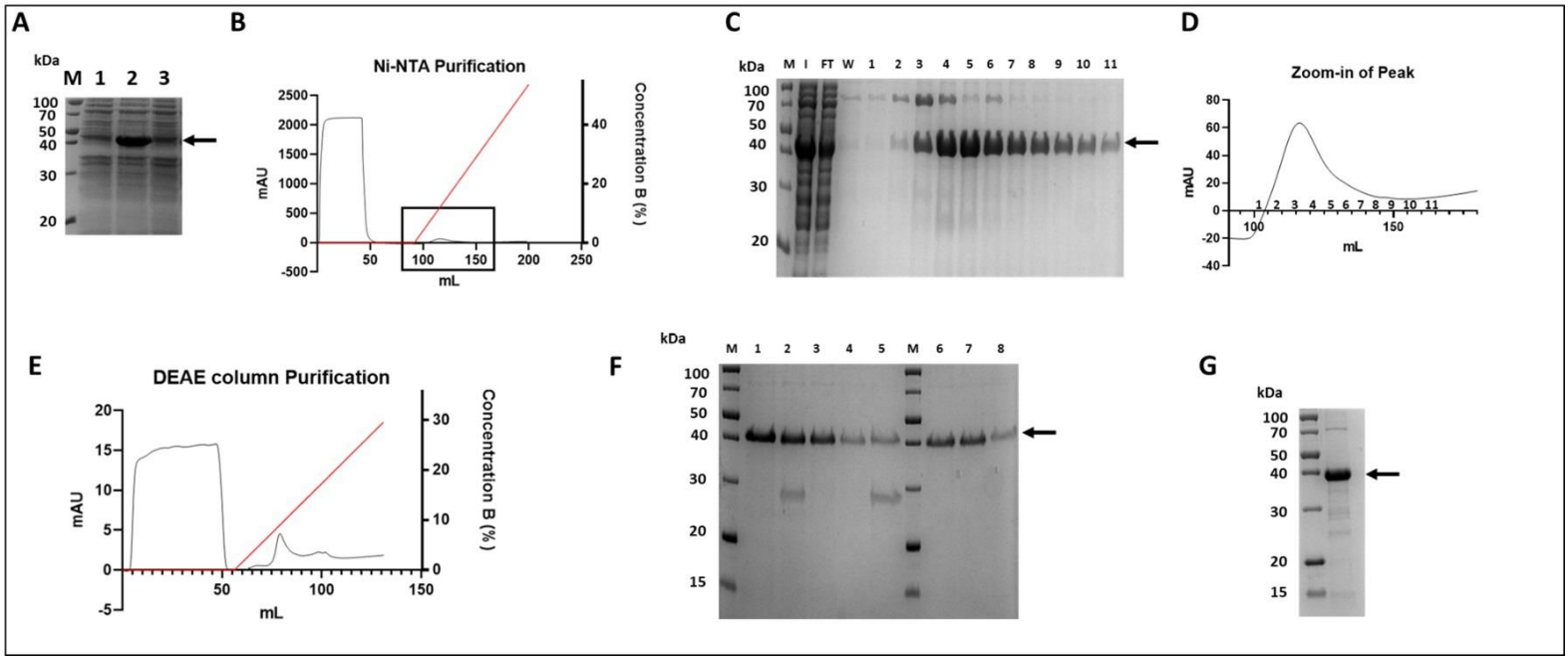


Figure 24. LARP4A L15AW22A expression and purification. *[A]: Expression of LARP4A L15AW22A protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of DEAE column purification. [F]: TEV protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before TEV cleavage; lane 2: after TEV cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: second Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8: DEAE elution peak. [G]: Final protein*

3.4 LARP4B constructs – Purification

3.4.1 LARP4B NTD, NTR and La-module

LARP4B 1-339 contains a Histidine tag followed by a GST tag which is 28 kDa, GST tagged LARP4B yielded purer final protein as compared to SUMO tagged LARP4B. This was cloned in pRSFDuet-1 vector by a former PhD in the Conte lab, Emily Baldwin, to include a 3C PreScission cleavage site. The 3C PreScission Protease is a fusion protein of human rhinovirus (HRV) 3C Protease (Ullah *et al.*, 2016). Our constructs contain a

HRV 3C protease cleavage site (LeuGluValLeuPheGln ↓ GlyPro, with the arrow indicating the cleavage position). Cleavage at this specific site allows the removal of the Histidine and GST tags which are relatively large and could interfere with our assays that are carried out after purification.

Figure 25A shows a clear overexpression below 70 kDa. However, looking at the SDS-PAGE from the Nickel column (Figure 25C), there seemed to be several bands close to the NTD that could be due to degradation, the protein may be degrading during the expression stage in the bacteria or during bacterial extraction.

LARP4B NTD 1-339 in detail:

MGSSHHHHHHSQDPNSSSSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYS
KDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFP
KLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSMTSDQDA
KVVAEPQTQRVQEGKDSAHLMNGPISQTTSQTSSIPPLSQVPATKVSELNPNAEVWGAPVLHLE
ASSAADGVSAAWEEVAGHHADRGPQGSDANGDGDQGHENAALPDPQESDPADMNALALGPSE
YDSLPENSETGGNESQPDSQEDPREVLKKTLEFCLSRENLASDMYLISQMDSDQYVPITTVANLD
HIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALFKGDNLPKFI
NCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPKNGFRPLDVS
LYAQQRYAT

(His-Tag) (GST-tag) (3C-protease cleavage site) (Construct)

Molecular Weight (Da): 65933.26

Extinction Coefficient (M^{-1} cm^{-1}): 71280

Theoretical pI: 5.00

Molecular Weight post-cleavage (Da): 37663.74

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 28420

Theoretical pI post-cleavage: 4.56

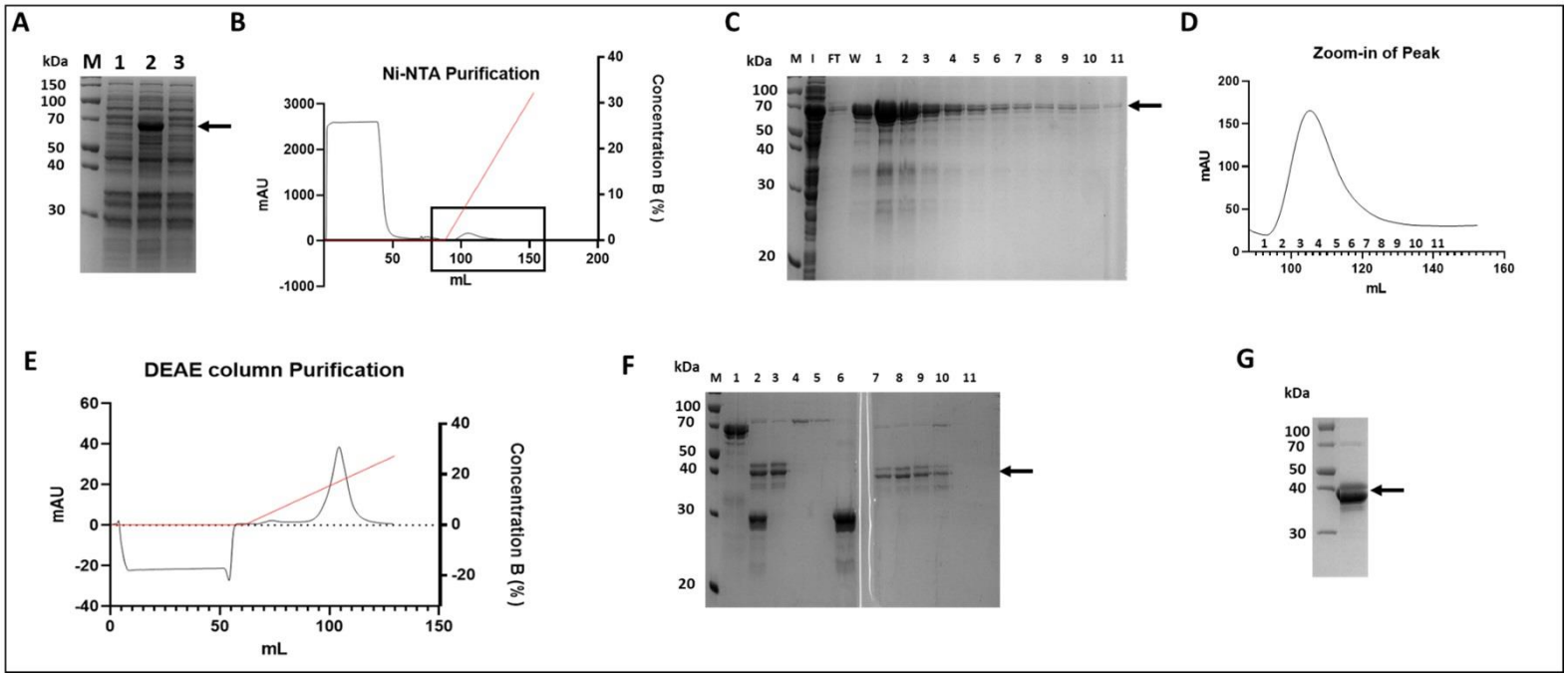


Figure 25. LARP4B NTD 1-339 expression and purification. *[A]: Expression of LARP4B NTD (1-339) protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of DEAE column purification. [F]: HRV 3C protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before 3C cleavage; lane 2: after 3C cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: second Ni-NTA wash 2; lane 6: Ni-NTA elute; lane 7: DEAE input; lane 8: DEAE flowthrough; lane 9: DEAE wash, lanes 10-11: elute fractions. [G]: Final protein*

The degradation issue became worse after each subsequent purification step, as in the SDS-PAGE showing the band after protease cleavage (Figure 25F, lane 2) there are 3 distinct bands around 40 kDa, and the same throughout the DEAE purification. The final protein gel shows a group of bands around 40 kDa, the degradation of 1-339 still persists in the final protein sample. To solve this issue, LARP4B NTD was shortened via traditional cloning to the 328 amino acid site (described previously). Figure 26 shows LARP4B NTD (1-328) expression and purification. When comparing it to the 1-339 construct, a clear difference can be seen, the protein seemed to be one distinct band throughout the purification. In both cases the protein is eluted during the Nickel purification, however more degradation can be seen for the 1-339 construct, in the SDS-

PAGE of the Nickel IMAC, there are bands around 50, 35 and 28 kDa, whereas the 328 construct, despite having degradation around 50 and 30 kDa, the main band corresponding to the actual 1-328 protein is much more prominent compared to the 339 protein.

Before cleavage, the protein with the His-GST tag is 66 kDa. After the addition of the protease overnight, the gel shows the cleaved protein slightly below 40 kDa, and the His-GST tag at 30 kDa, no tagged protein can be seen in the post-cleavage lane, which indicates that almost all the protein has been cleaved. The second IMAC nickel column is done to remove any protease that is in the mixture, the flowthrough and wash lanes contain the LARP4B NTD protein and is pooled to a DEAE column to further purify the nucleic acids from the protein.

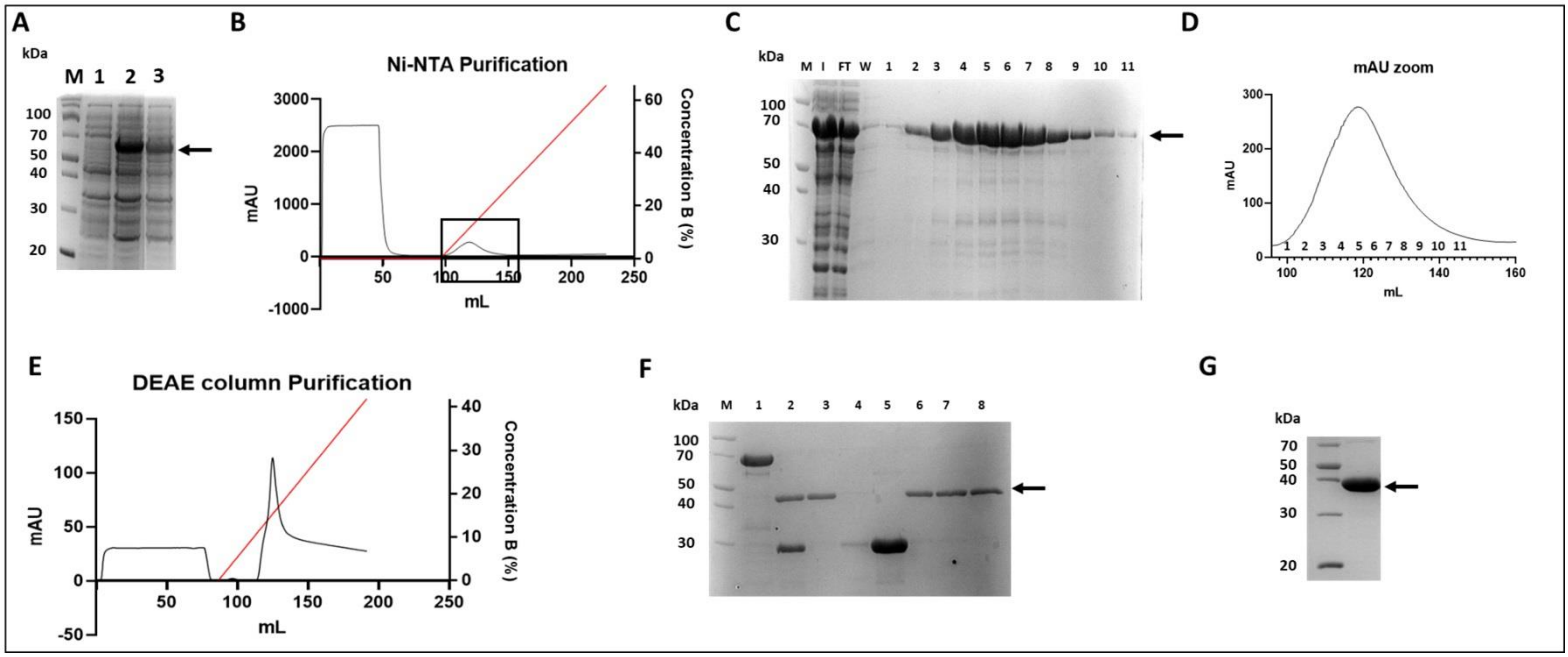


Figure 26. LARP4B NTD 1-328 expression and purification. *[A]: Expression of LARP4B NTD (1-328) protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of DEAE column purification. [F]: HRV 3C protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before 3C cleavage; lane 2: after 3C cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: second Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8: DEAE wash [G]: Final protein.*

After spin concentration, the two final constructs were compared on an SDS-PAGE gel, as shown in Figure 27, with the 1-328 construct appearing as a single band indicating protein purity, while most of the 1-339 construct has degraded to the to a band with a molecular weight which is compatible to that of the 1-328 construct. Hence the 328 construct was used for further studies.

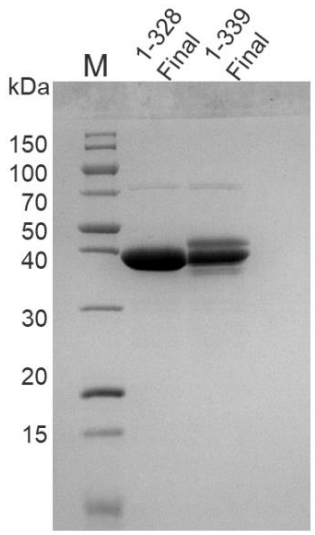


Figure 27. LARP4B NTD (1-339) shows degradation in the C-terminus. *SDS-PAGE analysis was used to compare final protein of LARP4B NTD ending in residue 328 and 339. The 1-328 lane shows a much purer protein band when compared to the 1-339 lane, which shows a significant amount of impurities and degradation slightly above and below the 40 kDa molecular weight.*

LARP4B NTD 1-328 in detail:

MGSSHHHHHHSQDPNSSSSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYS
KDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFP
KLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSMTSDQDA
KVVAEPQTQRVQEGKDSAHLMNGPISQTTSQTSSIPPLSQVPATKVSELNPNAEVWGAPVLHLE
ASSAADGVSAAWEEVAGHHADRGPQGSDANGDGDQGHENAALPDPQESDPADMNALALGPSE
YDSLPENSETGGNESQPDSQEDPREVLKKTLEFCLSRENLASDMYLISQMDSDQYVPITTVANLD
HIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALFKGDNLPKFI
NCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPKNGFRPLD

(His-Tag) (GST-tag) (3C-protease cleavage site) (Construct)

Molecular Weight (Da): 64651.82

Extinction Coefficient (M^{-1} cm^{-1}): 68300

Theoretical pI: 4.97

Molecular Weight post-cleavage (Da): 36382.31

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 25440

Theoretical pI post-cleavage: 4.52

For LARP4B NTR (Figure 28), the purification workflow is the same as LARP4A NTD, the pI of NTR after the protease cleavage is 4.05, which is similar to NTD, at 4.52, hence DEAE chromatography is chosen as a viable strategy to obtain the protein in the flowthrough during the IEX step. The purified protein is shown in Figure 28G, the protein seems to be pure as only one dominant band is present, representing LARP4B NTR shown just above 20 kDa.

LARP4B NTR in detail:

MGSSHHHHHHSQDPNSSSSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYS
KDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFP
KLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSMTSDQDA
KVVAEPQTQRVQEGKDSAHLMNGPISQTTSQTSSIPPLSQVPATKVSELNPNAEVWGAPVLHLE
ASSAADGVSAAWEEVAGHHADRGPQGSDANGDGDQGHENAALPDPQESDPADMNALALGPSE
YDSLPENSETGGNESQPD

(His-Tag) (GST-tag) (3C-protease cleavage site) (Construct)

Molecular Weight (Da): 44284.51

Extinction Coefficient (M^{-1} cm^{-1}): 55350

Theoretical pI: 4.92

Molecular Weight post-cleavage (Da): 16015.00

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 12490

Theoretical pI post-cleavage: 4.05

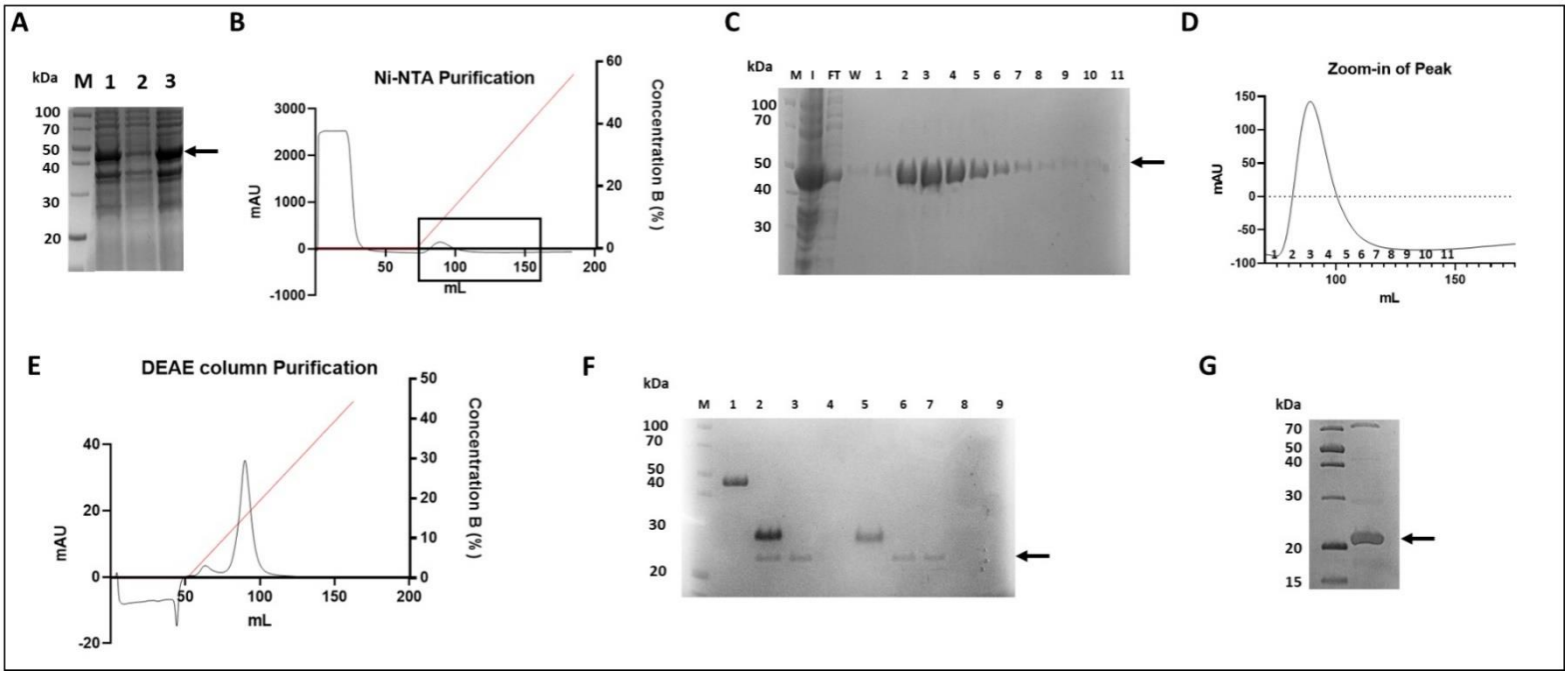


Figure 28. LARP4B NTR expression and purification. *[A]: Expression of LARP4B NTR protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of DEAE column purification. [F]: HRV 3C protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before 3C cleavage; lane 2: after 3C cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8-9: elute fractions. [G]: Final protein*

LARP4B La-module (Figure 29) contains a His-tag followed by a TEV-protease cleavage site. The La-module appears soluble indicated by a single band present at around 25 kDa (Figure 29A), the soluble fractions of the La-module bound to the Nickel IMAC column as an intense band around 25 kDa (Figure 29C), there are not many contaminations to be seen. After overnight dialysis with TEV-protease the La-module seemed to be cleaved, the size of the protein has been reduced to around 20 kDa on the SDS-PAGE gel, the protein mixture is subjected to a second nickel IMAC to get rid of the His-tagged protease and the uncleaved proteins. The flowthrough from the second nickel IMAC is then put

through a DEAE column to separate the protein from the nucleic acids, the flowthrough from the DEAE is dialysed overnight to the final buffer and concentrated. The concentrated protein seemed to be pure and without and contaminants and/or degradation (Figure 29G).

LARP4B La-module in detail:

MGSSHHHHHHSQDPNSSSENLYFQSDSQEDPREVLKKTLEFCLSRENLASDMYLISQMDSDQYV PITTVANLDHIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALF KGDNLPKFINCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPK NGFPLD

(His-Tag) (TEV-protease cleavage site) (Construct)

Molecular Weight (Da): 23370.36

Extinction Coefficient (M^{-1} cm^{-1}): 14440

Theoretical pI: 5.39

Molecular Weight post-cleavage (Da): 20587.50

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 12950

Theoretical pI post-cleavage: 5.00

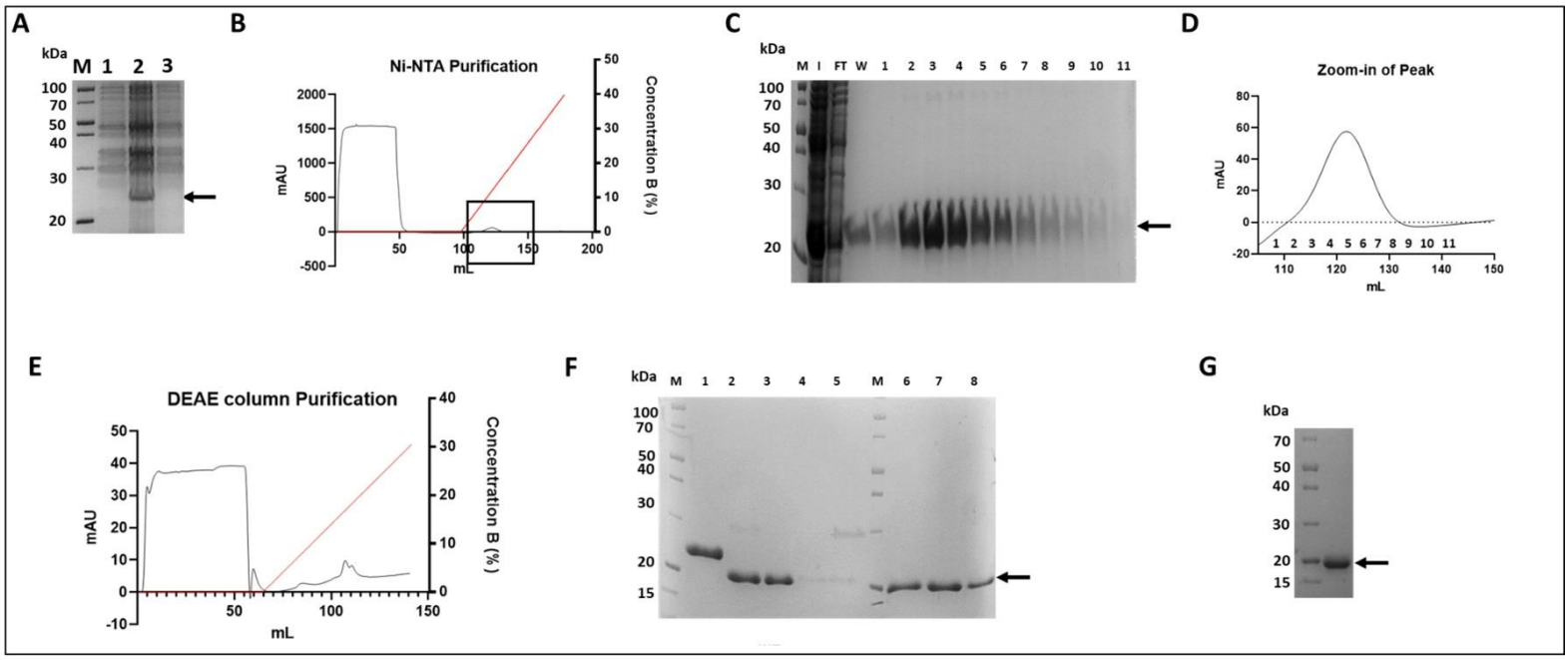


Figure 29. LARP4B La-module expression and purification. *[A]: Expression of LARP4B La-module protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of DEAE column purification. [F]: HRV 3C protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before TEV cleavage; lane 2: after TEV cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8: elution peak. [G]: Final protein*

3.4.2 LARP4B PAM2w mutant

The PAM2w domain of LARP4B is located within residues 56 to 63. The two conserved residues discussed above for LARP4A (L15 and W22) correspond to L56 and W63 in LARP4B. These two residues have been shown to affect LARP4B binding to PABP in a previous report (Schäffler *et al.*, 2010). We have generated a double mutant L56AW63A (in the context of 1-328) to assess the PAM2w ability to interact with RNA and PABP and compared its behaviour to LARP4A (Cruz-Gallardo *et al.*, 2019) The purification of LARP4B L56AW63A is shown in Figure 30.

LARP4B L56AW63A in detail:

MGSSHHHHHHSQDPNSSSSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE

LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYS

KDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFP

KLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSMTSDQDA

KVVAEPQTQRVQEGKDSAHLMNGPISQTTSQTSSIPPLSQVPATKVSEANPNAEVAGAPVLHLEA

SSAADGVSAAWEEVAGHHADRGPQGSDANGDGDQGHENAALPDPQESDPADMNALALGPSEY

DSLPENSETGGNESQPDSQEDPREVLKKTLEFCLSRENLASDMYLISQMDSDQYVPITTVANLDH

IKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALFKGDNLPKFIN

CEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPKNGFRPLD

(His-Tag) (GST-tag) (3C-protease cleavage site) (Construct)

Molecular Weight (Da): 64494.61

Extinction Coefficient (M^{-1} cm^{-1}): 62800

Theoretical pI: 4.97

Molecular Weight post-cleavage (Da): 36225.10

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 19940

Theoretical pI post-cleavage: 4.52

LARP4B L56AW63A (Figure 30) contains a His-GST-tag followed by a 3C-protease cleavage site. The protein appears soluble indicated by a single band present at around 70 kDa (Figure 30A), the soluble fractions of the protein bound to the Nickel IMAC column as one main band around 70 kDa, some degradation could be seen just below (Figure 30C). After overnight dialysis with 3C-protease LARP4B L56AW63A seemed to be cleaved, the size of the protein has been reduced to around 45 kDa on the SDS-PAGE gel, the protein mixture is subjected to a second nickel IMAC to get rid of the His-tagged protease, the uncleaved proteins and the His-GST tag. The flowthrough from the second nickel IMAC is then put through a DEAE column to separate the protein from the nucleic acids, the flowthrough from the DEAE is dialysed overnight to the final buffer and

concentrated. The concentrated protein seemed to have a main band at 45 kDa, and a couple below showing slight degradation (Figure 30G).

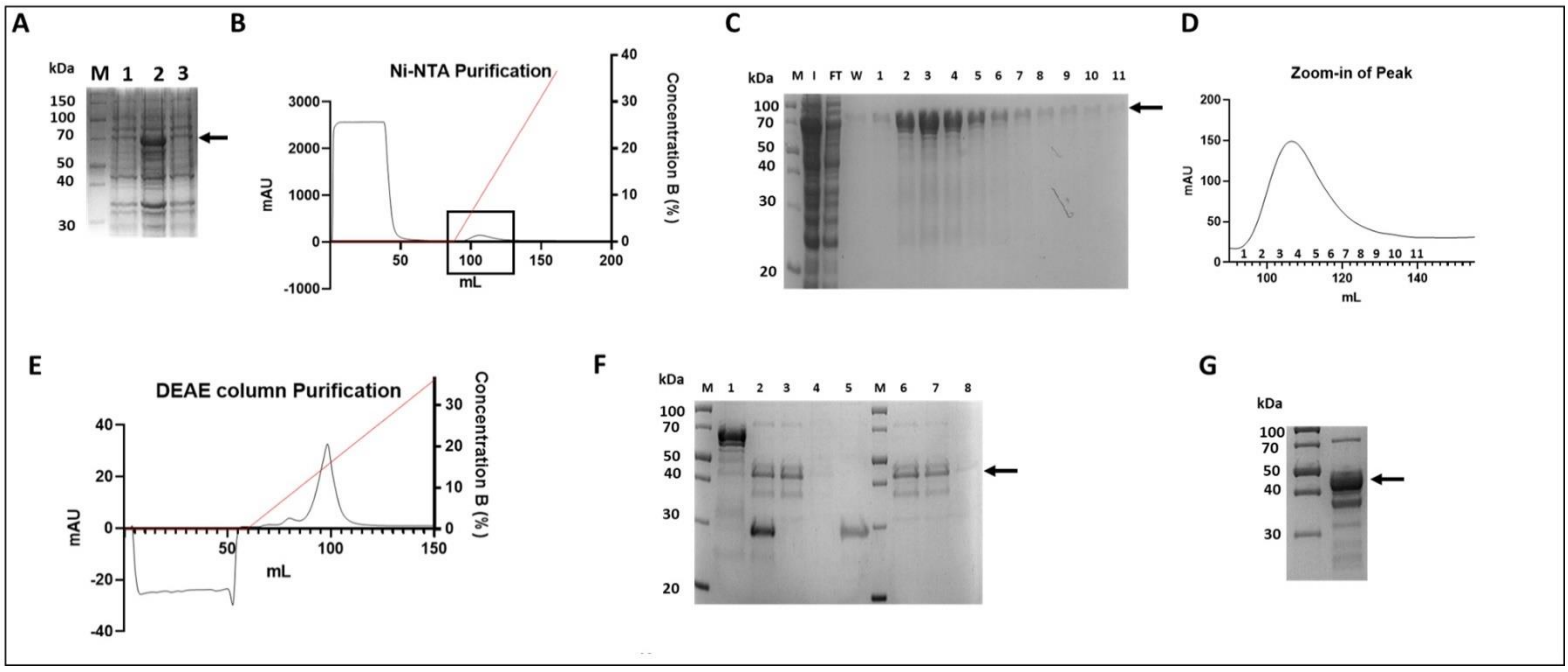


Figure 30. LARP4B L56AW63A expression and purification. *[A]: Expression of LARP4B L56AW63A protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of DEAE column purification. [F]: HRV 3C protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before 3C cleavage; lane 2: after 3C cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8: elution peak. [G]: Final protein*

3.4.3 LARP4B – N-terminal deletion mutants

N-terminal deletion mutants were designed and cloned by Neha Agrawal (*Conte Lab*) and purified by me using the same workflow as LARP4B NTD and NTR. The N-terminal truncation mutants (40-328, 71-328 and 95-328) were designed by sequence alignment to LARP4A (Cruz-Gallardo *et al.*, 2019) in the context of a comparative study of the RNA binding behaviour of LARP4A and LARP4B. In LARP4A the N-terminal regions had been shown to be involved in RNA recognition (Cruz-Gallardo *et al.*, 2019), and the aim in this study is to uncover the contribution of LARP4B N-terminal regions to RNA interaction (see Chapter 4). The mutant 40-328 was designed on the basis that the first 40 residues exist only in LARP4B tetrapod proteins (not LARP4A). Thus, the mutant 40-

328 would be important to assess the involvement of this region non-conserved across paralogs to RNA binding.

LARP4B 40-328 in detail:

MGSSHHHHHHSQDPNSSSSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYS
KDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFP
KLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSSSIPPLSQV
PATKVSELNPNAEVWGAPVLHLEASSAADGVSAAWEEVAGHHADRGPQGSDANGDGDQGHE
NAALPDPQESDPADMNALALGPSEYDSLPENSETGGNESQPDSQEDPREVLKKTLEFCLSRENLA
SDMYLISQMDSDQYVPITTVANLDHIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVI
LREISESTPVEEVEALFKGDNLPKFINCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIK
ARIKAKAIAINTFLPKNGFRPLD

(His-Tag) (GST-tag) (3C-protease cleavage site) (Construct)

Molecular Weight (Da): 60454.25

Extinction Coefficient (M^{-1} cm^{-1}): 68300

Theoretical pI: 4.98

Molecular Weight post-cleavage (Da): 32184.73

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 25440

Theoretical pI post-cleavage: 4.50

LARP4B 40-328 (Figure 31) contains a His-GST-tag followed by a 3C-protease cleavage site. The protein appears soluble indicated by a single band present at around 60 kDa (Figure 31A), the soluble fractions of the protein bound to the Nickel IMAC column as one main band around 60 kDa, there are not many contaminations to be seen (Figure 31C). After overnight dialysis with 3C-protease LARP4B 40-328 seemed to be cleaved, the size of the protein has been reduced to around 40 kDa on the SDS-PAGE gel, the protein mixture is subjected to a second nickel IMAC to get rid of the His-tagged protease, the uncleaved proteins and the tag. The flowthrough from the second nickel IMAC is then put through a DEAE column to separate the protein from the nucleic acids, the flowthrough from the DEAE is dialysed overnight to the final buffer and concentrated. The concentrated protein seemed to be pure and without and contaminants and/or degradation, at 40 kDa (Figure 31G).

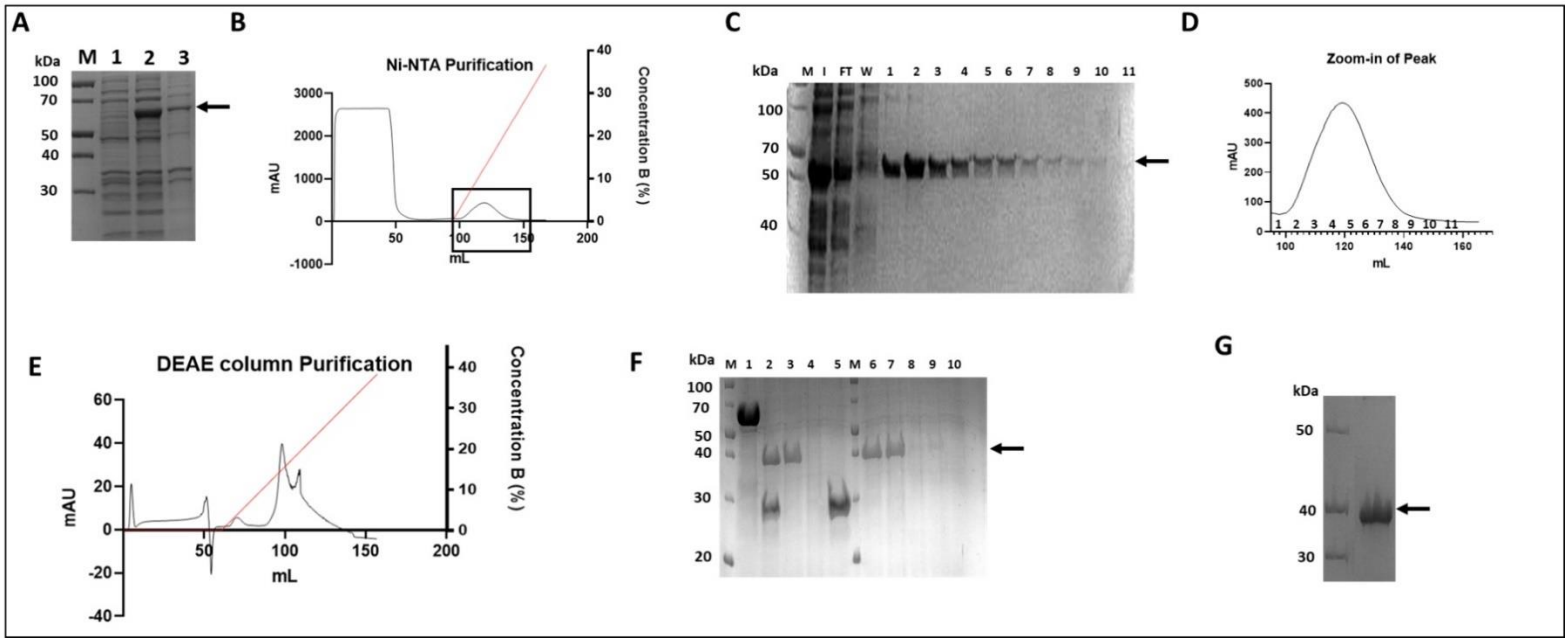


Figure 31. LARP4B 40-328 expression and purification*. [A]: Expression of LARP4B 40-328 protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of DEAE column purification. [F]: HRV 3C protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before 3C cleavage; lane 2: after 3C cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8-10: elution peak. [G]: Final protein*

LARP4B 71-328 (Figure 32) contains a His-GST-tag followed by a 3C-protease cleavage site. The protein appears soluble indicated by a single band present at around 55 kDa (Figure 32A), the soluble fractions of the protein bound to the Nickel IMAC column as one main band around 55 kDa, there is one band around 40 kDa that appears to be a contaminant (Figure 32C). After overnight dialysis with 3C-protease LARP4B 71-328 seemed to be cleaved, the size of the protein has been reduced to around 30 kDa on the SDS-PAGE gel, the protein mixture is subjected to a second nickel IMAC to get rid of the His-tagged protease and the uncleaved proteins. The flowthrough from the second nickel IMAC is then put through a DEAE column to separate the protein from the nucleic acids, the flowthrough from the DEAE is dialysed overnight to the final buffer and concentrated. The concentrated protein seemed to have little degradation, which might be from previous steps of purification, the main band of 71-328 is just under 30 kDa (Figure 32G).

LARP4B 71-328 in detail:

MGSSHHHHHHSQDPNSSSSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYS
KDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFP
KLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSEASSAADG
VSAAWEEVAGHHADRGPQGSDANGDGDQGHENAALPDPQESDPADMNALALGPSEYDSLPEN
SETGGNESQPDSQEDPREVLKKTLEFCLSRENLASDMYLISQMDSDQYVPITTVANLDHIKKLST
DVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALFKGDNLPKFINCEFAY
NDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPKNGFRPLD

(His-Tag) (GST-tag) (3C-protease cleavage site) (Construct)

Molecular Weight (Da): 57220.53

Extinction Coefficient (M^{-1} cm^{-1}): 62800

Theoretical pI: 4.97

Molecular Weight post-cleavage (Da): 28908.98

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 19440

Theoretical pI post-cleavage: 4.48

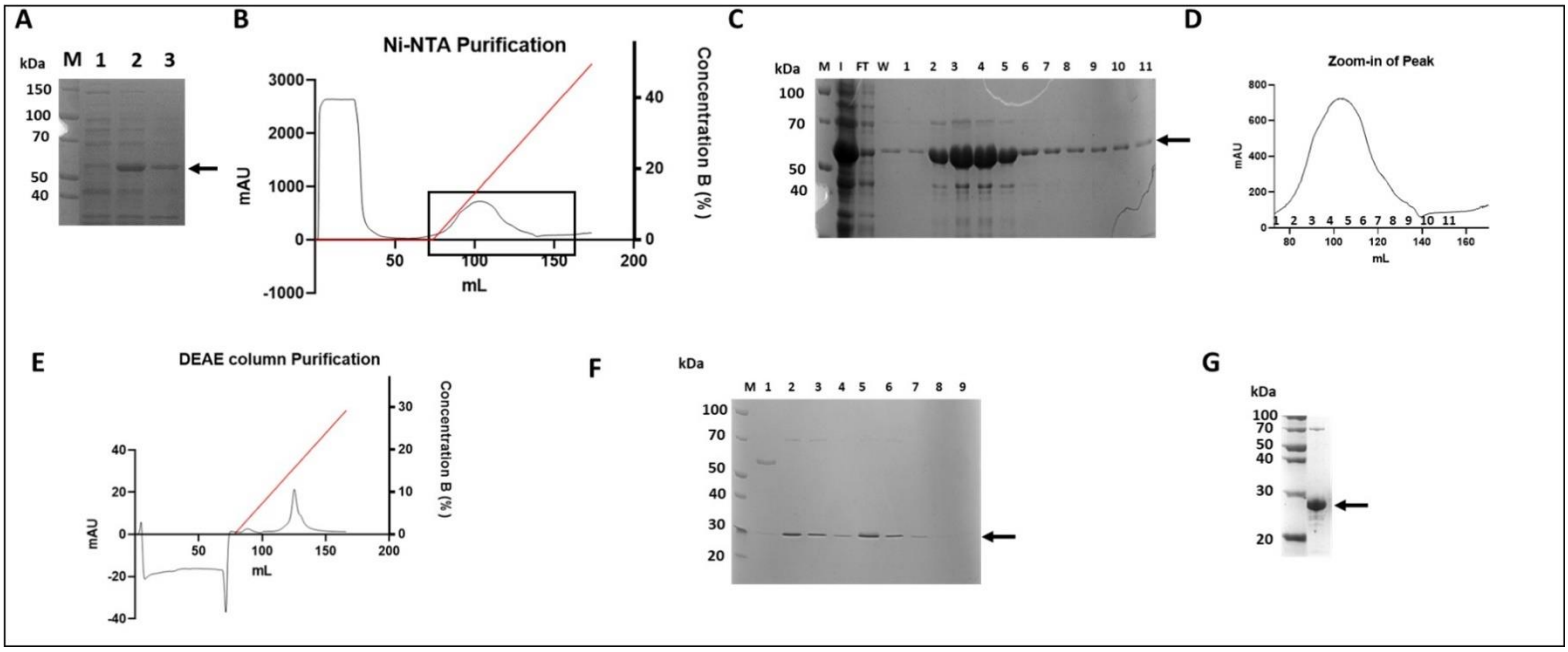


Figure 32. LARP4B 71-328 expression and purification*. [A]: Expression of LARP4B 71-328 protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of DEAE column purification. [F]: HRV 3C protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before 3C cleavage; lane 2: after 3C cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8-9: elution peak. [G]: Final protein*

LARP4B 95-328 (Figure 33) contains a His-GST-tag followed by a 3C-protease cleavage site. The protein appears soluble indicated by a single band present at around 50 kDa (Figure 33A), the soluble fractions of the protein bound to the Nickel IMAC column as one main band around 50 kDa, there is a band under 40 kDa that could be a contaminant (Figure 33C). After overnight dialysis with 3C-protease LARP4B 40-328 seemed to be cleaved, the size of the protein has been reduced to around 30 kDa on the SDS-PAGE gel, the protein mixture is subjected to a second nickel IMAC to get rid of the His-tagged protease and the uncleaved proteins. The flowthrough from the second nickel IMAC is then put through a DEAE column to separate the protein from the nucleic acids, the

flowthrough from the DEAE is dialysed overnight to the final buffer and concentrated. The concentrated protein seemed to have a contaminant just under 30 kDa, which could be from the previous purification steps, the final protein is around 25 kDa (Figure 33G).

LARP4B 95-328 in detail:

MGSSHHHHHHSQDPNSSSSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYS
KDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFP
KLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSQGSDANG
DGDQGHENAALPDPQESDPADMNALALGPSEYDSLPENSETGGNESQPDSQEDPREVLKKTLEF
CLSRENLASDMYLISQMDSDQYVPITTVANLDHIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRP
NQNRCIVILREISESTPVEEVEALFKGDNLPKFINCEFAYNDNWFITFETEADAQQAYKYLREEVK
TFQGKPIKARIKAKAIAINTFLPKNGFRPLD

(His-Tag) (GST-tag) (3C-protease cleavage site) (Construct)

Molecular Weight (Da): 54761.00

Extinction Coefficient (M^{-1} cm^{-1}): 57300

Theoretical pI: 5.02

Molecular Weight post-cleavage (Da): 26491.49

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 14440

Theoretical pI post-cleavage: 4.46

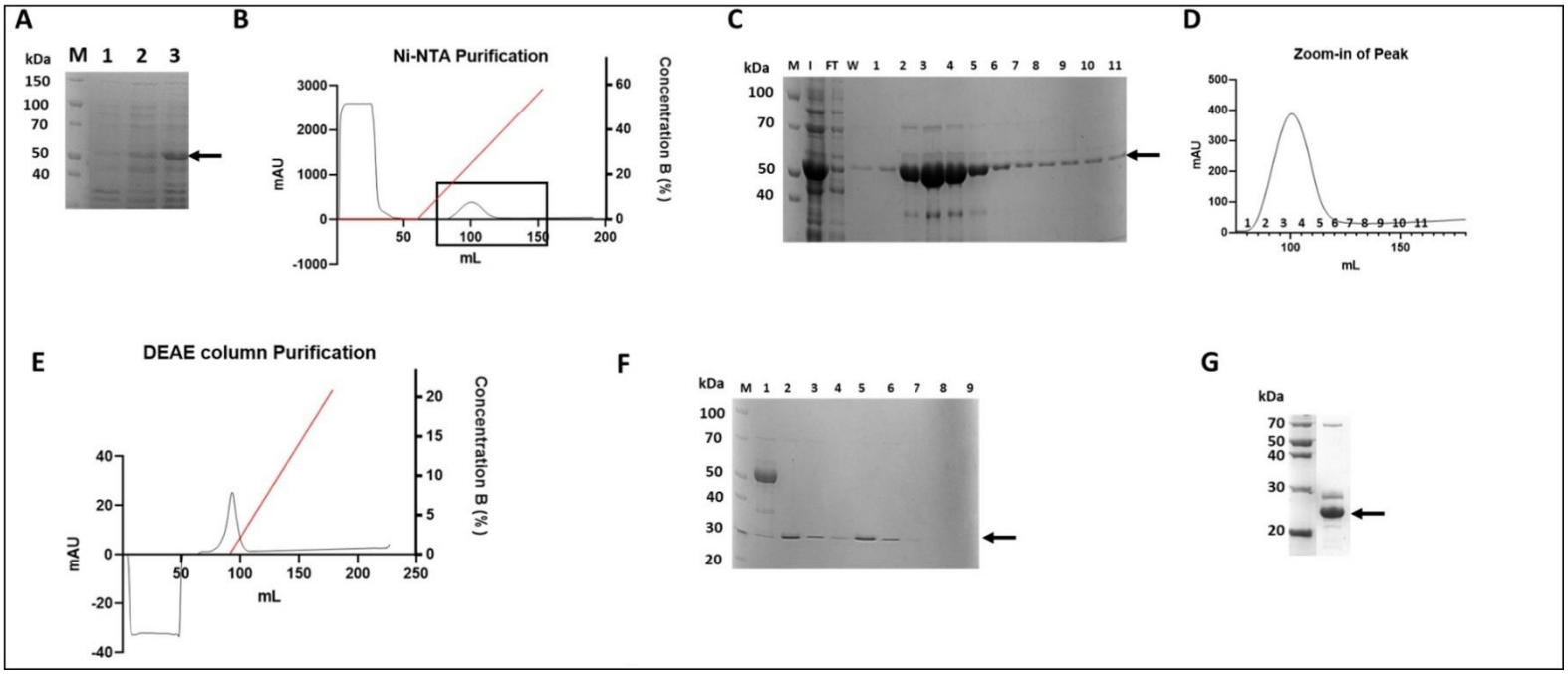


Figure 33. LARP4B 95-328 expression and purification*. [A]: Expression of LARP4B 95-328 protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of DEAE column purification. [F]: HRV 3C protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before 3C cleavage; lane 2: after 3C cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8-9: elution peak. [G]: Final protein*

3.4.4 LARP4B – LaM point mutants

In the La motif of La, specific and non-specific contacts are formed between La and UUU-3'-OH RNA ligand and in particular mediated by 6 key conserved amino acid residues present in the La motif: Q20, Y23, Y24, D33, F35 and F55 (hLa numbering). (Maraia *et al.*, 2017) These residues are extremely conserved throughout LARPs and it is therefore hypothesised that these residues could be important to bind to AU-rich RNA, for LARP4B. The corresponding amino acid in LARP4B has been found by sequence alignment (Figure 3) and mutated to alanine. One set of the LaM mutants were in the context of the NTD, and the other set is in the context of the La-module, this is to eliminate the contribution of the NTR from influencing the binding affinity towards RNA. Hypothetically speaking if the binding affinity towards RNA has changed because of a point mutation in the La motif, we could have 2 sets of results, one set would be the La-

module context, where the binding affinity is changed solely based on the contribution of the La-module; the other set would be the NTD context, where the binding affinity could be changed based on the interaction in the N-terminal region.

3.4.4.1 LARP4B – LaM point mutants (NTD context)

LARP4B NTD T163A (Figure 34) contains a His-GST-tag followed by a 3C-protease cleavage site. The protein appears to be extremely abundant, the Nickel IMAC column has one main smeared band just under 70 kDa, there is a band around 40 kDa that could be a contaminant (Figure 34A). After overnight dialysis with 3C-protease LARP4B NTD T163A seemed to be cleaved, the size of the protein has been reduced to around 40 kDa on the SDS-PAGE gel (before cleavage sample not taken), the protein mixture is subjected to a second nickel IMAC to get rid of the His-tagged protease and the uncleaved proteins. The flowthrough from the second nickel IMAC is then put through a DEAE column to separate the protein from the nucleic acids, the flowthrough from the DEAE is dialysed overnight to the final buffer and concentrated. The concentrated protein seemed to have a contaminant just around 30 and 25 kDa, which could be from the previous purification steps or degradation, the final protein is around 40 kDa (Figure 34F).

LARP4B NTD T163A in detail:

MGSSHHHHHHSQDPNSSSSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYS
KDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFP
KLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSMTSDQDA
KVVAEPQTQRVQEGKDSAHLMNGPISQTTSQTSSIPPLSQVPATKVSELNPNAEVWGAPVLHLE
ASSAADGVSAAWEEVAGHHADRGPQGSDANGDGDQGHENAALPDPQESDPADMNALALGPSE

YDSLPENSETGGNESQPDSQEDPREVLKKALEFCLSRENLASDMYLISQMDSDQYVPITTVANLD

HIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALFKGDNLPKFI

NCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPKNGFRPLD

(His-Tag) (GST-tag) (3C-protease cleavage site) (Construct)

Molecular Weight (Da): 64621.80

Extinction Coefficient (M^{-1} cm^{-1}): 68300

Theoretical pI: 4.97

Molecular Weight post-cleavage (Da): 36352.28

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 25440

Theoretical pI post-cleavage: 4.52

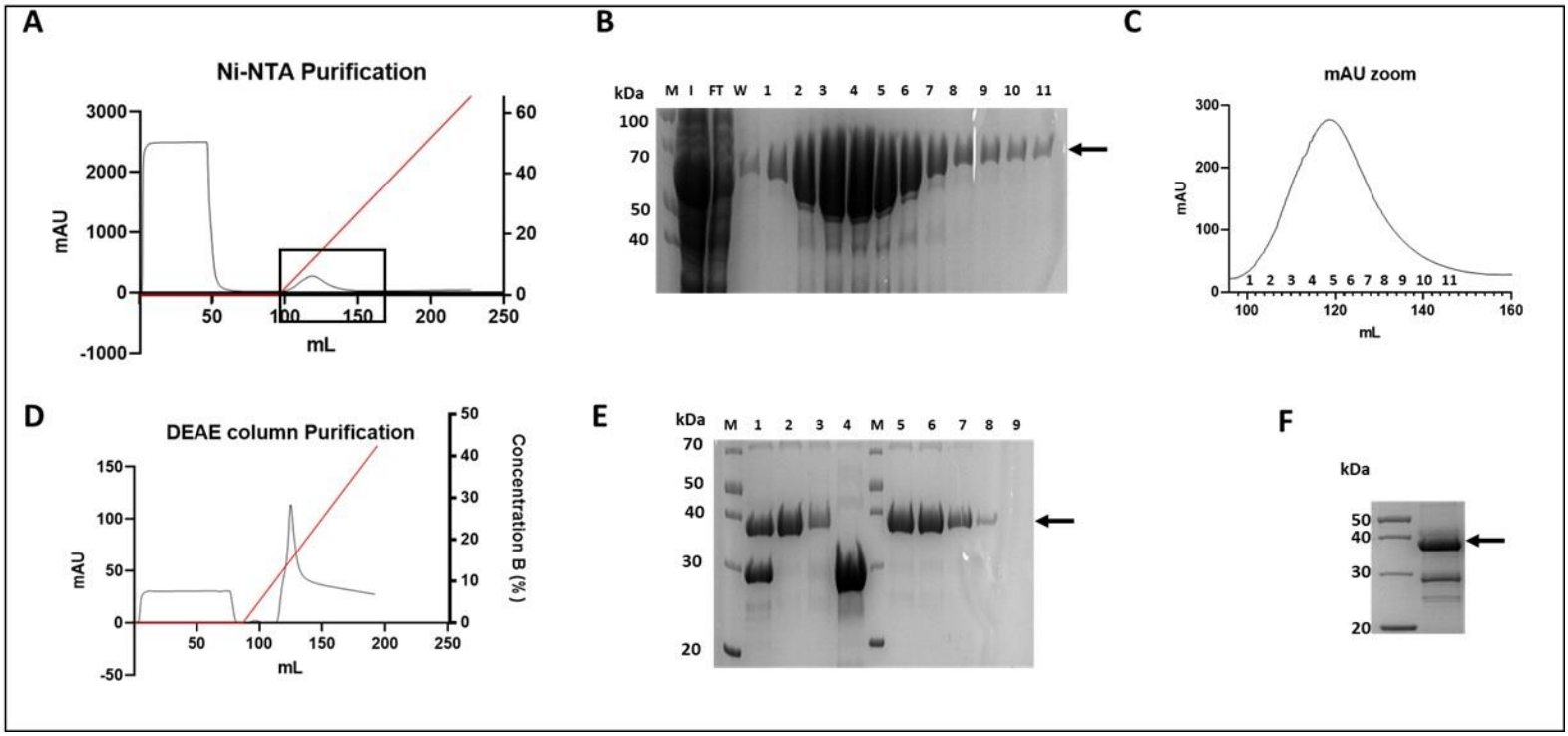


Figure 34. LARP4B NTD T163A expression and purification. *[A]: Chromatogram of Ni-NTA purification. [B]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [C]: elute fraction peaks zoom in from Ni-NTA purification [D]: Chromatogram of DEAE column purification. [E]: HRV 3C protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: after 3C cleavage and second Ni-NTA purification input; lane 2: second Ni-NTA flowthrough; lane 3: second Ni-NTA wash; lane 4: Ni-NTA elute; lane 5: DEAE input; lane 6: DEAE flowthrough; lane 7-9: elution peak. [F]: Final protein*

LARP4B NTD F166A (Figure 35) contains a His-GST-tag followed by a 3C-protease cleavage site. The Nickel IMAC column has one main band just under 70 kDa, there is a band above 70 kDa that could be a contaminant (Figure 35A). After overnight dialysis with 3C-protease LARP4B NTD F166A seemed to be cleaved, the size of the protein has been reduced to around 40 kDa on the SDS-PAGE gel, however it is quite faint, the protein mixture is subjected to a second nickel IMAC to get rid of the His-tagged protease and the uncleaved proteins. The flowthrough from the second nickel IMAC is then put through a DEAE column to separate the protein from the nucleic acids, the flowthrough from the DEAE is dialysed overnight to the final buffer and concentrated. The concentrated protein seemed to have a contaminant just around 70, 30 and one under 30 kDa, which could be from the previous purification steps or degradation, the final protein is around 40 kDa (Figure 35F).

LARP4B NTD F166A in detail:

MGSSHHHHHHSQDPNSSSSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYS
KDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFP
KLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSMTSDQDA
KVVAEPQTQRVQEGKDSAHLMNGPISQTTSQTSSIPPLSQVPATKVSELNPNAEVWGAPVLHLE
ASSAADGVSAAWEEVAGHHADRGPQGSDANGDGDQGHENAALPDPQESDPADMNALALGPSE
YDSLPENSETGGNESQPDSQEDPREVLKKTLEACLSRENLASDMYLISQMDSDQYVPITTVANLD
HIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALFKGDNLPKFI
NCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPKNGFRPLD

(His-Tag) (GST-tag) (3C-protease cleavage site) (Construct)

Molecular Weight (Da): 64575.73

Extinction Coefficient (M^{-1} cm^{-1}): 68300

Theoretical pI: 4.97

Molecular Weight post-cleavage (Da): 36306.21

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 25440

Theoretical pI post-cleavage: 4.52

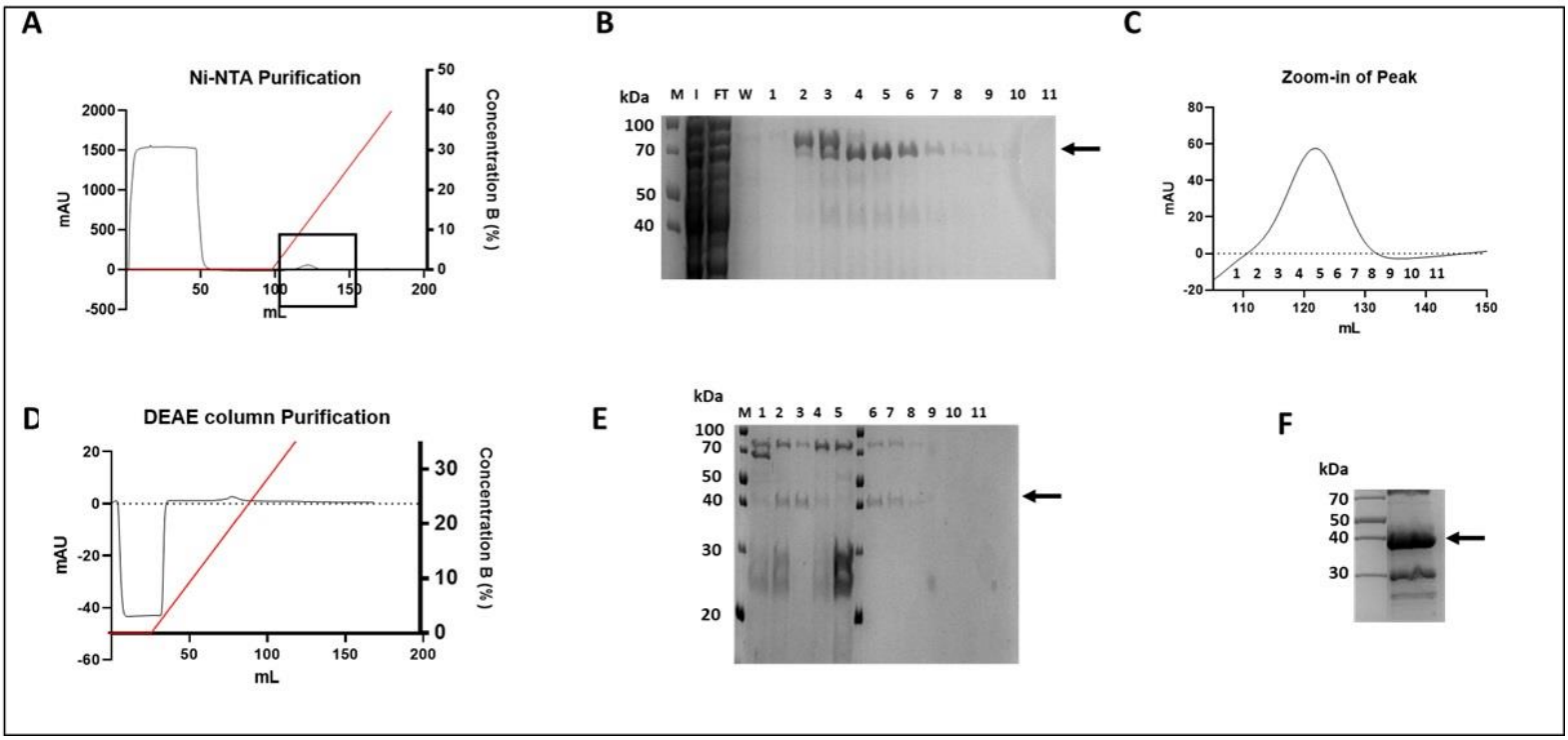


Figure 35. LARP4B NTD F166A expression and purification*. [A]: Chromatogram of Ni-NTA purification. [B]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [C]: elute fraction peaks zoom in from Ni-NTA purification [D]: Chromatogram of DEAE column purification. [E]: HRV 3C protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before 3C cleavage; lane 2: after 3C cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8-11: elution peak. [F]: Final protein*

The rest of the mutants C167A (Figure 36), D176A (Figure 37), Y178A (Figure 38) and L197A (Figure 39) are all purified using the same purification workflow as the last two mutants, and they all seem to have resulted in some contamination or degradation around 30 kDa (with the exception of Y178A, which have a single band). This 30 kDa contamination is likely to be from previous purification, it could very likely be a small concentration of the 3C-protease, since it is also conveniently at 30 kDa. This could have happened due to the over-saturation of the second Ni-IMAC column, one improvement

of the purification could be to utilize a higher volume of Nickel resin (10-15 mL instead of 5 mL).

LARP4B NTD C167A in detail:

MGSSHHHHHHSQDPNSSSSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYS
KDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFP
KLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSMTSDQDA
KVVAEPQTQRVQEGKDSAHLMNGPISQTTSQTSSIPPLSQVPATKVSELNPNAEVWGAPVLHLE
ASSAADGVSAAWEEVAGHHADRGPQGSDANGDGDQGHENAALPDPQESDPADMNALALGPSE
YDSLPENSETGGNESQPDSQEDPREVLKKTLEFALSRENLASDMYLISQMDSDQYVPITTVANLD
HIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALFKGDNLPKFI
NCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPKNGFRPLD

(His-Tag) (GST-tag) (3C-protease cleavage site) (Construct)

Molecular Weight (Da): 64619.76

Extinction Coefficient (M^{-1} cm^{-1}): 68300

Theoretical pI: 4.97

Molecular Weight post-cleavage (Da): 36350.25

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 25440

Theoretical pI post-cleavage: 4.52

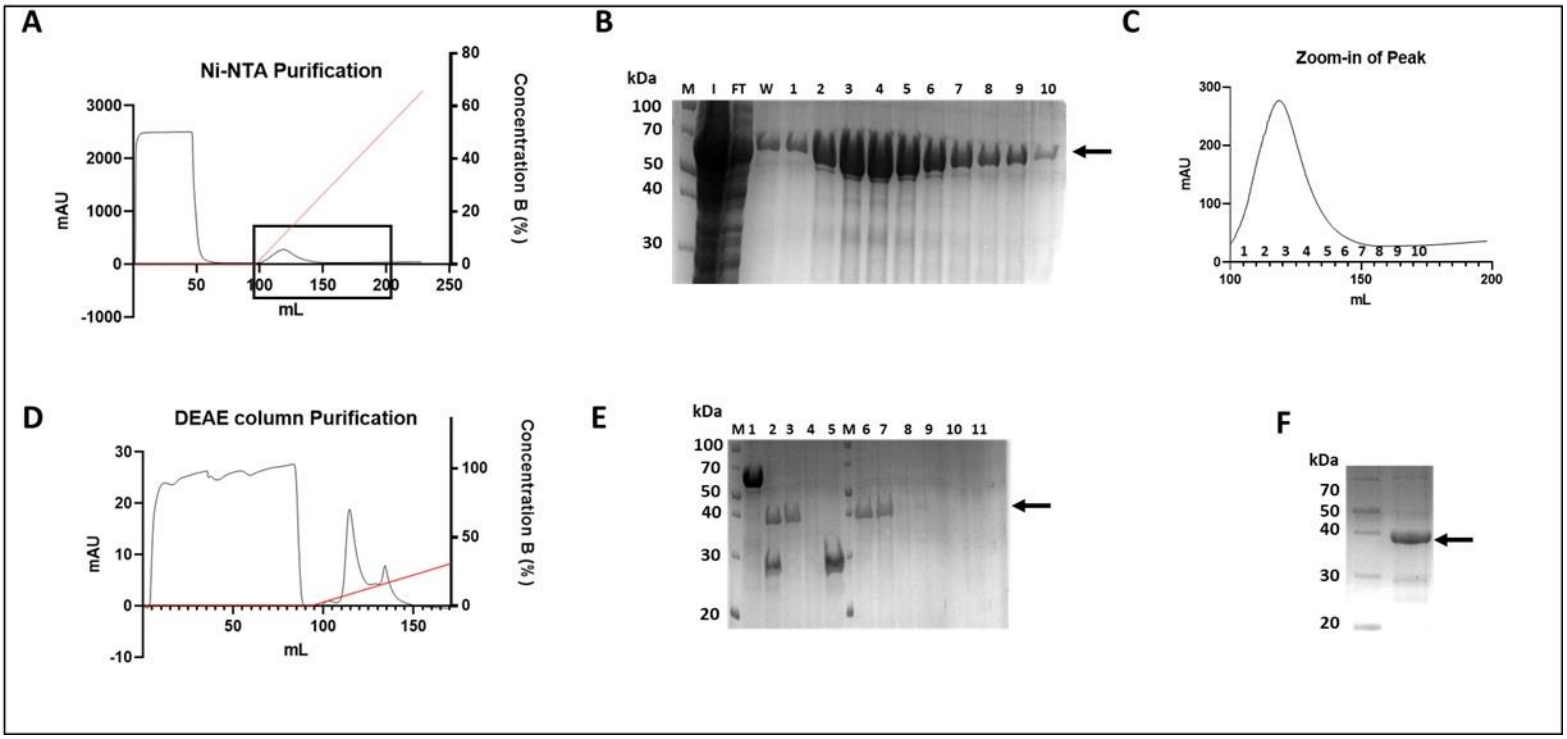


Figure 36. LARP4B NTD C167A expression and purification. *[A]: Chromatogram of Ni-NTA purification. [B]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-10: elute fractions. [C]: elute fraction peaks zoom in from Ni-NTA purification [D]: Chromatogram of DEAE column purification. [E]: HRV 3C protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before 3C cleavage; lane 2: after 3C cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8-11: elution peak. [F]: Final protein*

LARP4B NTD D176A in detail:

MGSSHHHHHHSQDPNSSSSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE

LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYS

KDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFP

KLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSMTSDQDA

KVVAEPQTQRVQEGKDSAHLMNGPISQTTSQTSSIPPLSQVPATKVSELNPNAEVWGAPVLHLE

ASSAADGVSAAWEEVAGHHADRGPQGSDANGDGDQGHENAALPDPQESDPADMNALALGPSE

YDSLPENSETGGNESQPDSQEDPREVLKKTLEFCLSRENLASAMYLISQMDSDQYVPITTVANLD

HIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALFKGDNLPKFI

NCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPKNGFRPLD

(His-Tag) (GST-tag) (3C-protease cleavage site) (Construct)

Molecular Weight (Da): 64607.81

Extinction Coefficient (M^{-1} cm^{-1}): 68300

Theoretical pI: 5.00

Molecular Weight post-cleavage (Da): 36338.30

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 25440

Theoretical pI post-cleavage: 4.55

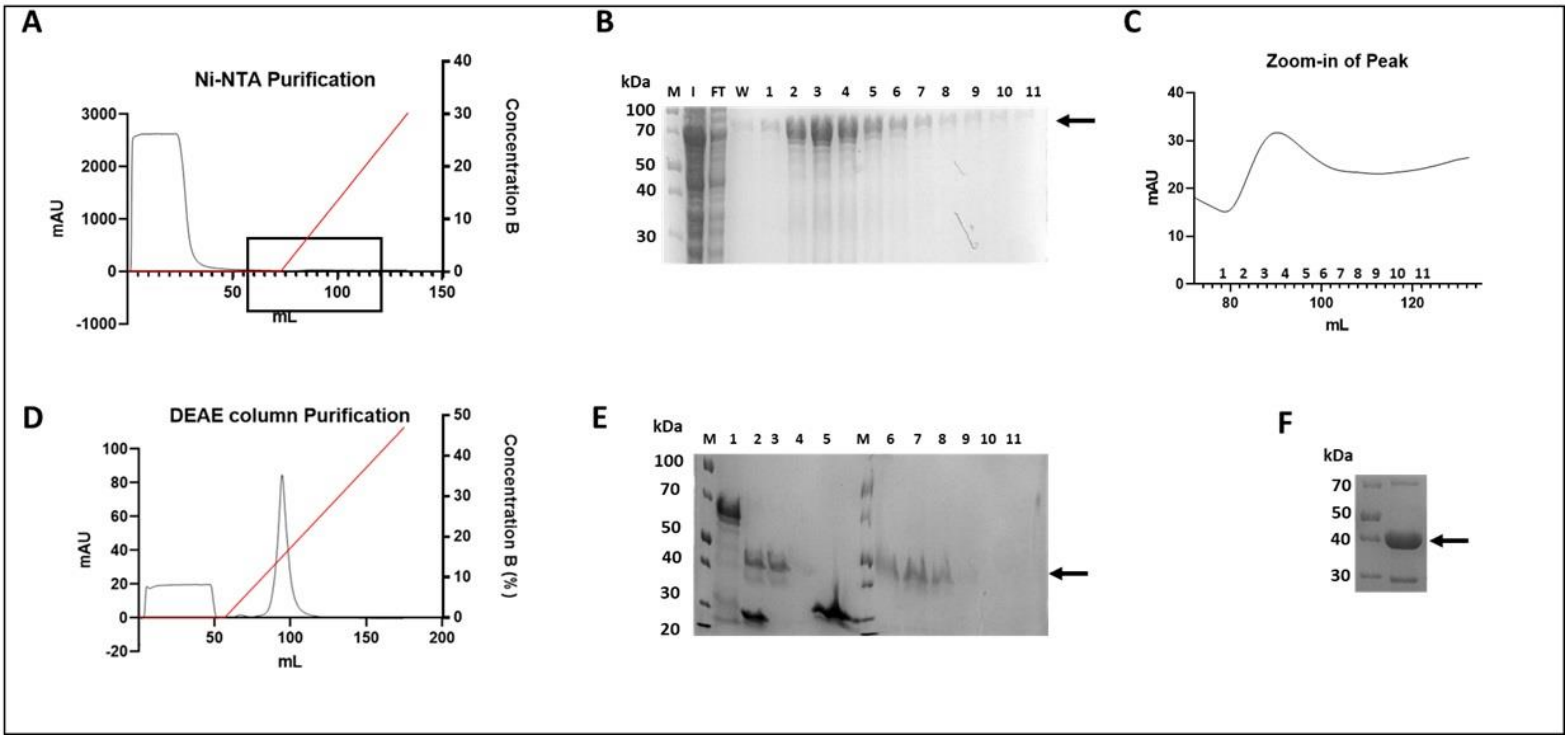


Figure 37. LARP4B NTD D176A expression and purification*. [A]: Chromatogram of Ni-NTA purification. [B]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [C]: elute fraction peaks zoom in from Ni-NTA purification [D]: Chromatogram of DEAE column purification. [E]: HRV 3C protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before 3C cleavage; lane 2: after 3C cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8-11: elution peak. [F]: Final protein*

LARP4B NTD Y178A in detail:

MGSSHHHHHHSQDPNSSSSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE

LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYS

KDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFP

KLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSMTSDQDA

KVVAEPQTQRVQEGKDSAHLMNGPISQTTSQTSSIPPLSQVPATKVSELNPNAEVWGAPVLHLE

ASSAADGVSAAWEEVAGHHADRGPQGSDANGDGDQGHENAALPDPQESDPADMNALALGPSE

YDSLPENSETGGNESQPDSQEDPREVLKKTLEFCLSRENLASDMALISQMDSDQYVPITTVANLD

HIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALFKGDNLPKFI

NCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPKNGFRPLD

(His-Tag) (GST-tag) (3C-protease cleavage site) (Construct)

Molecular Weight (Da): 64559.73

Extinction Coefficient (M^{-1} cm^{-1}): 66810

Theoretical pI: 4.97

Molecular Weight post-cleavage (Da): 36290.21

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 23950

Theoretical pI post-cleavage: 4.52

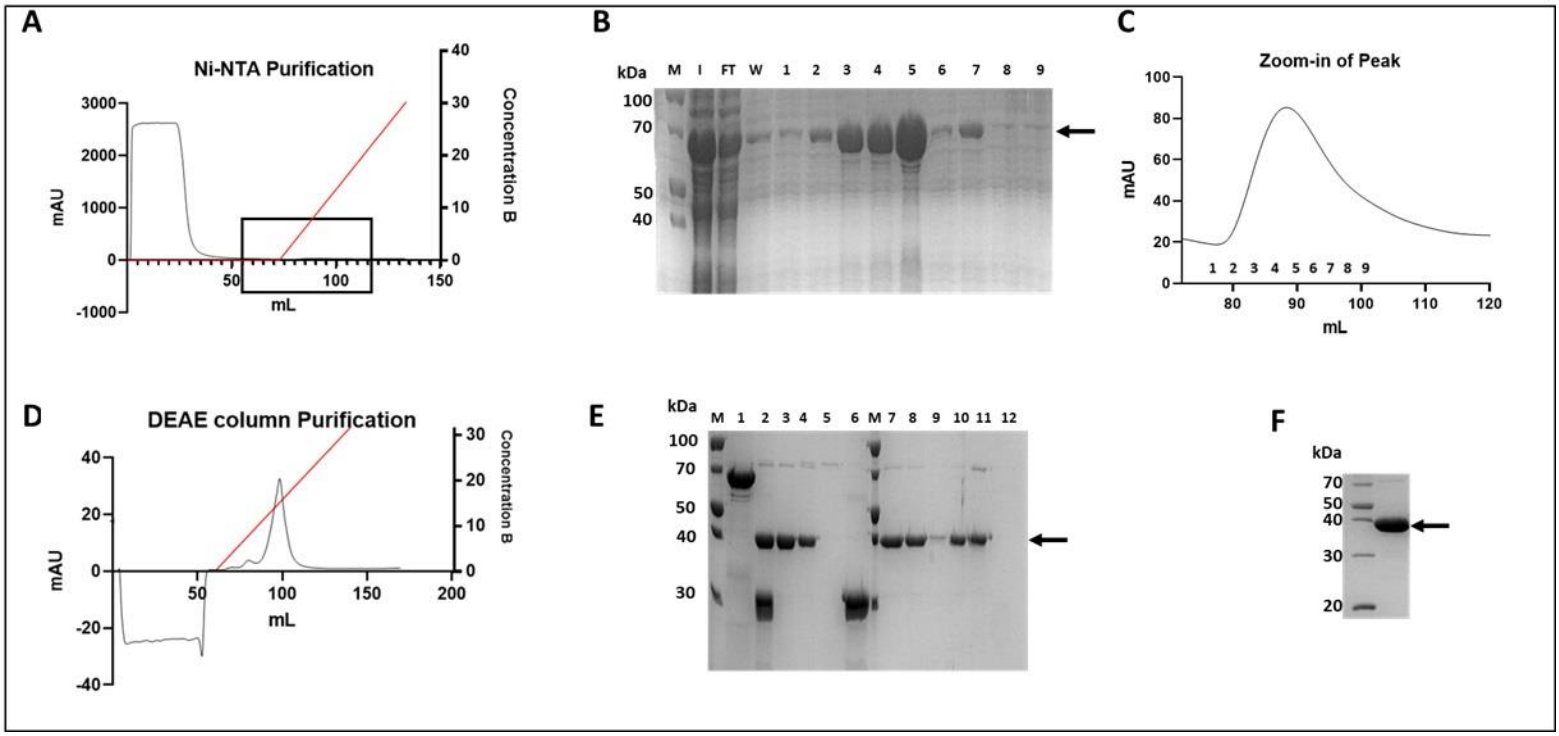


Figure 38. LARP4B NTD Y178A expression and purification. *[A]: Chromatogram of Ni-NTA purification. [B]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-9: elute fractions. [C]: elute fraction peaks zoom in from Ni-NTA purification [D]: Chromatogram of DEAE column purification. [E]: HRV 3C protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before 3C cleavage; lane 2: after 3C cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8-12: elution peak. [F]: Final protein*

LARP4B NTD L197A in detail:

MGSSHHHHHHSQDPNSSSSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYS
KDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFP
KLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSMTSDQDA
KVVAEPQTQRVQEGKDSAHLMNGPISQTTSQTSSIPPLSQVPATKVSELNPNAEVWGAPVLHLE
ASSAADGVSAAWEEVAGHHADRGPQGSDANGDGDQGHENAALPDPQESDPADMNALALGPSE
YDSLPENSETGGNESQPDSQEDPREVLKKTLEFCLSRENLASDMYLISQMDSDQYVPITTVANAD
HIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALFKGDNLPKFI
NCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPKNGFRPLD

(His-Tag) (GST-tag) (3C-protease cleavage site) (Construct)

Molecular Weight (Da): 64609.74

Extinction Coefficient (M^{-1} cm^{-1}): 68300

Theoretical pI: 4.97

Molecular Weight post-cleavage (Da): 36340.23

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 25440

Theoretical pI post-cleavage: 4.52

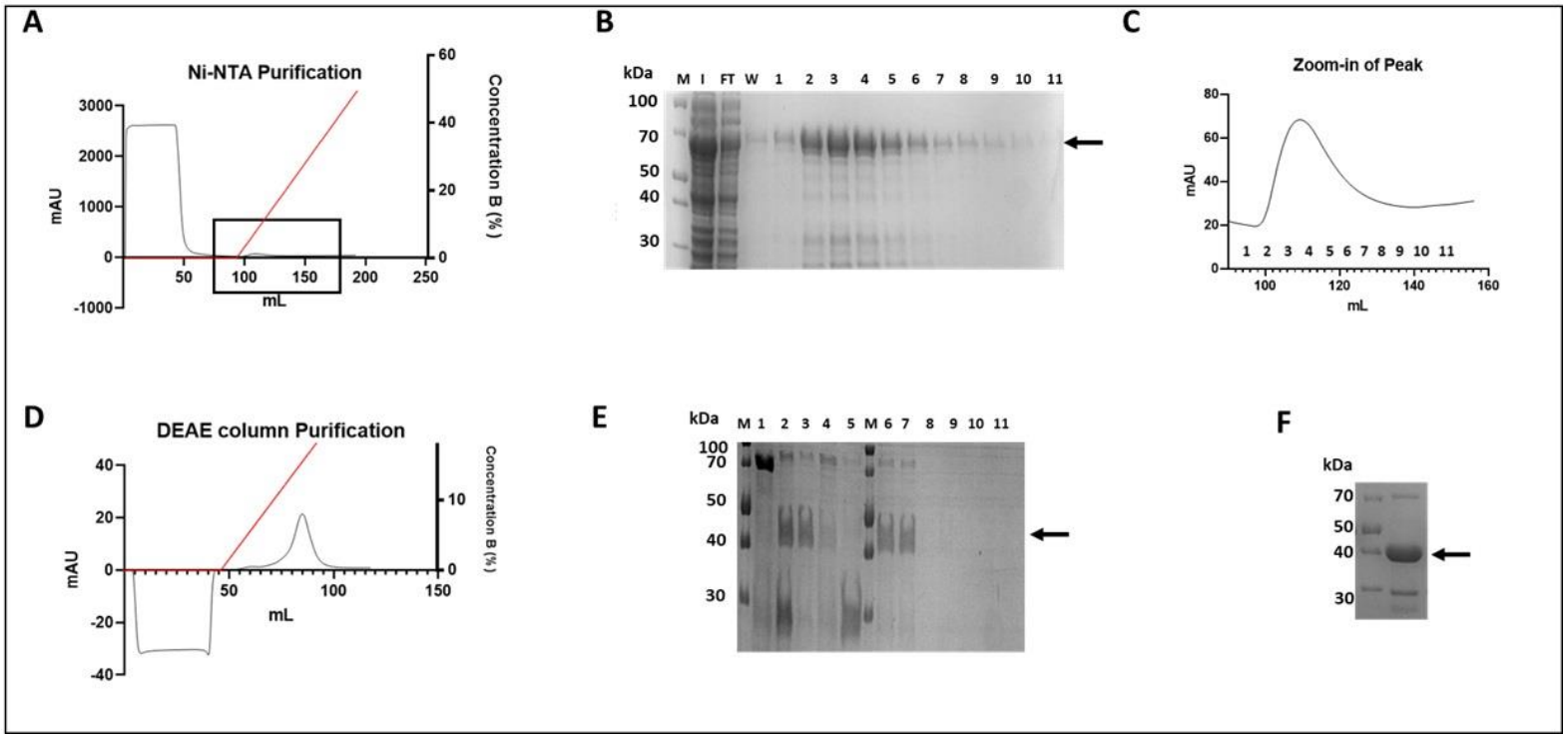


Figure 39. LARP4B NTD L197A expression and purification. *[A]: Chromatogram of Ni-NTA purification. [B]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [C]: elute fraction peaks zoom in from Ni-NTA purification [D]: Chromatogram of DEAE column purification. [E]: HRV 3C protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before 3C cleavage; lane 2: after 3C cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8-11: elution peak. [F]: Final protein*

3.4.4.2 LARP4B – LaM point mutants (La-module context)

LARP4B La-module T163A (Figure 40) contains only a His-tag followed by a TEV-protease cleavage site, the La-module mutants are significantly smaller in size compared to the mutants in the context of the NTD. The protein appears to be extremely abundant, the Nickel IMAC column has one main smeared band around 25 kDa, (Figure 40A). After overnight dialysis with TEV-protease, LARP4B La-module T163A seemed to be cleaved, the size of the protein has been reduced to just under 20 kDa on the SDS-PAGE gel, the protein mixture is subjected to a second nickel IMAC to get rid of the His-tagged protease and the uncleaved proteins. The flowthrough from the second nickel IMAC is then put through a DEAE column to separate the protein from the nucleic acids, the flowthrough from the DEAE (gel not shown) is dialysed overnight to the final buffer and concentrated.

The concentrated protein seemed to be a single band, indicating that it is pure (Figure 40F).

LARP4B La-module T163A in detail:

MGSSHHHHHHSQDPNSSSENLYFQSDSQEDPREVLKKALEFCLSRENLASDMYLISQMDSDQYV PITTVANLDHIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALF KGDNLPKFINCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPK NGFPLD

(His-Tag) (TEV-protease cleavage site) (Construct)

T163A: Molecular Weight (Da): 23184.15

Extinction Coefficient (M^{-1} cm^{-1}): 14440

Theoretical pI: 5.27

Molecular Weight post-cleavage (Da): 20529.41

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 12950

Theoretical pI post-cleavage: 4.90

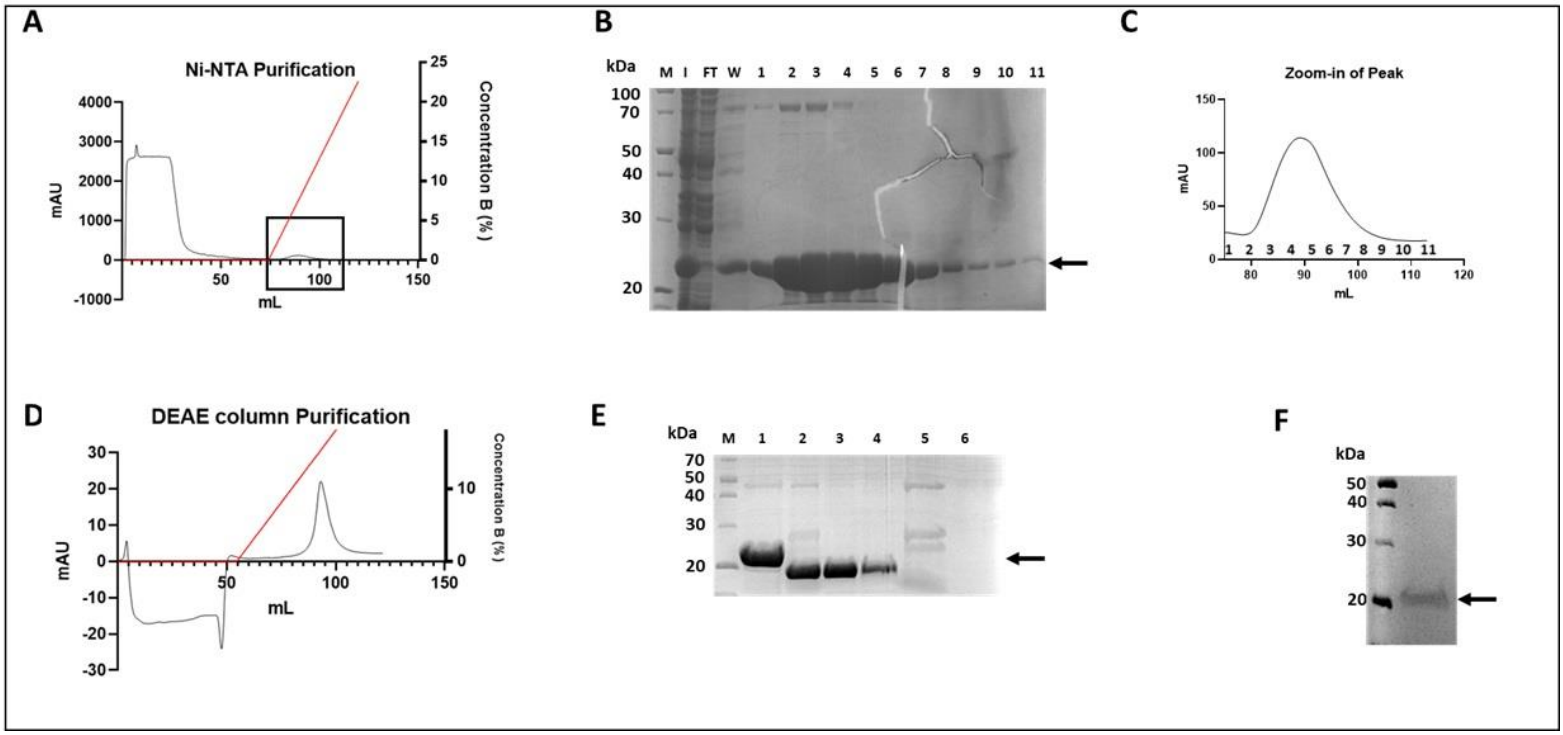


Figure 40. LARP4B La-module T163A expression and purification*. [A]: Chromatogram of Ni-NTA purification. [B]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [C]: elute fraction peaks zoom in from Ni-NTA purification [D]: Chromatogram of DEAE column purification. [E]: TEV protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before TEV cleavage; lane 2: after TEV cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute. [F]: Final protein*

LARP4B La-module F166A (Figure 41) contains only a His-tag followed by a TEV-protease cleavage site. The protein appears to be extremely abundant, the Nickel IMAC column has one main smeared band around 25 kDa, (Figure 41A). After overnight dialysis with TEV-protease, LARP4B La-module F166A seemed to be cleaved, the size of the protein has been reduced to just under 20 kDa on the SDS-PAGE gel, the protein mixture is subjected to a second nickel IMAC to get rid of the His-tagged protease and the uncleaved proteins. The flowthrough from the second nickel IMAC is then put through a DEAE column to separate the protein from the nucleic acids, the flowthrough from the DEAE is dialysed overnight to the final buffer and concentrated. The concentrated protein seemed to be a single band, indicating that it is pure (Figure 41F).

LARP4B La-module F166A in detail:

MGSSHHHHHHSQDPNSSSENLYFQSDSQEDPREVLKKTLEACLSRENLASDMYLISQMDSDQYV
PITTVANLDHIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALF
KGDNLPKFINCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPK
NGFPLD

(His-Tag) (TEV-protease cleavage site) (Construct)

Molecular Weight (Da): 23138.07

Extinction Coefficient (M^{-1} cm^{-1}): 14440

Theoretical pI: 5.27

Molecular Weight post-cleavage (Da): 20483.34

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 12950

Theoretical pI post-cleavage: 4.90

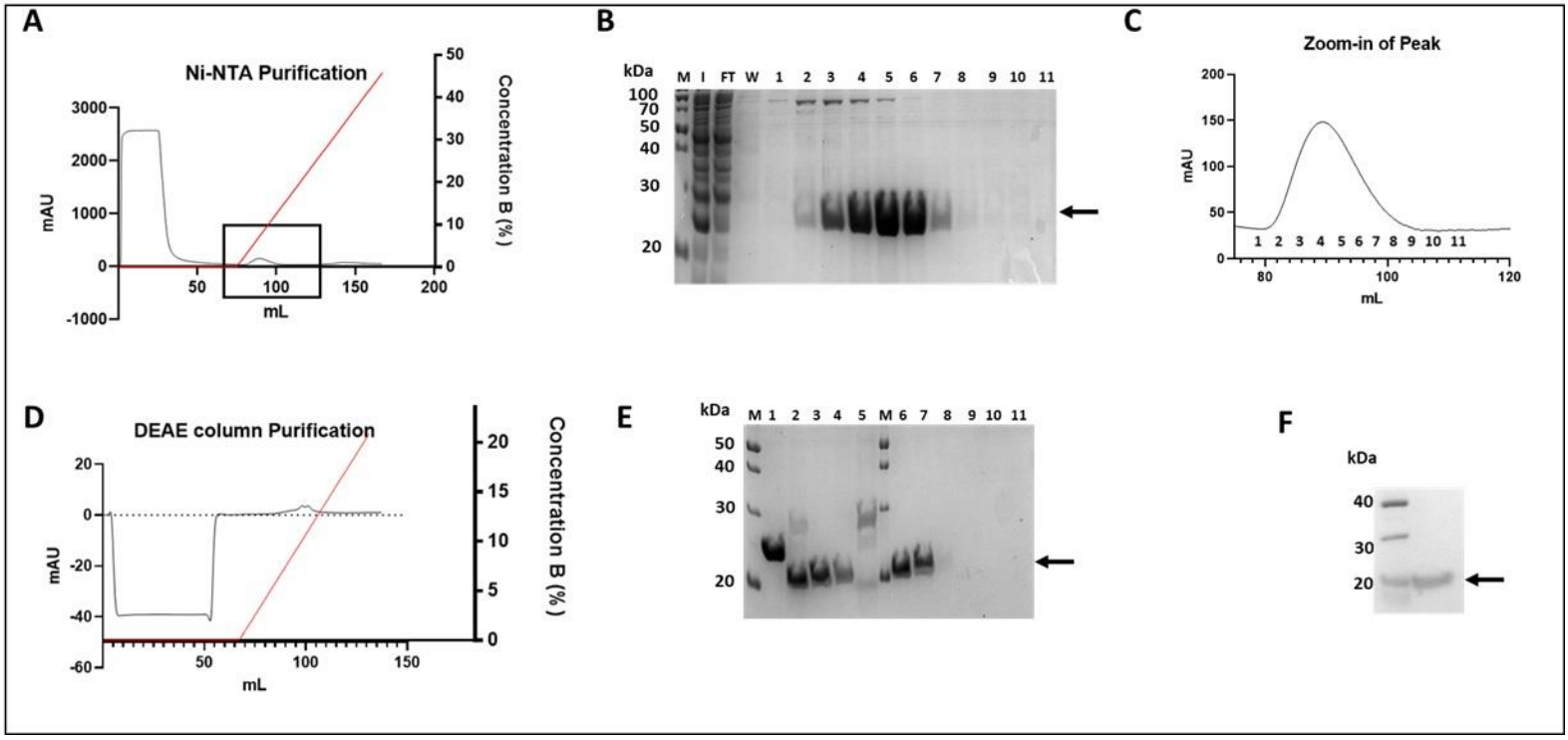


Figure 41. LARP4B La-module F166A expression and purification*. [A]: Chromatogram of Ni-NTA purification. [B]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [C]: elute fraction peaks zoom in from Ni-NTA purification [D]: Chromatogram of DEAE column purification. [E]: TEV protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before TEV cleavage; lane 2: after TEV cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8-11: elution peak. [F]: Final protein*

The rest of the mutants C167A (Figure 42), D176A (Figure 43), Y178A (Figure 44) and L197A (Figure 45) are all purified using the same purification workflow as the last two mutants, and of the constructs resulted in a single band in the final protein, which indicated that the protein is pure. The exception to this is Y178A, where another band was seen associating with the protein (Figure 44E).

LARP4B La-module C167A in detail:

MGSSHHHHHHSQDPNSSSENLYFQSDSQEDPREVLKKTLEFALSRENLASDMYLISQMDSDQYV PITTVANLDHIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALF KGDNLPKFINCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPK NGFPLD

(His-Tag) (TEV-protease cleavage site) (Construct)

Molecular Weight (Da): 23182.11

Extinction Coefficient (M^{-1} cm^{-1}): 14440

Theoretical pI: 5.27

Molecular Weight post-cleavage (Da): 20527.38

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 12950

Theoretical pI post-cleavage: 4.90

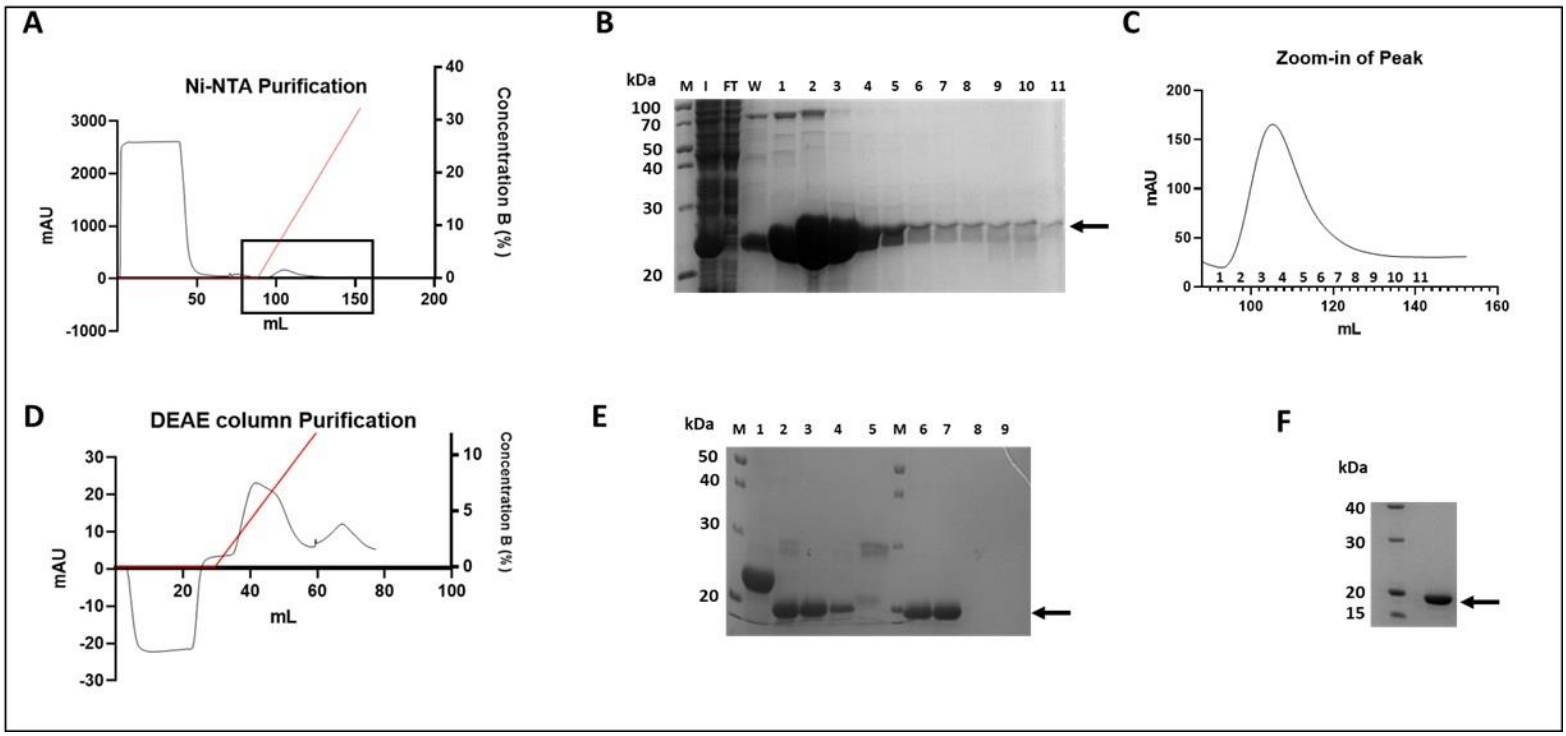


Figure 42. LARP4B La-module C167A expression and purification. *[A]: Chromatogram of Ni-NTA purification. [B]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [C]: elute fraction peaks zoom in from Ni-NTA purification [D]: Chromatogram of DEAE column purification. [E]: TEV protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before TEV cleavage; lane 2: after TEV cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8-9: elution peak. [F]: Final protein*

LARP4B La-module D176A in detail:

MGSSHHHHHHSQDPNSSSENLYFQSDSQEDPREVLKKTLEFCLSRENLASAMYLISQMDSDQYV
PITTVANLDHIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALF
KGDNLPKFINCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPK
NGFPLD

(His-Tag) (TEV-protease cleavage site) (Construct)

Molecular Weight (Da): 23170.16

Extinction Coefficient (M^{-1} cm^{-1}): 14440

Theoretical pI: 5.39

Molecular Weight post-cleavage (Da): 20515.43

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 12950

Theoretical pI post-cleavage: 4.99

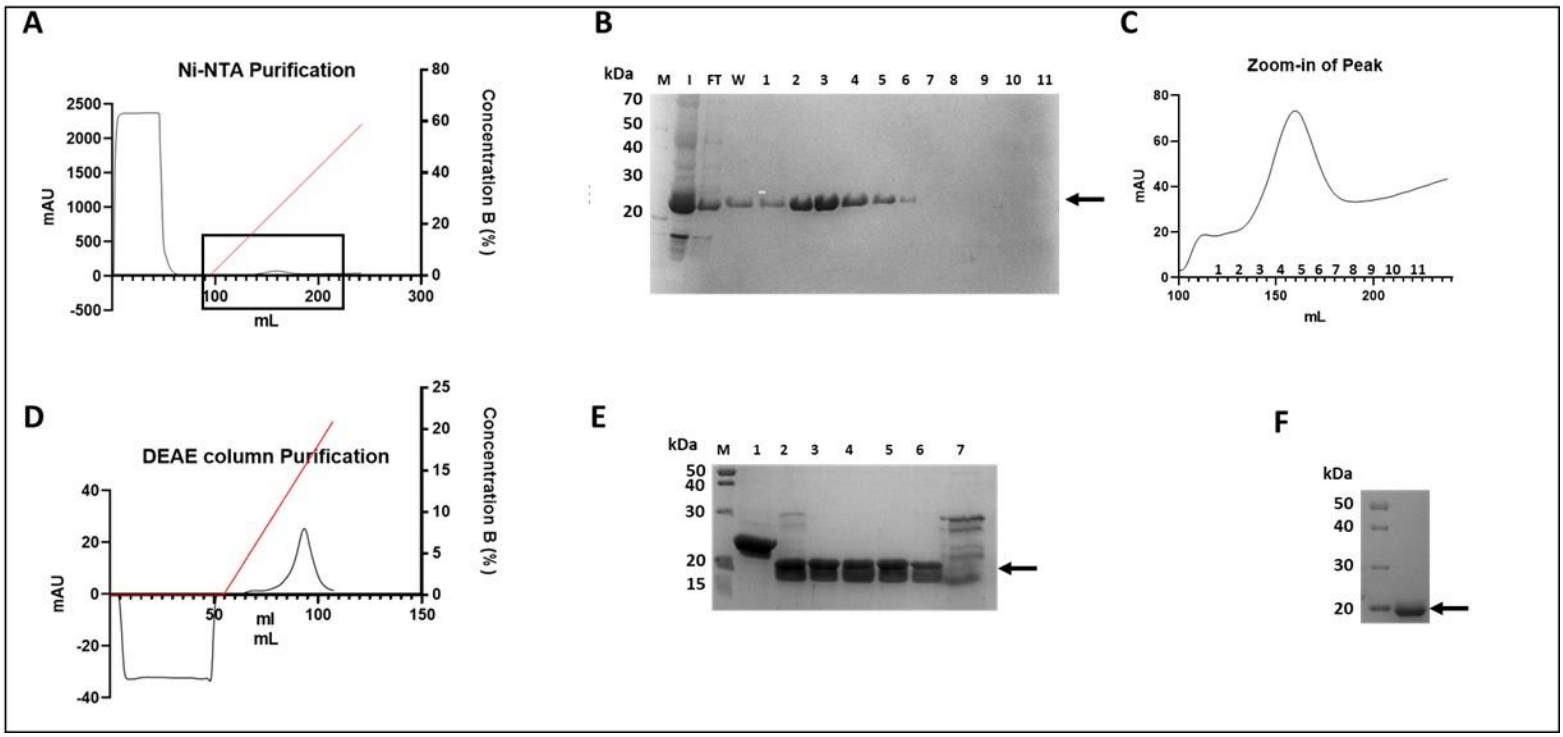


Figure 43. LARP4B La-module D176A expression and purification. *[A]: Chromatogram of Ni-NTA purification. [B]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [C]: elute fraction peaks zoom in from Ni-NTA purification [D]: Chromatogram of DEAE column purification. [E]: TEV protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before TEV cleavage; lane 2: after TEV cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: DEAE input; lane 5: DEAE flowthrough; lane 6: DEAE elution peak; lane 7: Elution from Ni-NTA. [F]: Final protein*

LARP4B La-module Y178A in detail:

MGSSHHHHHHSQDPNSSSENLYFQSDSQEDPREVLKKTLEFCLSRENLASDMALISQMDSDQYV
PITTVANLDHIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALF
KGDNLPKFINCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPK
NGFPLD

(His-Tag) (TEV-protease cleavage site) (Construct)

Molecular Weight (Da): 23122.07

Extinction Coefficient (M^{-1} cm^{-1}): 12950

Theoretical pI: 5.27

Molecular Weight post-cleavage (Da): 20467.34

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 11460

Theoretical pI post-cleavage: 4.90

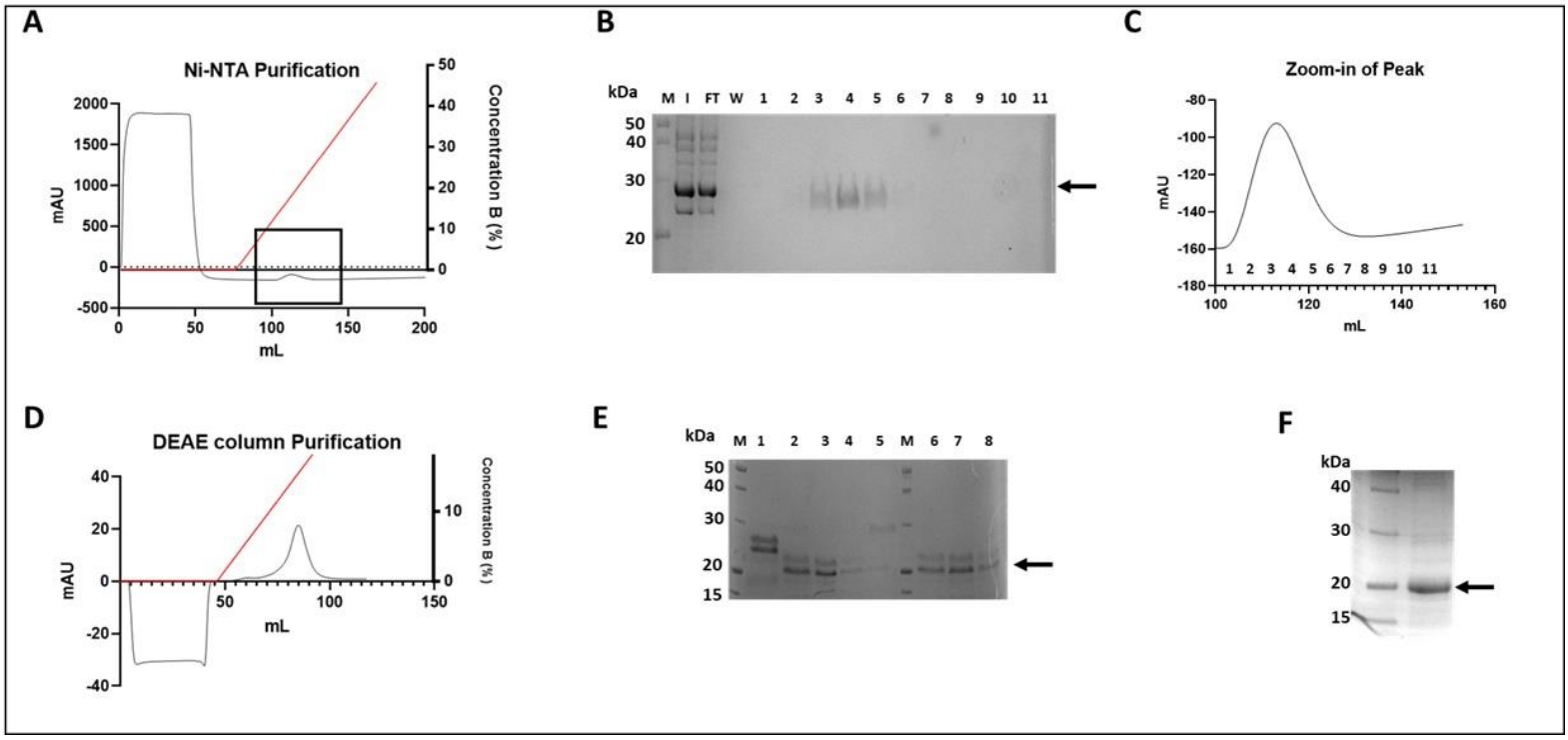


Figure 44. LARP4B La-module Y178A expression and purification*. [A]: Chromatogram of Ni-NTA purification. [B]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [C]: elute fraction peaks zoom in from Ni-NTA purification [D]: Chromatogram of DEAE column purification. [E]: TEV protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before TEV cleavage; lane 2: after TEV cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8: elution peak. [F]: Final protein*

LARP4B La-module L197A:

MGSSHHHHHHSQDPNSSSENLYFQSDSQEDPREVLKKTLEFCLSRENLASDMYLISQMDSDQYV
PITTVANADHIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALF
KGDNLPKFINCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPK
NGFPLD

(His-Tag) (TEV-protease cleavage site) (Construct)

Molecular Weight (Da): 23172.09

Extinction Coefficient (M^{-1} cm^{-1}): 14440

Theoretical pI: 5.27

Molecular Weight post-cleavage (Da): 20517.36

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 12950

Theoretical pI post-cleavage: 4.90

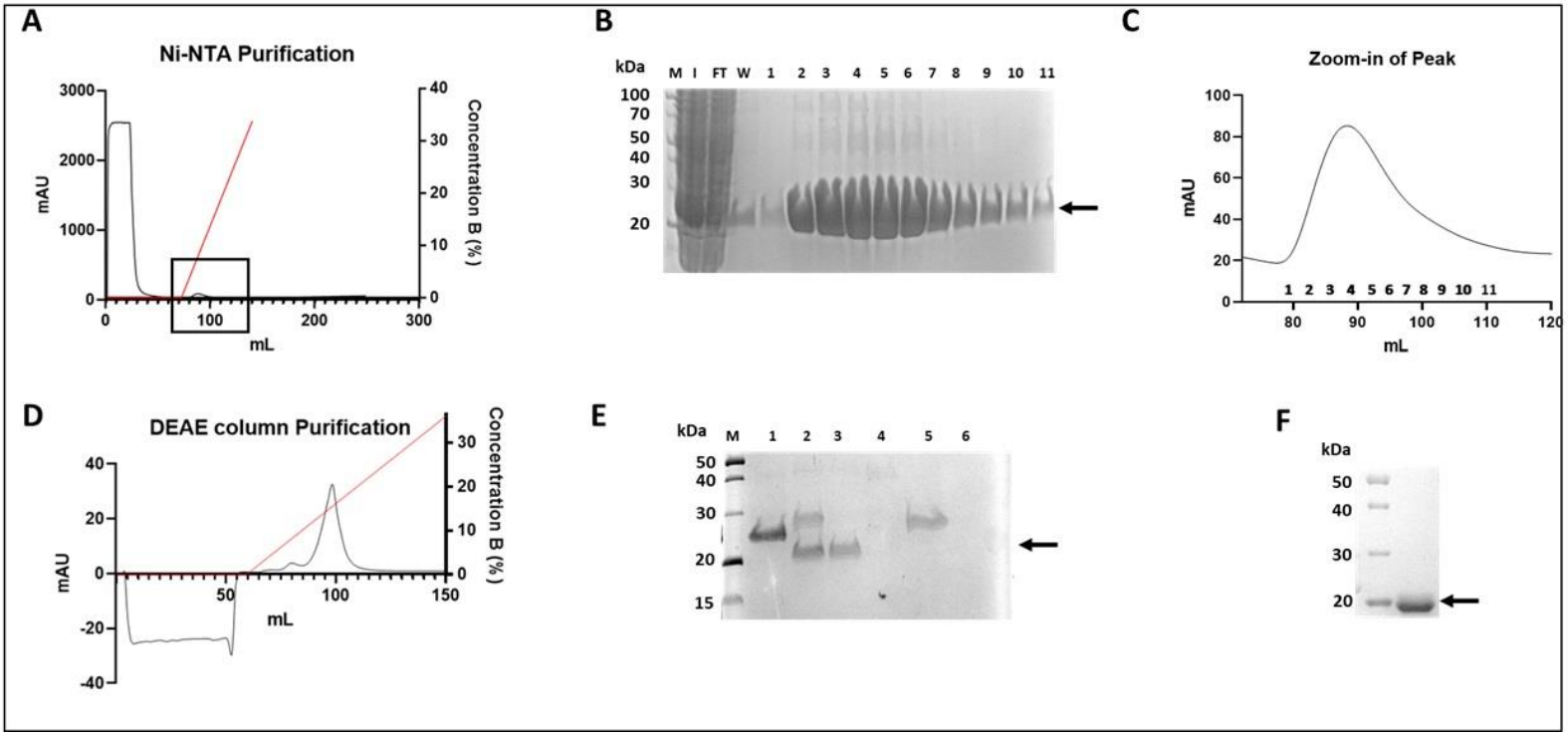


Figure 45. LARP4B La-module L197A expression and purification*. [A]: Chromatogram of Ni-NTA purification. [B]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [C]: elute fraction peaks zoom in from Ni-NTA purification [D]: Chromatogram of DEAE column purification. [E]: TEV protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before TEV cleavage; lane 2: after TEV cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute. [F]: Final protein*

3.4.5 LARP4B – La motif and RRM

As part of this study to understand the molecular mechanisms of LARP4B binding to RNA (Chapter 4), I also purified the individual domains within the La-module of LARP4B, namely the La motif (LaM) and RNA recognition motif (RRM).

LARP4B La motif (LaM) (Figure 46) contains only a His-tag followed by a TEV-protease cleavage site. The protein appears to be expressed in the soluble fraction and is between 15 and 20 kDa, appearing as a smear (Figure 46A), the Nickel IMAC column has one main smeared band around the same molecular weight, (Figure 46C). After overnight dialysis with TEV-protease, LARP4B LaM seemed to be cleaved, the size of the protein has been reduced to just around 10 kDa on the SDS-PAGE gel (Figure 46F), the protein

mixture is subjected to a second nickel IMAC to get rid of the His-tagged protease and the uncleaved proteins. The flowthrough from the second nickel IMAC is then put through a DEAE column to separate the protein from the nucleic acids, the flowthrough from the DEAE is dialysed overnight to the final buffer and concentrated. The concentrated protein seemed to be a single band, indicating that it is pure (Figure 46G)

LARP4B LaM in detail:

MGSSHHHHHHSQDPNSSSENLYFQSDSQEDPREVLKKTLEFCLSRENLASDMYLISQMDSDQYV

PITTVANLDHIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQ

(His-Tag) (TEV-protease cleavage site) (Construct)

Molecular Weight (Da): 12371.74

Extinction Coefficient (M^{-1} cm^{-1}): 4470

Theoretical pI: 5.18

Molecular Weight post-cleavage (Da): 9501.80

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 2980

Theoretical pI post-cleavage: 4.59

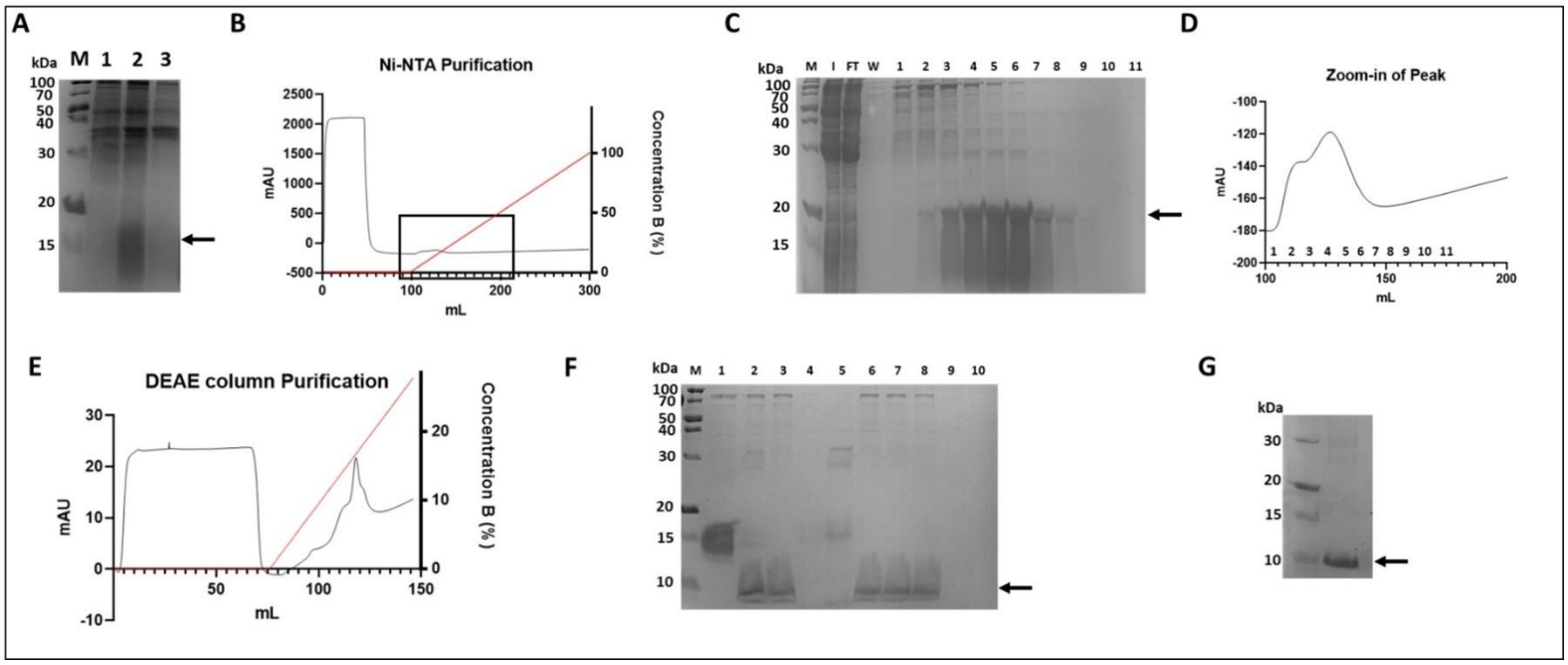


Figure 46. LARP4B La motif expression and purification. *[A]: Expression of LARP4B LaM protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of DEAE column purification. [F]: TEV protease cleavage and DEAE column purification, lane M: protein ladder; lane 1: before TEV cleavage; lane 2: after TEV cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: DEAE input; lane 7: DEAE flowthrough; lane 8-10: elution peak. [G]: Final protein*

LARP4B RNA recognition motif (RRM) (Figure 47) contains only a His-tag followed by a TEV-protease cleavage site. The protein appears to be expressed in the soluble fraction and is around 15 kDa (Figure 47A), the Nickel IMAC column has one main smeared band around the same molecular weight, (Figure 47C). After overnight dialysis with TEV-protease, LARP4B RRM seemed to have not been cleaved, the size of the protein has remained to be 15 kDa on the SDS-PAGE gel (Figure 47F). The cleavage was re-tried and even done at room temperature overnight, but the cleavage is still not efficient (not shown). The protein mixture is subjected to a second nickel IMAC to get rid of the His-tagged protease and the uncleaved proteins. The flowthrough from the second nickel IMAC is then put through a Heparin column to separate the protein from the nucleic acids, since the pI of RRM is quite high, at 6.51. The fractions from the Heparin elution (Figure 47F, lanes 8-14) is pooled and dialysed overnight to the final buffer and concentrated. The concentrated protein seemed to be a single smeared band, at 15 kDa.

LARP4B RRM:

MGSSHHHHHHSQDPNSSSENLYFQSNRCIVILREISESTPVEEVEALFKGDNLPKFINCEFAYNDN

WFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPKNGFRPLD

(His-Tag) (TEV-protease cleavage site) (Construct)

Molecular Weight (Da): 13886.57

Extinction Coefficient (M^{-1} cm^{-1}): 11460

Theoretical pI: 6.29

Molecular Weight post-cleavage (Da): 11016.63

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 9970

Theoretical pI post-cleavage: 6.51

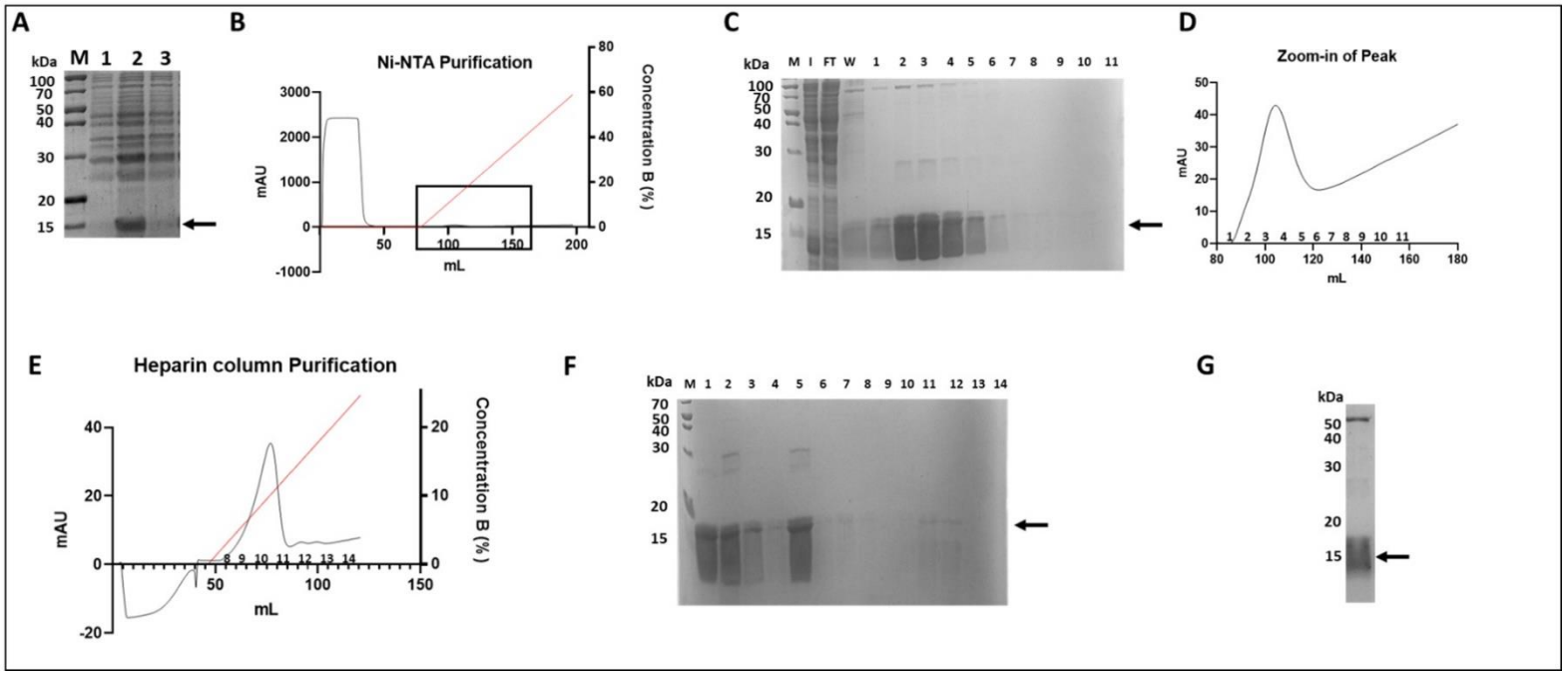


Figure 47. LARP4B RRM expression and purification. *[A]: Expression of LARP4B RRM protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction, soluble fraction; lane 3: after IPTG induction, insoluble fraction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of Heparin column purification. [F]: TEV protease cleavage and Heparin column purification, lane M: protein ladder; lane 1: before TEV cleavage; lane 2: after TEV cleavage and second Ni-NTA purification input; lane 3: second Ni-NTA flowthrough; lane 4: second Ni-NTA wash; lane 5: Ni-NTA elute; lane 6: Heparin input; lane 7: Heparin flowthrough; lane 8-14: elution peak. [G]: Final protein*

3.5 Full length proteins

Full length LARP4A (Figure 48) and LARP4B (Figure 49) were purified to study the protein interaction with PABPC1. Most of the constructs being studied are only within the N-terminal domain, and the C-terminal domain has not been considered, however there are known regions in the CTD which interacts with PABPC1 and other proteins such as RACK1.

Full length LARP4A (Figure 48) contains only a His-SUMO-tag. The protein appears to be expressed in the soluble fraction and has a molecular weight between 100 and 150 kDa (Figure 48A), the Nickel IMAC column has the Full length LARP4A band at the top, however there are many other bands below it that could be contaminants and/or degradation (Figure 48C). Notable bands include the one in 70 kDa, a double band around 50 kDa and a band at 20 kDa. We decided not to cleave the protein since there are too many alternative species during all stages of the purification, and cleavage using the SUMO Protease usually renders the protein to be insoluble (Hickey, Wilson and Hochstrasser, 2012). The protein is put through a Heparin column to separate the protein from the nucleic acids, and to try to separate the contaminants and degradation products based on a difference in charge. Using the heparin column, it seems like the pure FL-protein could be selected out (Figure 48F, lanes 7-9). The fractions from the Heparin elution containing only the FL-protein are pooled and dialysed overnight to the final buffer and concentrated. The concentrated protein seemed to be a single band, between 100 and 150 kDa.

LARP4A FL:

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIK
KTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGSMLLFVE
QVASKGTGLNPNAKVWQEIAPGNTDATPVTHGTESSWHEIAATSGAHPEGNAELSEDICKEYEV
MYSSSCETTRNTTGIEESTDGMILGPEDLSYQIYDVSGESNSAVSTEDLKECLKKQLEFCFSRENL
SKDLYLISQMDSDQFIPIWTVANMEEIKKLTTDPDLILEVLRSSPMVQVDEKGEKVRPSHKRCIVI
LREIPEXTPIEEVKGLFKSENCPKVISCEFAHNSNWYITFQSDTDAQQAFKYLREEVKTFQGKPIM
ARIKAINTFFAKNGYRLMDSSIYSHPIQTQAQYASPVFMQPVYNPHQQYSVYSIVPQSWSPNPTP
YFETPLAPFPNGSFVNGFNSPGSYKTNAAAMNMGRPFQKNRVKPQFRSSGGSEHSTEGSVSLGD
GQLNRYSSRNFPAERHNPTVTGHQEQTYLQKETSTLQVEQNGDYGRGRRTLFRGRRRREDDRIS
RPHPSTAESKAPTPKFDLLASNFPPLPGSSSRMPGELVLENRMSDVVKGVYKEKDNEELTISCPVP
ADEQTECTSAQQLNMSTSSPCAAELTALSTTQQEKDLIEDSSVQKDGLNQTTIPVSPPSTTKPSRA
STASPCNNNINAATAVALQEPRKLSYAEVCQKPPKEPSSVLVQPLRELRSNVVSPTKNEDNGAPE
NSVEKPHEKPEARASKDYSGFRGNIIPRGAAGKIREQRRQFSHRAIPQGVTRRNGKEQYVPPRSP
K

(His-Tag) (SUMO-Tag) (Construct)

Molecular Weight (Da): 94746.97

Extinction Coefficient (M^{-1} cm^{-1}): 61770

Theoretical pI: 6.18

Molecular Weight post-cleavage (Da): 80606.22 (the protein was not cleaved)

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 60280

Theoretical pI post-cleavage: 6.20

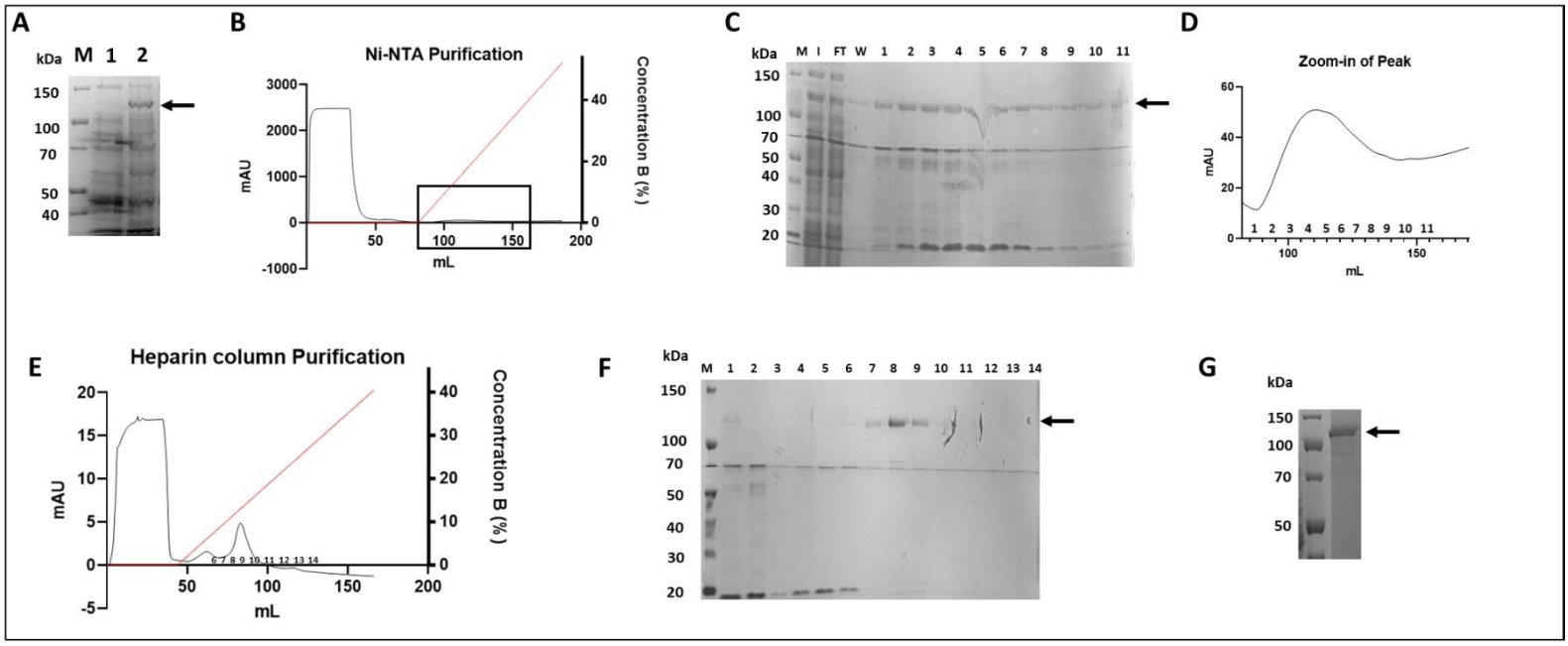


Figure 48. Full length LARP4A expression and purification. *[A]: Expression of FL LARP4A protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of Heparin column purification. [F]: Heparin column purification, lane M: protein ladder; lane 1: Heparin purification input; lane 2: Heparin flowthrough; lane 3: Heparin wash; lane 6-14: elution peak. [G]: Final protein*

Full length LARP4B (Figure 49) contains only a His-SUMO-tag. The protein appears to be expressed in the soluble fraction and is between 100 and 150 kDa (Figure 49A), the Nickel IMAC column has the Full length LARP4B band at the top, however there are many other bands below it that could be contaminants and/or degradation (Figure 49C). Notable bands include the one in 70 kDa, a double band between 50 and 70 kDa, a band at 35 kDa and one below 30 kDa. The protein is decided not to be cleaved since there is too many alternative species, and cleavage using the SUMO Protease usually renders the protein to be insoluble (Hickey, Wilson and Hochstrasser, 2012). The protein is put through a Heparin column just like LARP4A to separate the protein from the nucleic acids, and to try to separate the contaminants and degradation products based on a difference in charge. Using the heparin column, it seems like the pure FL-protein could be selected out (Figure 49F, lanes 9-13). The fractions from the Heparin elution containing only the FL-protein are pooled and dialysed overnight to the final buffer and

concentrated. The concentrated protein seemed to be a single band, between 100 and 150 kDa.

LARP4B FL:

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIK
KTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGSMTSDQ
DAKVVAEPQTQRVQEGKDSAHLMNGPISQTTSQTSSIPPLSQVPATKVSELNPNAEVWGAPVLH
LEASSAADGVSAAWEEVAGHHADRGPQGSDANGDGDQGHENAALPDPQESDPADMNALALGP
SEYDSLPENSETGGNESQPDSQEDPREVLKKTLEFCLSRENLASDMYLISQMDSDQYVPITTVAN
LDHIKKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQNRCIVILREISESTPVEEVEALFKGDNLP
KFINCEFAYNDNWFITFETEADAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPKNGFRPLD
VSLYAQQRYATSFYFPPMYSPQQQFPLYSLITPQTWSATHSYLDPPLVTPFPNTGFINGFTSPAFKP
AASPLTSLRQYPPRSRNPSKSHLRHAIPSAERGPGLLESPSIFNFTADRLINGVRSPQTRQAGQTRT
RIQNPSAYAKREAGPGRVEPGSLESSPGLGRGRKNSFGYRKKREEKFTSSQTQSPTPPKPPSPSFEL
GLSSFPPLPGAAGNLKTEDLFENRLSSLIIGPSKERTLSADASVNTLPVVVSREPSVPASCAVSATY
ERSPSPAHLPDDPKVAEKQRETHSVDRLPSALTATACKSVQVNGAATELRKPSYAEICQRTSKEP
PSSPLQPQKEQKPNTVGCGKEEKKLAEPAERYREPPALKSTPGAPRDQRRPAGGRPSPSAMGKR
LSREQSTPPKSPQ

(His-Tag) (SUMO-Tag) (Construct)

Molecular Weight (Da): 94692.76

Extinction Coefficient (M^{-1} cm^{-1}): 50310

Theoretical pI: 6.31

Molecular Weight post-cleavage (Da): 80552.01 (the protein was not cleaved)

Extinction Coefficient post-cleavage (M^{-1} cm^{-1}): 48820

Theoretical pI post-cleavage: 6.48

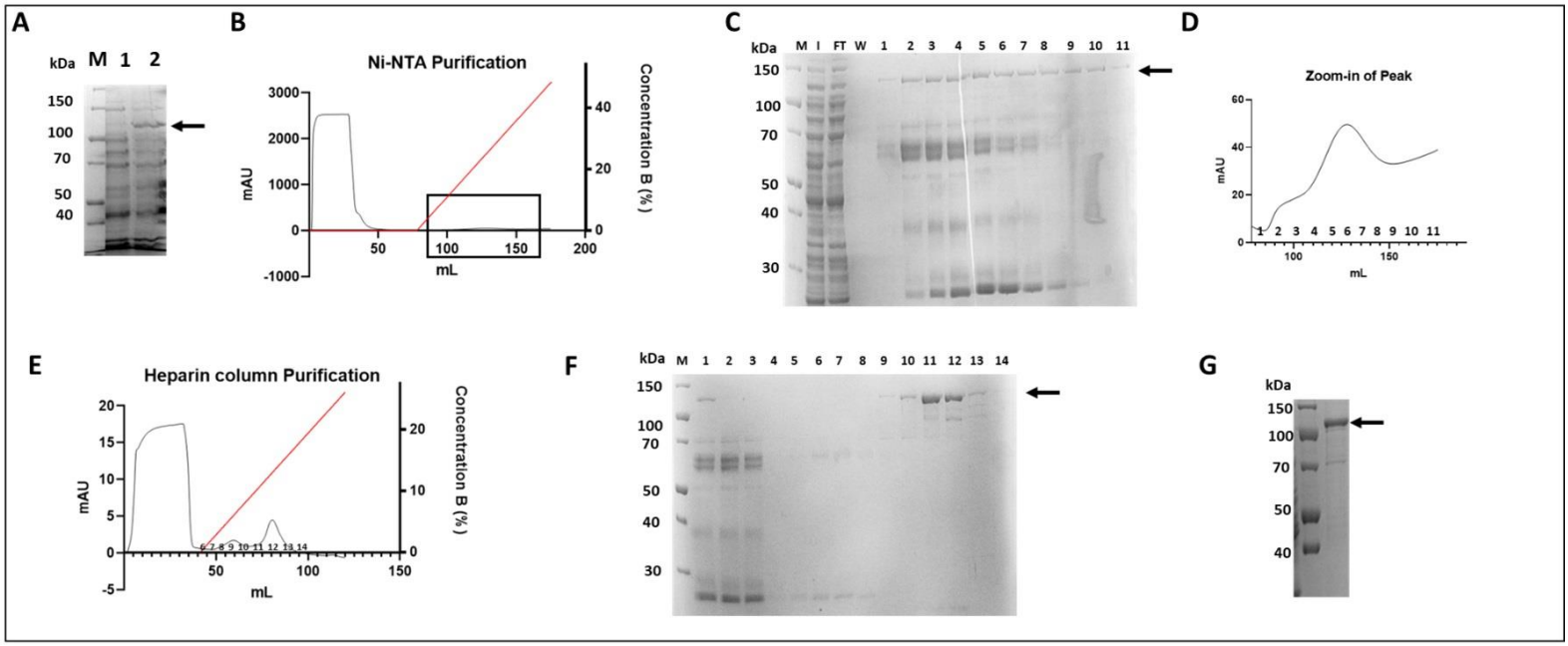


Figure 49. Full length LARP4B expression and purification. *[A]: Expression of FL LARP4B protein, lane M: protein ladder; lane 1: before IPTG induction; lane 2: after IPTG induction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of Heparin column purification. [F]: Heparin column purification, lane M: protein ladder; lane 1: Heparin purification input; lane 2: Heparin flowthrough; lane 3: Heparin wash; lane 6-14: elution peak. [G]: Final protein*

3.6 Other proteins

LARP4A and LARP4B are known to bind the MLLE domain of PABPC1 (Yang *et al.*, 2011; Grimm *et al.*, 2020) using a variant PAM2w domain located on the N-terminus. One of the aims of the project is to test the association of the proteins via a pulldown assay. Wild type LARP4A and LARP4B have been tested for their interaction with MLLE, and then by mutating the PAM2w domain, it is known that LARP4A loses the ability to bind to MLLE (Yang *et al.*, 2011), however this has not been tested with LARP4B.

His-SUMO MLLE construct has a His-tag followed by a SUMO-tag, purification of this construct only involves one round of nickel IMAC (Figure 50). The protein seems to be abundant after the purification, located just under 30 kDa. After dialysis in the final buffer and concentrated the final gel shows a single band just below 30 kDa and looks pure.

His SUMO MLLE:

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSS
EIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHR
EQIGGSPLTASMLASAPPQEQKQMLGERLFPLIQAMHPTLAGKITGMLLEIDNSELLHM
LESPESLRSKVDEAVAVLQAHQAKEAAQKA

(His-Tag) (SUMO-Tag) (Construct)

Molecular Weight (Da): 23141.23

Extinction Coefficient (M^{-1} cm^{-1}): 1490

Theoretical pI: 5.49

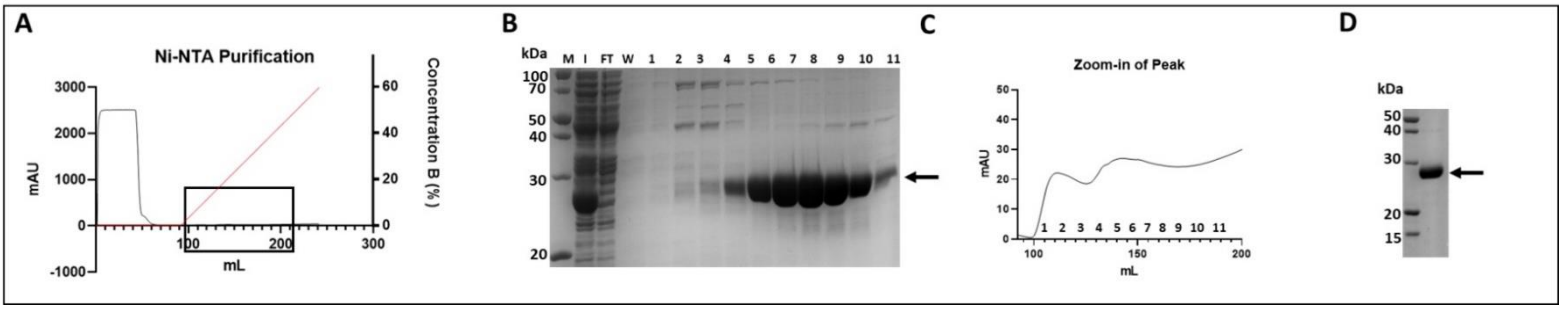


Figure 50. His-SUMO MLLE purification. *[A]: Chromatogram of Ni-NTA purification. [B]: Ni-NTA protein purification of PABPC1 His-SUMO MLLE protein, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [C]: elute fraction peaks zoom in from Ni-NTA purification. [D]: Final protein*

The SUMO tag alone is purified as a control for our protein-protein studies to show the SUMO tag alone does not bind any protein by itself. The purification is shown in Figure 51. The protein appears to be expressed in the soluble fraction and is just above 20 kDa (Figure 51A), the Nickel IMAC column has one main smeared band around the same molecular weight, (Figure 51C). The sample is pooled and dialysed overnight in 5 L nickel dialysis buffer and put through a DEAE column to try to eliminate some of the contaminants as shown in the Nickel column purification gel, the flowthrough from the DEAE is dialysed overnight to the final buffer and concentrated. The concentrated protein

seemed to be a single band, with a lighter band just above 20 kDa that seemed to be a contaminant present in very low quantity. (Figure 51G)

SUMO alone:

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSS

EIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHR

EQIGGSEFELRRQACGRTRAPPPPPLRSGC

(His-Tag) (SUMO-Tag)

Molecular Weight (Da): 16835.87

Extinction Coefficient (M^{-1} cm^{-1}): 1490

Theoretical pI: 6.53

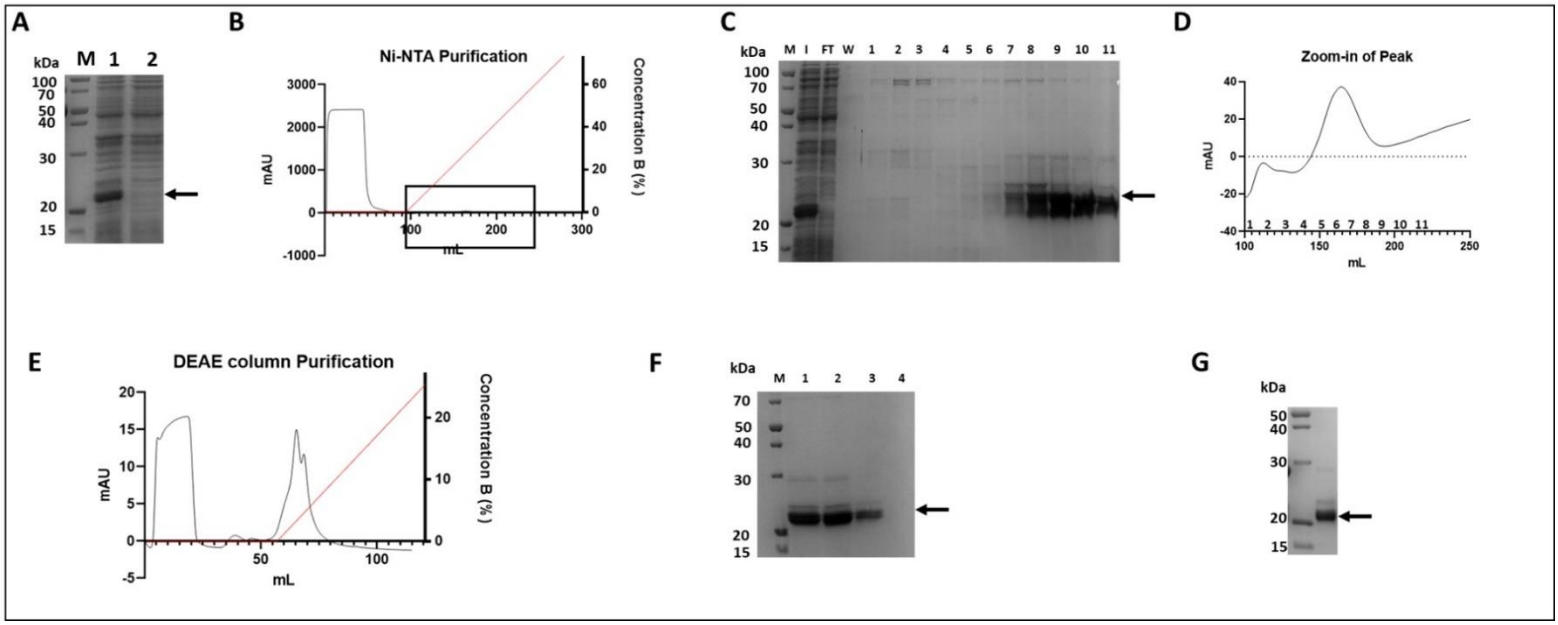


Figure 51. SUMO protein expression and purification. *Expression of SUMO protein, lane M: protein ladder; lane 1: after IPTG induction; lane 2: before IPTG induction. [B]: Chromatogram of Ni-NTA purification. [C]: Ni-NTA protein purification, lane M: protein ladder; lane I: input; lane FT: flowthrough; lane W: wash, lanes 1-11: elute fractions. [D]: elute fraction peaks zoom in from Ni-NTA purification [E]: Chromatogram of DEAE column purification. [F]: DEAE column purification, lane M: protein ladder; lane 1: DEAE input; lane 2: DEAE flowthrough; lane 3: wash; lane 4: elute peak. [G]: Final protein*

3.7 Conclusions

Many truncation mutants and point mutants needed to be generated in order to systematically identify the RNA/protein binding locus of LARP4A and LARP4B. LARP4A is used many times as the 'control' since a lot more is known about the protein, how it is expressed, the conditions for purification etc. LARP4B required more troubleshooting, and the purifications were repeated multiple times to reach a satisfactory state. Most of the problems faced in expression and purification is having insoluble protein, contaminants or sample degradation in the end, which defeats the purpose of purification. To improve the expression, I have done expression tests for all the constructs before moving on to a larger culture induction. Temperatures such as 18°C, 25°C, and 37°C was tested; induction times were also done in 3 hour/overnight timings. In certain cases, re-cloning the construct (see LARP4B NTD 1-328) has been done due to degradation. Other methods used to reduce degradation are to reduce dead times during the purification, since proteases in the cell can rapidly degrade proteins if left for too long (Gopal and Kumar, 2013).

I was involved in the later stages of a publication for LARP4A (Cruz-Gallardo *et al.*, 2019) when I first arrived at the lab. A lot of emphasis is put on the comparison of these two proteins that are so similar in amino acid sequence, but nevertheless bind to different target RNA. Using these mutants expressed and purified above, the RNA binding signature of LARP4B towards its target AU-rich/CKB RNA has been characterised in the next section, and the locus of RNA specificity in LARP4B has been explored. At the same time, the PAM2w motif of LARP4B, known as the protein-binding locus (which in LARP4A proved to be useful in both protein and RNA interactions) was studied.

Chapter 4. Interaction studies

4.1 Overview

LARP4A has a role in promoting mRNA stability and enhancing translation by its association with PABP and oligoA RNA (Yang *et al.*, 2011) and LARP4B also interacts with PABP and AU-rich/CKB RNA to promote mRNA accumulation, and translation stimulation (Schäffler *et al.*, 2010). The work focuses on understanding the molecular mechanisms and binding affinity of the protein with its targets. LARP4A has been thoroughly studied in the paper (Cruz-Gallardo *et al.*, 2019), where it has been shown to bind to oligoA using the PAM2w N-terminus, in which is located in an intrinsically disordered region that do not contain any recognisable RNA-binding domains and never known to have any associations with RNA. The La-module of LARP4A, initially anticipated to be the main locus of RNA binding has been found only to have a minor role. LARP4A PAM2w has also been shown to interact with its protein binding partner – PABPC1, and the double mutant L15AW22A not only disrupts the binding to PABPC1, but also affects LARP4A's ability to bind RNA (Cruz-Gallardo *et al.*, 2019). These new findings for LARP4A have given insight to how it interacts with its partners, and couple of questions falls upon LARP4B. Is the La-module of LARP4B unconventional like LARP4A? Does the PAM2w play a role in RNA binding, or it is only functional to bind to PABPC1? How does the PAM2w double mutant change the binding activity of LARP4B? Where exactly is the binding locus of LARP4B for RNA, and does it have more RNA targets? These questions can be answered using a systematic approach, by creating truncation mutants and mutants spanning known LARP4B domains, using biochemical and biophysical techniques such as Electrophoretic mobility shift assay (EMSAs), Microscale thermophoresis (MST) and Circular Dichroism (CD), each region and domain of LARP4B will be tested against AU-rich/CKB RNA, which was first

discovered to be a target using PAR-CLIP analysis (Küspert *et al.*, 2015), interestingly, they have found many more RNA targets for LARP4B, but the only RNA verified using *in vitro* assays is CKB, so LARP4B could potentially bind to other RNA targets that have not been discovered.

4.2 A contiguous stretch of the N-terminal region of LARP4A is important in RNA binding

In the investigations of LARP4A interactions with oligoA RNA, several deletion mutants of LARP4A involving the full N-terminal Domain encompassing residues 1-287, and deletion mutants 24-287, 50-287 and 79-287 were employed in EMSA (Cruz-Gallardo *et al.*, 2019). When I joined the lab, I contributed to this research in investigating the importance of the region encompassing residues 1-24 to RNA binding.

Given that the PAM2w region (residues 13-26) is imperative in RNA binding as shown by previous experiments(Cruz-Gallardo *et al.*, 2019), the question was whether the short fragment spanning residues 1-24 and encompassing the PAM2w would be sufficient for oligoA RNA binding, I cloned 2 deletion mutants, spanning residues 1-50 and 1-79. These are longer than just the first 24 residues, as it was hypothesised that the longer fragments may be more stable for *E.coli* expression.

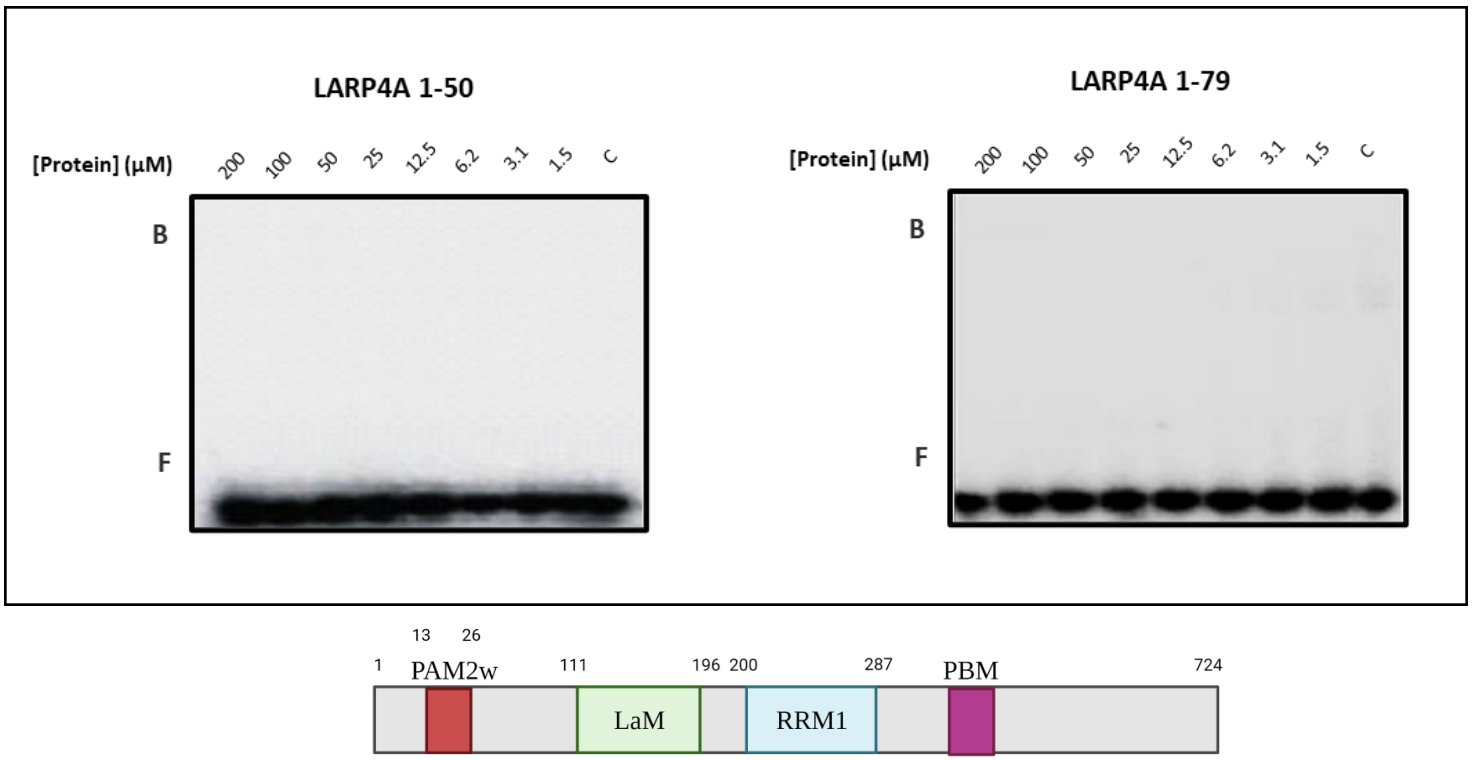


Figure 52. LARP4A 1-50 and LARP4A 1-79 do not bind to oligoA RNA (5'-AAAAAAAAAAAAAAA-3') alone. *EMSA binding assays of LARP4A 1-50 and LARP4A 1-79 with 32P-oligoA15. Representative autoradiograms are shown for 1-50 (left) and 1-79 (right). The protein concentrations used were 0, 1.5, 3.1, 6.2, 12.5, 25, 50, 100 and 200 μM. Bound (B) and free (F) RNA populations are labelled. A diagram of LARP4A is shown with amino acid residues numbering the domains.*

LARP4A constructs 1-50 and 1-79 containing the PAM2w domain (present in residues 13-26) both showed no binding towards the target RNA (Figure 52), even in concentrations of 200 μM, which might indicate that PAM2w alone can't bind to the RNA alone and requires the help of residues in the N-terminal domain for binding. To determine the secondary structure of the constructs 1-50 and 1-79, Circular Dichroism was performed.

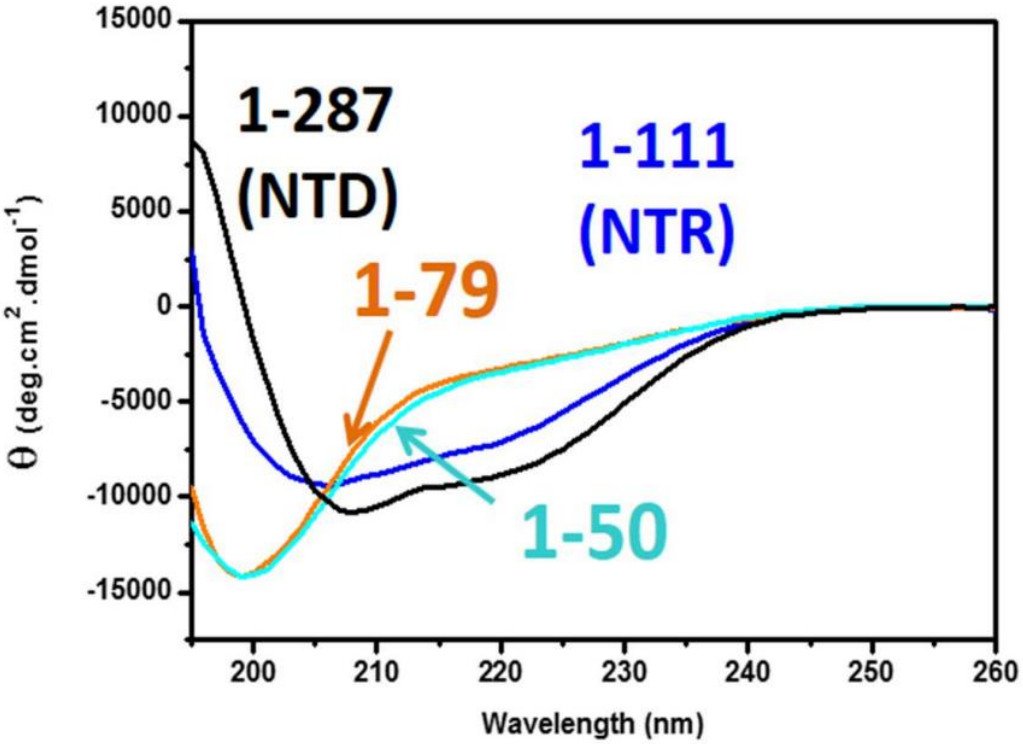


Figure 53. LARP4A 1-50 and 1-79 show a disordered orientation and having characteristics that is more similar to random coils. *Far-UV CD analysis. Far-UV CD spectra of: LARP4A 1-50 (teal), 1-79 (orange) compared with LARP4A NTD (black) and NTR (blue).*

Proteins produce a distinctive shape in the CD spectra due to their secondary structure (Greenfield, 2007) using this we can deduce the secondary structure of unknown proteins such as 1-50 and 1-79. The secondary structure of the newly obtained short constructs was compared with the previously known NTD and NTR of LARP4A (Figure 53). The spectrum of α-helices in general have two negative bands of similar magnitude at 222 and 208 nm, and a positive band at around 190 nm. On the other hand, for β-sheets they have a negative band between 210-220 nm and a positive band between 195-200 nm (Greenfield, 2007). The NTR does not contain any commonly known structured domains but according to the CD spectra it is not totally unstructured but rather contain a mixture of disordered regions and secondary structures, both α-helical and β-strand, this structure could potentially be caused by the variant PAM2w motif (13-26). While the signatures of the NTD and NTR display evidence of clear secondary structure with the combination

of α-helices and β-sheets, the smaller constructs lose that and show a disordered orientation and having characteristics that is more similar to random coils. This suggested that a stretch of residues towards the NTD is required to maintain the secondary structure of LARP4A which is required for RNA binding and suggests why there is a lack of binding in the EMSA since there is a lack of secondary structure.

4.3 LARP4B N-terminal domain is required for maximum CKB RNA-binding

To investigate whether the LARP4B has RNA-binding specificity for AU-rich/CKB RNA, Electromobility Shift assays (EMSAs) of LARP4B N-terminal Domain (NTD), N-terminal Region (NTR) and La-module were performed with the target AU-rich RNA (CKB) to determine the contribution of the separate domains of LARP4B (Figure 54).

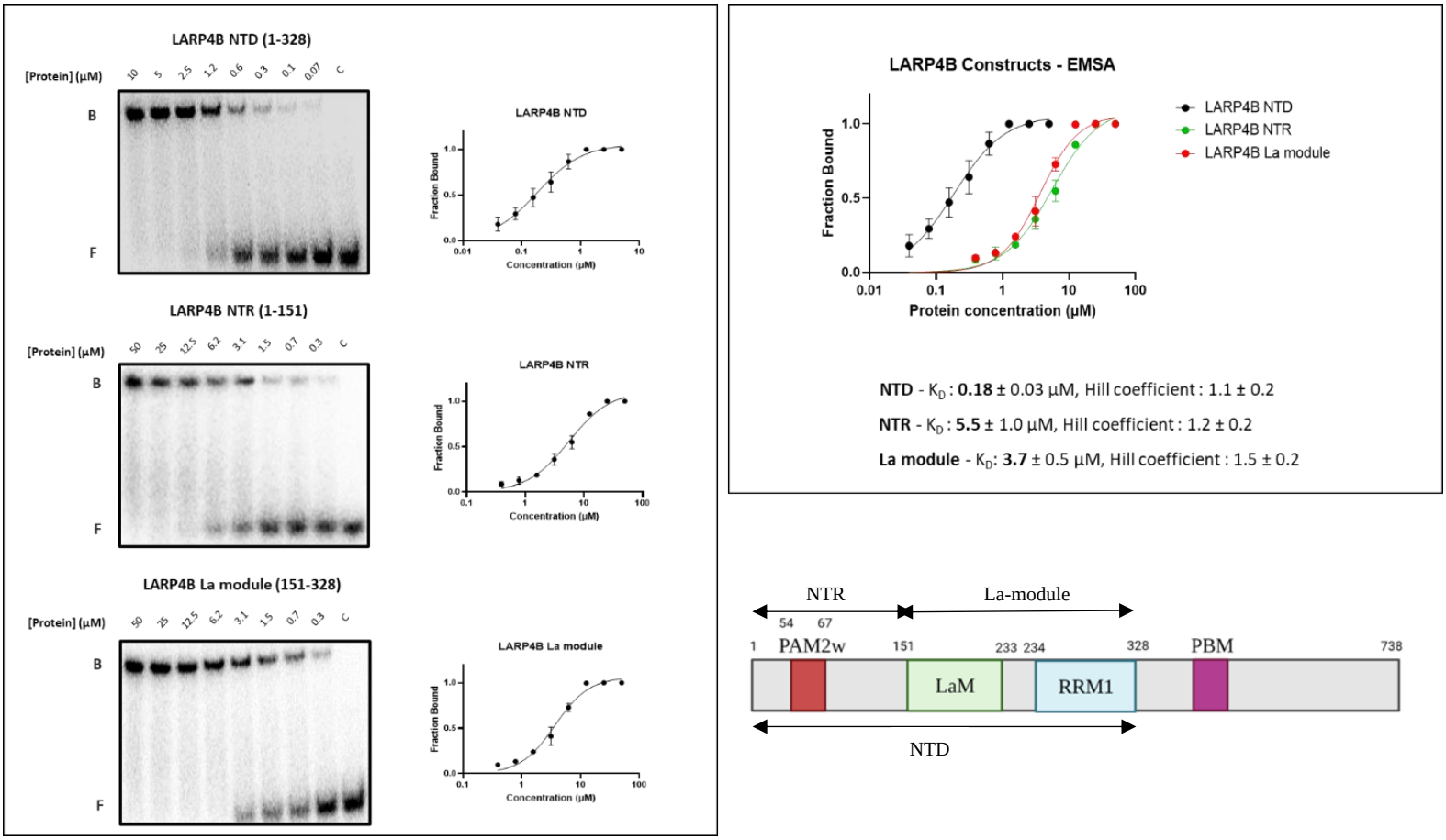


Figure 54. EMSA binding assays of LARP4B N-terminal domain (NTD), N-terminal region (NTR) and La-Module with 32P-AU-rich RNA (5'-UGGUGAGUUUAUUUUUUUGA-3'). *Representative autoradiograms are shown for LARP4B NTD (top), NTR (centre) and La-Module (bottom) and their binding curves (right). The protein concentrations used were 0, 0.07, 0.1, 0.3, 0.6, 1.2, 2.5, 5 and 10 μM for NTD and 0, 0.3, 0.7, 1.5, 3.1, 6.2, 12.5, 25 and 50 μM for La-module and NTR. Bound (B) and free (F) RNA populations are labelled. At least three biological replicates were used to calculate average values for KD. For each concentration, the fraction bound has been reported with error bars reporting standard deviation. KDs have been calculated from a non-linear regression curve with Hill fitting, reported with an error range that denotes 95% confidence interval, as determined by the fitting of the data to the indicated equation in Prism 9.*

LARP4B NTD EMSAs was performed with a protein starting concentration of 10 μM, serially diluted 1:1 after each subsequent data point, whereas the NTR and the La-module after optimising the experiment used a higher starting concentration of 50 μM since the NTD binds considerably tighter than the other two. Results show that LARP4B NTD binds CKB RNA with a dissociation constant (K_D) of 0.18 ± 0.03 μM and a Hill coefficient of 1.1 ± 0.2, LARP4B NTR binds CKB RNA with a dissociation constant (K_D) 5.5 ± 1.0 μM and a Hill coefficient of 1.2 ± 0.2, and LARP4B La-module binds CKB RNA with a dissociation constant (K_D) 3.7 ± 0.5 μM and a Hill coefficient of 1.5 ± 0.2.

(For each concentration, K_Ds have been calculated from a non-linear regression curve with Hill fitting, reported with an error range that denotes 95% confidence interval).

The whole N-terminal Domain that includes both the N-terminal Region and the La-module (see Figure 54, lower right) binds the tightest to CKB, while the La-module alone binds 20 times less tight than the NTD, and the NTR binds 30 times less tight than the NTD. This suggests that the whole of the NTD is required for the maximum binding affinity towards CKB. Although the La-module is historically known as the most important domain for RNA-binding (Maraia *et al.*, 2017; Dock-Bregeon, Lewis and Conte, 2021), in the case of LARP4B, our EMSA results suggest that the NTR is still a vital region that aids the binding, albeit requires the La-module working in conjunction with the N-terminal region to achieve maximum binding. This is further emphasized by the fact that the NTR alone has the weakest affinity towards RNA.

As a comparison to LARP4A, the NTD holds the maximum binding to oligoA RNA, at a dissociation constant of 3.3 μM, and the NTR binds at 4.5 μM, which is similar to NTD, however the La-module showed extremely weak binding (>100 μM) (Cruz-Gallardo *et al.*, 2019), which is a main difference to LARP4B.

The Hill coefficient is a measure of cooperativity (Abeliovich, 2005), it is a coefficient devised by A. V. Hill (1886-1977) to deduce the cooperativity in a binding process, it is commonly used to study the kinetics of reactions that exhibit a sigmoidal behaviour. A Hill coefficient of 1 indicates independent binding, the Hill equation is reduced to its simpler form known as the Michaelis-Menten equation (Choi, Rempala and Kim, 2017), a coefficient greater than 1 suggests that two or more binding sites exist in the protein and that there is positive cooperativity with respect to substrate binding. Positive cooperativity refers to a scenario when the binding of one substrate facilitates the binding of another substrate to the protein, a coefficient of less than 1 indicates negative

cooperativity. Our Hill coefficients of LARP4B all seem to be around 1-1.5, which suggests independent binding with a slight positive cooperativity.

MST is an alternative method used in this study in order to verify results from the EMSA, it is always important in science to reach the same conclusion using several methods that utilises different physical properties. In this way we can validate our results. LARP4B NTD, NTR and La-module were all tested with CKB RNA labelled at the 5′ end with 5-FAM. A representative MST trace for each construct tested is shown in Figure 55, and the subsequent binding plots showing the fractions of protein bound RNA as a function of protein concentration shown in Figure 56.

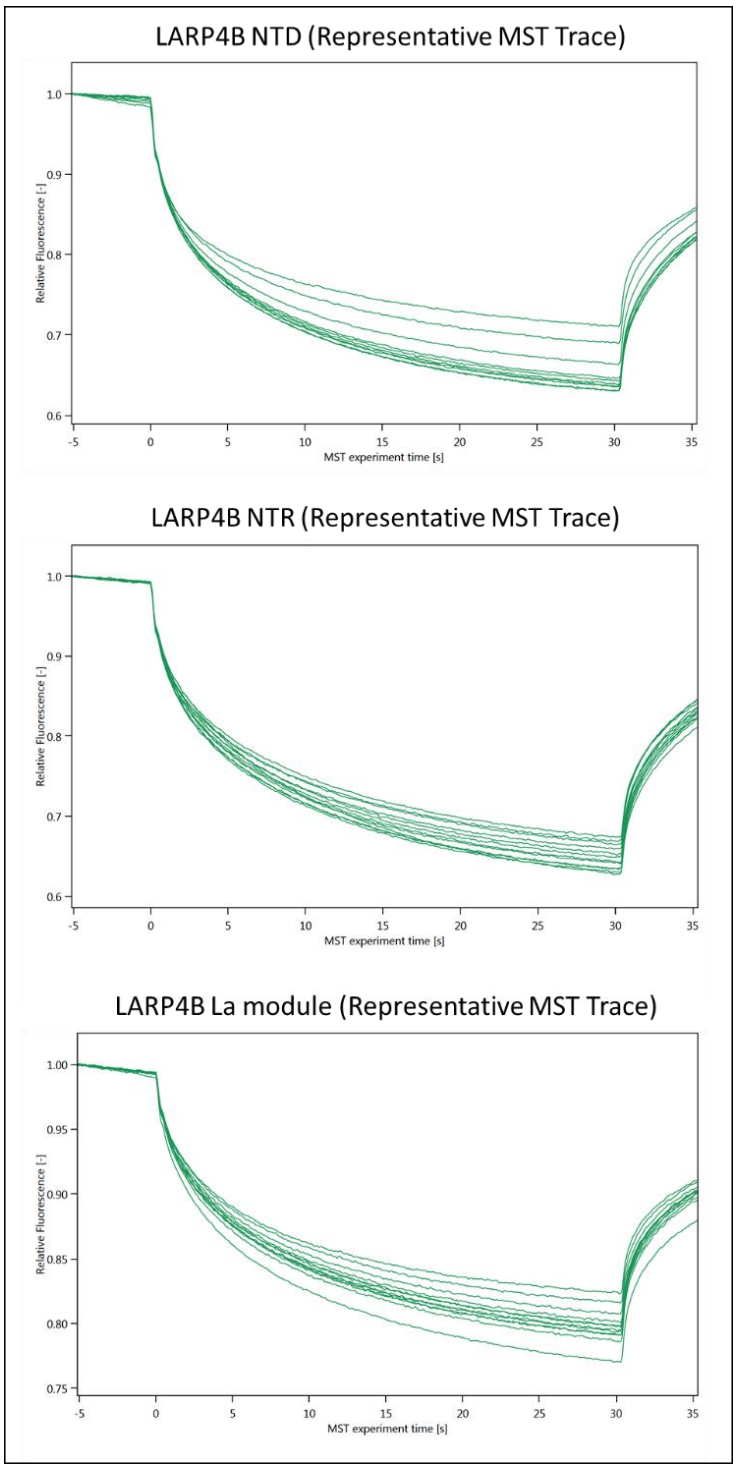


Figure 55. Normalized thermophoretic time-traces from one representative curve of LARP4B NTD, NTR and La-module binding to 5' FAM AU-rich RNA (5'-UGGUGAGUUUAUUUUUUUGA-3')*. Relative fluorescence is plotted against experimental time. When performing an MST experiment, a microscopic temperature gradient is induced by an infrared laser, and temperature related intensity change as well as thermophoresis are detected. Each line represents the change of relative fluorescence of the molecule with different concentrations of the protein, and the RNA concentration is fixed.*

Results show that LARP4B NTD binds CKB RNA with a dissociation constant (K_D) of 0.47 ± 0.11 μM, and a Hill coefficient of 1.2 ± 0.3, LARP4B NTR binds CKB RNA with a dissociation constant (K_D) 6.1 ± 2.9 μM, and a Hill coefficient of 1.0 ± 0.3 and LARP4B La-module binds CKB RNA with a dissociation constant (K_D) of 5.3 ± 1.1 μM and a Hill coefficient of 1.0 ± 0.1. The binding constants derived from the MST and EMSAs have slightly different values, which is expected since EMSAs rely on the rate of RNA migration slowing down when bound to proteins, and MSTs is based on the detection of a temperature-induced change in fluorescence of a target, they are different techniques that utilizes a different way to obtain their results. However, they follow the same trend, which strengthens the conclusion that the LARP4B N-terminal domain is required for maximum CKB RNA-binding.

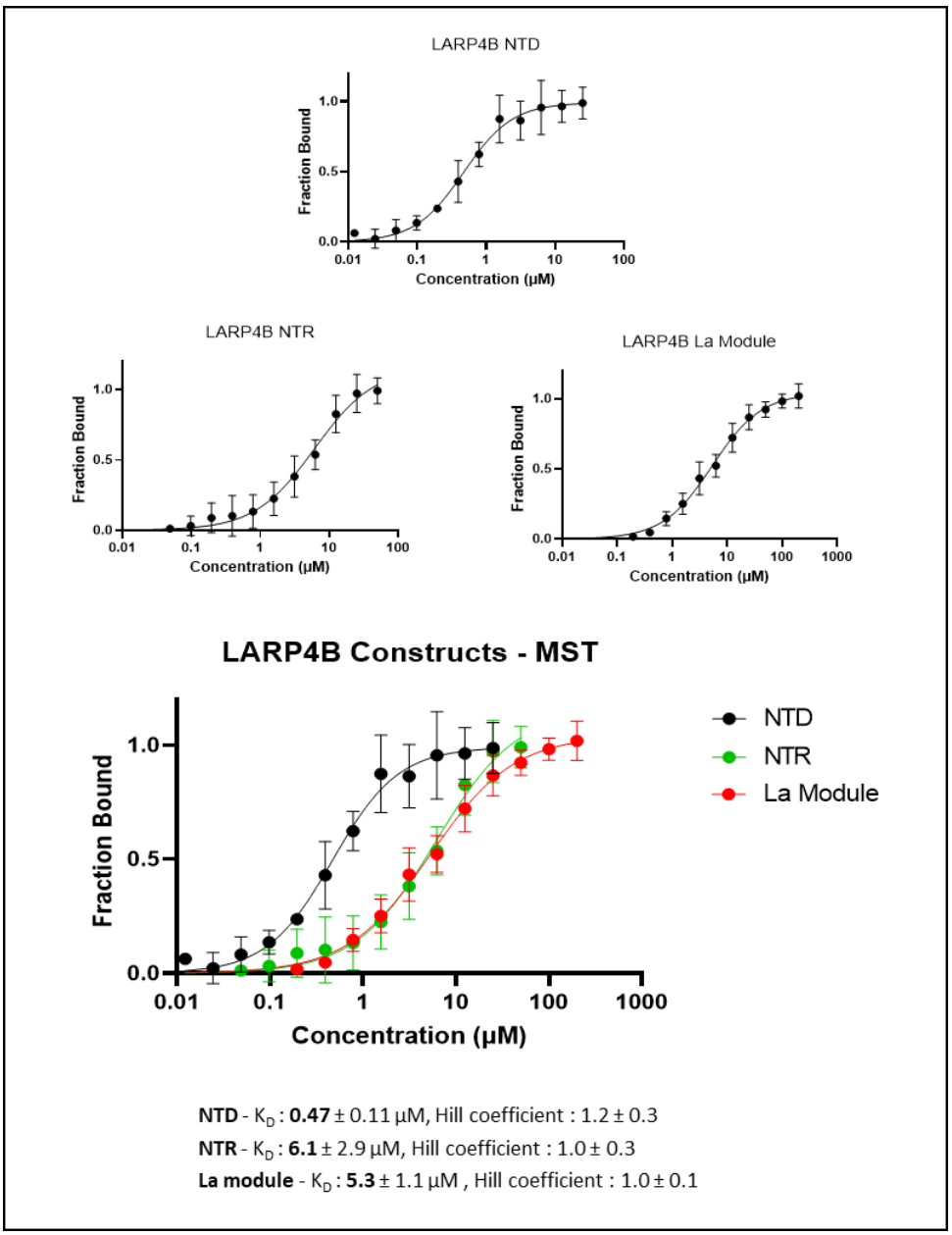


Figure 56. MST binding curves for the interaction of LARP4B NTD, NTR, La-Module and with 5'FAM-AU-rich RNA (5'-UGGUGAGUUUAUUUUUUUGA-3'). *Binding plots showing the fractions of protein-bound RNA as a function of protein concentrations for the three constructs tested. At least three biological replicates were used to calculate average values for K_D. For each concentration, the fraction bound has been reported with error bars reporting standard deviation. K_Ds have been calculated from a non-linear regression curve with Hill fitting, reported with an error range that denotes 95% confidence interval, as determined by the fitting of the data to the indicated equation in Prism 9.*

4.4 Different LARP4B N-terminal deletion mutants shows a different preference for CKB binding

Since LARP4B N-terminal domain is required for maximum CKB RNA-binding, to map the extent of the N-terminal region needed for high affinity for CKB, LARP4B EMSAs of the N-terminal Region truncation mutants (40-328, 71-328 and 95-328) (Figure 57) were performed.

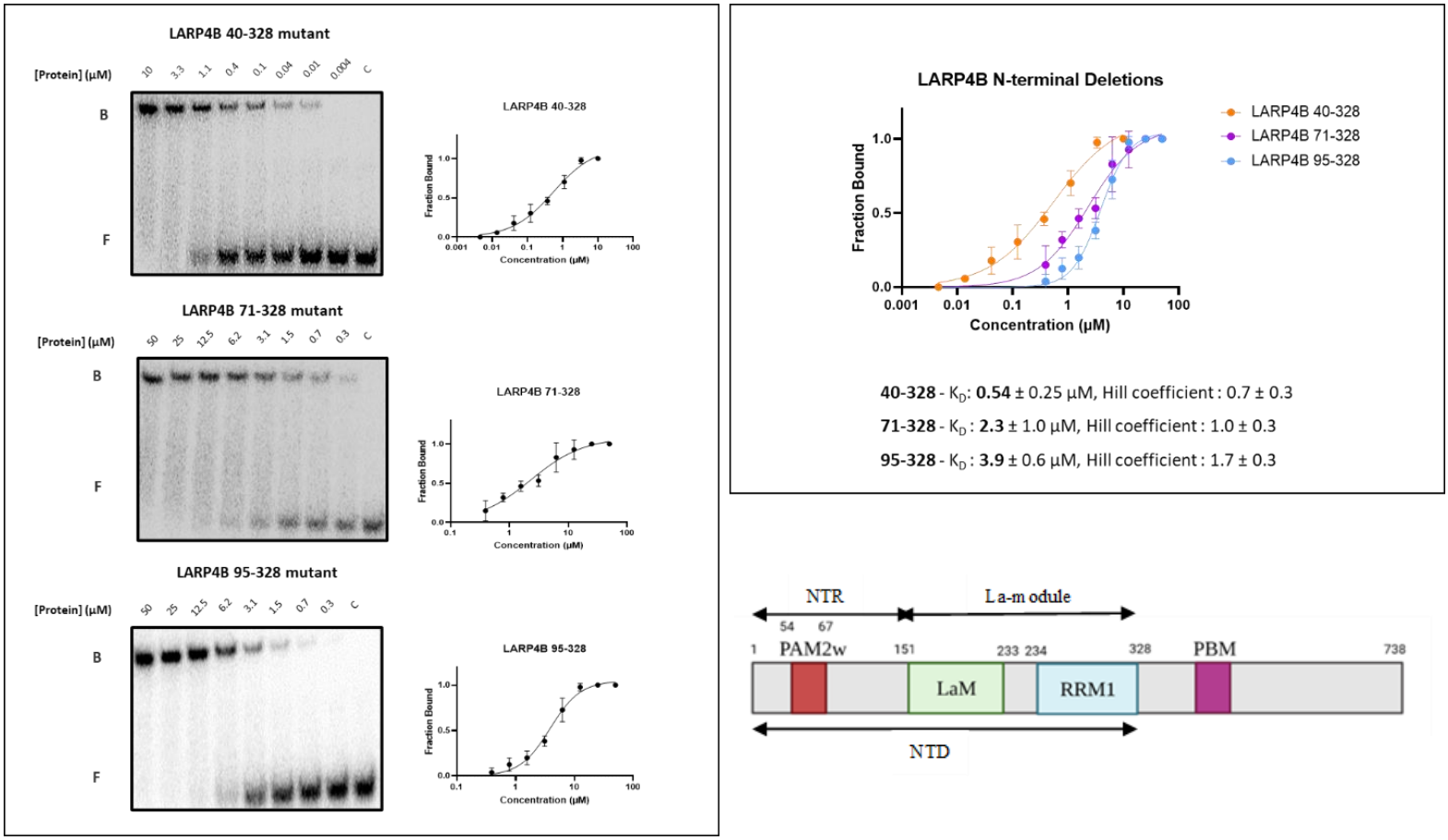


Figure 57. EMSA binding assays of LARP4B N-terminal deletion mutants 40-328, 71-328 and 95-328 with 32P-AU-rich RNA (5'-UGGUGAGUUUAUUUUUUUGA-3'). *Representative autoradiograms are shown for 40-328 (top), 71-328 (centre) and 95-328 (bottom) and their binding curves (right). The protein concentrations used were 0, 0.004, 0.01, 0.04, 0.1, 0.4, 1.1, 3.3 and 10 μM for 40-328 and 0, 0.3, 0.7, 1.5, 3.1, 6.2, 12.5, 25 and 50 μM for 71-328 and 95-328. Bound (B) and free (F) RNA populations are labelled. At least three biological replicates were used to calculate average values for K_D. For each concentration, the fraction bound has been reported with error bars reporting standard deviation. K_Ds have been calculated from a non-linear regression curve with Hill fitting, reported with an error range that denotes 95% confidence interval, as determined by the fitting of the data to the indicated equation in Prism 9.*

The N-terminal truncation mutants were designed by sequence alignment to LARP4A (Cruz-Gallardo et al., 2019) in the context of a comparative study of the RNA binding behaviour of LARP4A and LARP4B. LARP4B contains 40 extra amino acid residues in the N-terminus that is not present in LARP4A, this could be interesting to see if the additional residues allow LARP4B to bind to RNA differently compared to LARP4A.

LARP4B 40-328 binds CKB RNA with a dissociation constant (K_D) of 0.54 ± 0.25 μM, and a Hill coefficient of 0.7 ± 0.3, LARP4B 71-328 binds CKB RNA with a dissociation constant (K_D) of 2.3 ± 1.0 μM, and a Hill coefficient of 1.0 ± 0.3, LARP4B 95-328 binds CKB RNA with a dissociation constant (K_D) of 3.9 ± 0.06 μM, and a Hill coefficient of 1.7 ± 0.3. It was noticed that the Hill coefficients are quite different, but it could be less of an issue with cooperativity and more to do with sample variability and the technique. During MST runs it was noted that the Hill coefficient could be different even from using the same proteins, therefore the Hill coefficient was also used as a control for our experiments, as when we get Hill coefficients of greater than 2 or less than 0.5, or when obtaining an inconsistent Hill coefficient between repeats we can be sure that there was an error in the experiments.

The general trend is that as more residues are truncated in the N-terminal region, the less strongly the protein binds CKB. Binding of LARP4B 40-328 ($K_{D:}$ 0.54 μM) to CKB is almost comparable to the NTD ($K_{D:}$ 0.18 μM), suggesting that the first 40 residues of LARP4B do not affect the binding drastically. However, our binding experiments indicate that at least for CKB these N-terminal 40 residues are not significantly implicated in the molecular recognition. Trimming an additional 30 N-terminal residues, in the LARP4B 71-328 mutant ($K_{D:}$ 2.3 μM), causes a further loss of binding affinity when compared to 40-328 of about 4-fold ($K_{D:}$ 0.54 μM), and 5-fold compared with NTD, $K_{D:}$ 0.47 μM.

Trimming the N-terminal region to residue 95 has a similar binding affinity ($K_{D:}$ 3.9 μM) to the La-module (151-328) ($K_{D:}$ 5.3 μM). So, a reduction of 1.5-fold compared to 71-

328 ($K_{D:}$ 2.3 μM), and a reduction of 7-fold compared to 40-328 ($K_{D:}$ 0.54 μM) and 8-fold compared to NTD ($K_{D:}$ 0.47 μM).

4.5 LARP4B La motif and RRM alone have a weak contribution to CKB binding

LARP4B La-module retains some RNA binding activity (EMSA $K_{D:}$ 3.7 μM) to CKB RNA, contrary to LARP4A La-module ($K_{D:}$ >100 μM) to oligoA RNA. La-modules are known to bind quite tightly to their target RNA e.g. LARP6 $K_{D:}$ 48 nM (Martino *et al.*, 2015b), LARP1: $K_{D:}$ 40 nM (Al-Ashtal *et al.*, 2021). La-modules of LARPs have been extensively studied in other systems where biophysical studies of LARP6 provided initial evidence to suggest that the two domains of the La-module may be capable of synergistic RNA recognition via different topological arrangements in different LARPs (Martino *et al.*, 2015b).

To understand the contribution of the individual domains within the La-module towards RNA binding, La motif and the RRM alone have been tested in EMSAs towards CKB RNA (Figure 55). LaM has a K_D of 30.9 ± 13.6 μM, with a Hill coefficient of 1.0 ± 0.3, and the RRM alone has a K_D of 13.0 ± 2.2 μM, with a Hill coefficient of 1.5 ± 0.3. When comparing to the La-module, the LaM alone binds 8 times weaker than the La-module, and the RRM alone binds 3.5 times weaker than the La-module. This suggests that both

domains are important towards RNA binding, however the contribution of the RRM seems to be more important with respect to that of the LaM.

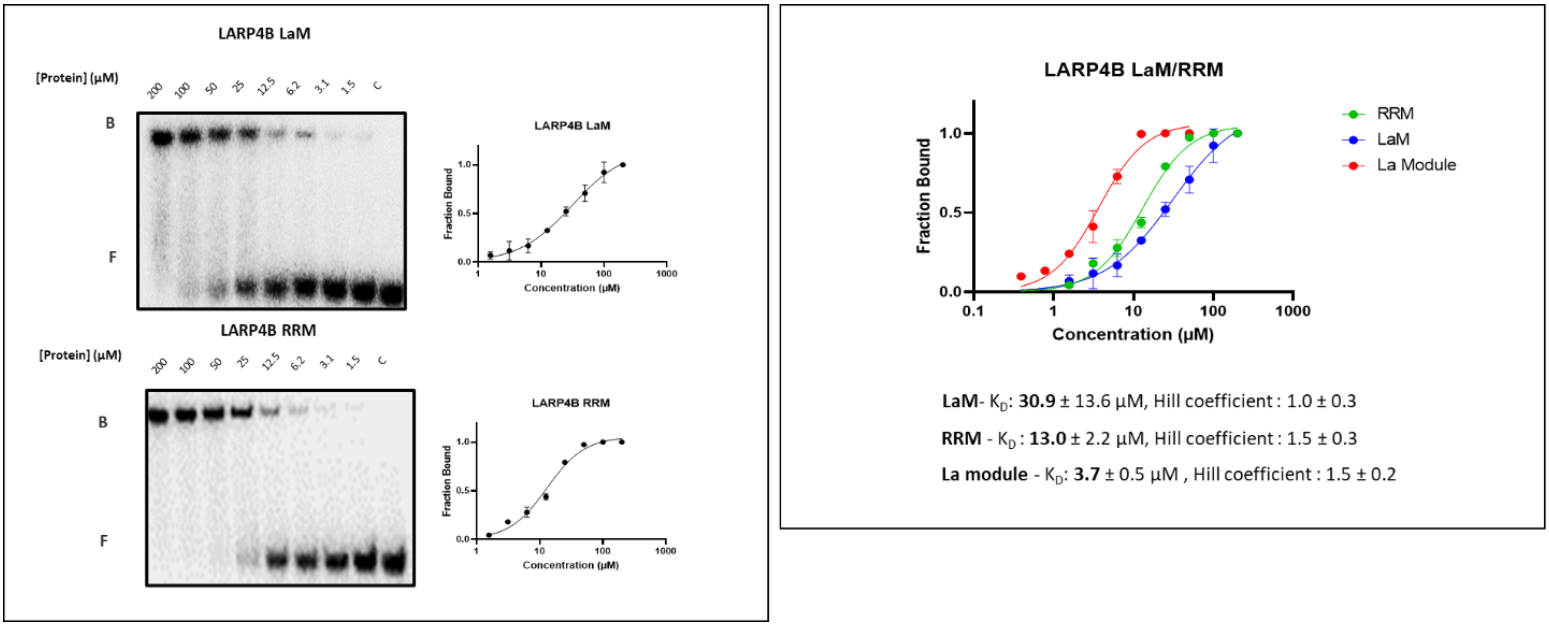


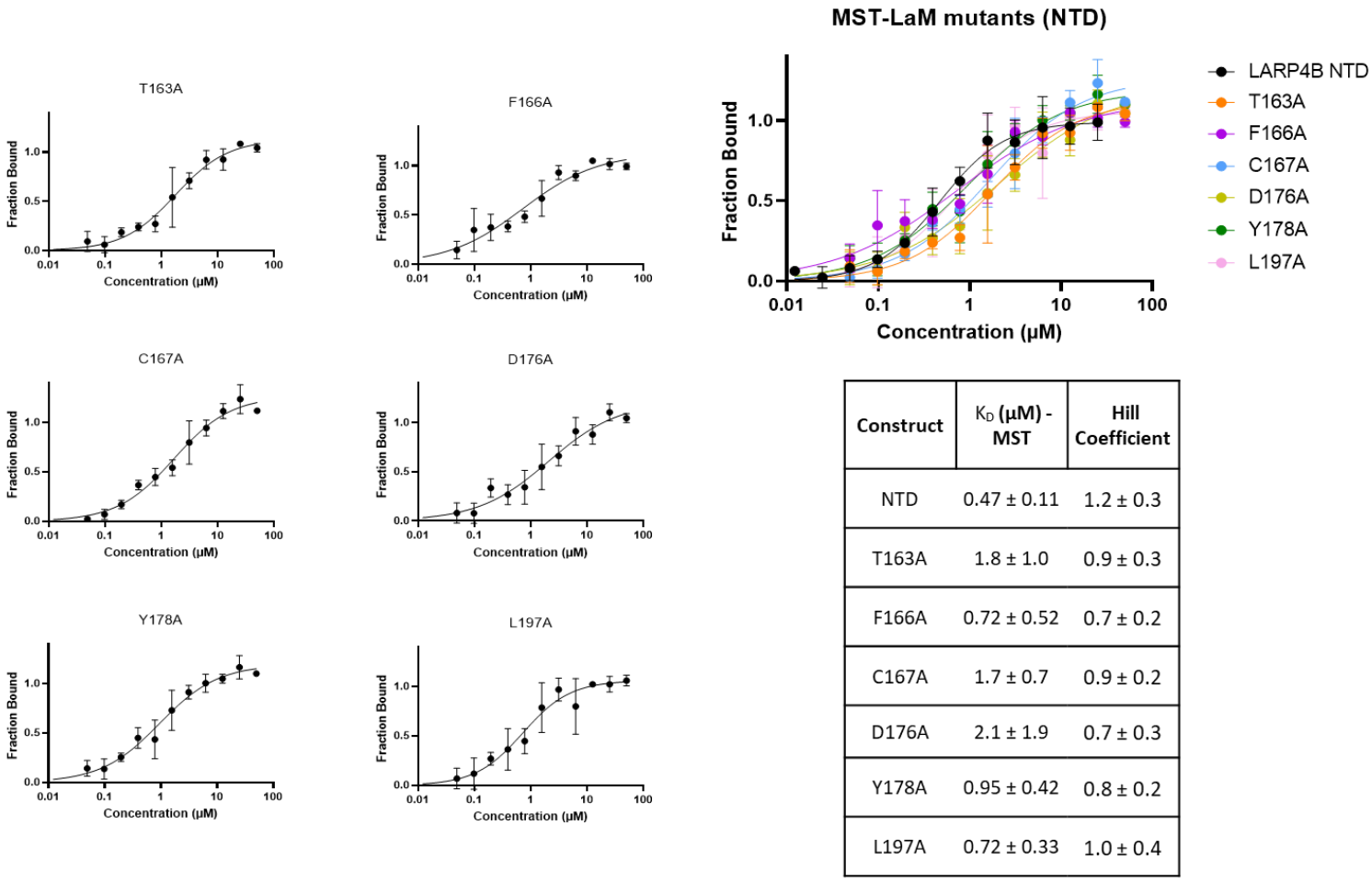
Figure 58. EMSA binding assays of LARP4B La motif (LaM) and RRM with 32P-AU-rich RNA (5'-UGGUGAGUUUAUUUUUUUGA-3'). *Representative autoradiograms are shown for LaM (top) and RRM (bottom) and their binding curves (right). The protein concentrations used were 0, 1.5, 3.1, 6.2, 12.5, 25, 50, 100 and 200 μM. Bound (B) and free (F) RNA populations are labelled. At least three biological replicates were used to calculate average values for K_D. For each concentration, the fraction bound has been reported with error bars reporting standard deviation. K_Ds have been calculated from a non-linear regression curve with Hill fitting, reported with an error range that denotes 95% confidence interval, as determined by the fitting of the data to the indicated equation in Prism 9.*

4.6 LARP4B La motif point mutants does not influence the RNA-binding of LARP4B

The La motif (LaM), originally discovered in La, is a modified 'winged-helix motif,' that contains three α-helices (Alfano *et al.*, 2004; Dong *et al.*, 2004) which allow the formation of a hydrophobic cavity that provides La, LARP7 and LARP6 with RNA-binding capabilities(Dong *et al.*, 2004; Martino *et al.*, 2015b; Uchikawa *et al.*, 2015). La is the most studied protein (Wolin and Cedervall, 2002; Cardinali *et al.*, 2003; Bayfield, Yang and Maraia, 2010; Maraia *et al.*, 2017). In this cavity, specific and non-specific contacts are formed between La and UUU-3'-OH RNA ligand and in particular by six key conserved amino acid residues present in the La motif: Q20, Y23, Y24, D33, F35 and

F55. Using biophysical and mutagenesis techniques, this hydrophobic pocket proved to be important in LARP7, having an importance in interaction to its target RNA which is the 7SK RNA UUU-3'OH (Uchikawa *et al.*, 2015), and also in LARP6 (Martino *et al.*, 2015c), where it binds to the 48 nt stem-loop of the 5' UTR in collagen α1 mRNA.

Conservation analysis is one of the most widely used methods for predicting these functionally important residues in protein sequences (Capra and Singh, 2007). The sequence conservation tends to lead us to mutate the key amino acid residues to alanine (see Figure 3, page 20 for the sequence alignment of the LaM and their N-terminal regions in LARPs) and test them in LARP4B-RNA interactions using MST. MST is the technique chosen for the LaM mutants because it is more suited to deal with such a large set of protein, results and repeats could be obtained much faster than using EMSAs, it is also much safer than EMSAs since we do not work with radioactive substances. In this case we were only expecting a drastic change in binding affinity if one residue is more important than the other, so we opted for MST.



Construct	K_D (μM) - MST	Hill Coefficient
NTD	0.47 ± 0.11	1.2 ± 0.3
T163A	1.8 ± 1.0	0.9 ± 0.3
F166A	0.72 ± 0.52	0.7 ± 0.2
C167A	1.7 ± 0.7	0.9 ± 0.2
D176A	2.1 ± 1.9	0.7 ± 0.3
Y178A	0.95 ± 0.42	0.8 ± 0.2
L197A	0.72 ± 0.33	1.0 ± 0.4

Figure 59. MST binding curves for the interaction of LaM point mutants T163A, F166A, C167A, D176A, Y178A and L197A in the context of NTD (1-328) with 5'FAM-AU-rich RNA (5'-UGGUGAGUUUAUUUUUUUGA-3')*. Binding plots showing the fractions of protein-bound RNA as a function of protein concentrations for the six constructs tested. Two biological replicates were used for T163A and at least three for the other constructs to calculate average values for K_D. For each concentration, the fraction bound has been reported with error bars reporting standard deviation. K_Ds have been calculated from a non-linear regression curve with Hill fitting, reported with an error range that denotes 95% confidence interval, as determined by the fitting of the data to the indicated equation in Prism 9.*

Figure 59 shows the MST data of the LaM mutants in the context of the NTD (1-328). When compared to the NTD, the mutant with the highest change in binding affinity is the D176A, where there is around a 4.5x increase in K_D, where the rest of the mutants have around a 2-4x difference when compared to the NTD. It is important to note that the errors in some of the mutants are quite high, this is due to the nature of MST where the output is quantified using fluorescence which can sometimes be quite inconsistent depending on the fluorescent molecule (in our case the RNA) used.

These results suggest that these 6 residues residing in the LaM hydrophobic pocket may have some involvement in the RNA recognition, however no mutant showed a big reduction in binding affinity towards RNA, although quality control with CD (shown in Figure 60 and 61) showed nucleic acid contamination in these samples. Although these data will need to be repeated for a quantitative rigorous analysis, in any case we can conclude that the LaM of LARP4B behaves differently that other LARPs where such mutations caused drastic reduction in affinity (Teplova et al., 2006; Martino et al., 2015).

Mutant samples were subjected to circular dichroism (CD) analysis to verify the integrity of the secondary structure and determine if there was any difference in folding upon mutation. A normalized Far-UV CD spectra of these LaM mutants is shown in Figure 60. The general shape of the curves compared to one another is quite comparable, however in the case of NTD the intensity of the signal is greater. This was quite odd since the data was already normalized to mean residue ellipticity, which then prompted us to double check the UV signal in the near-UV region, specifically around the 260-280 region where any nucleic acids and proteins are present (Figure 61).

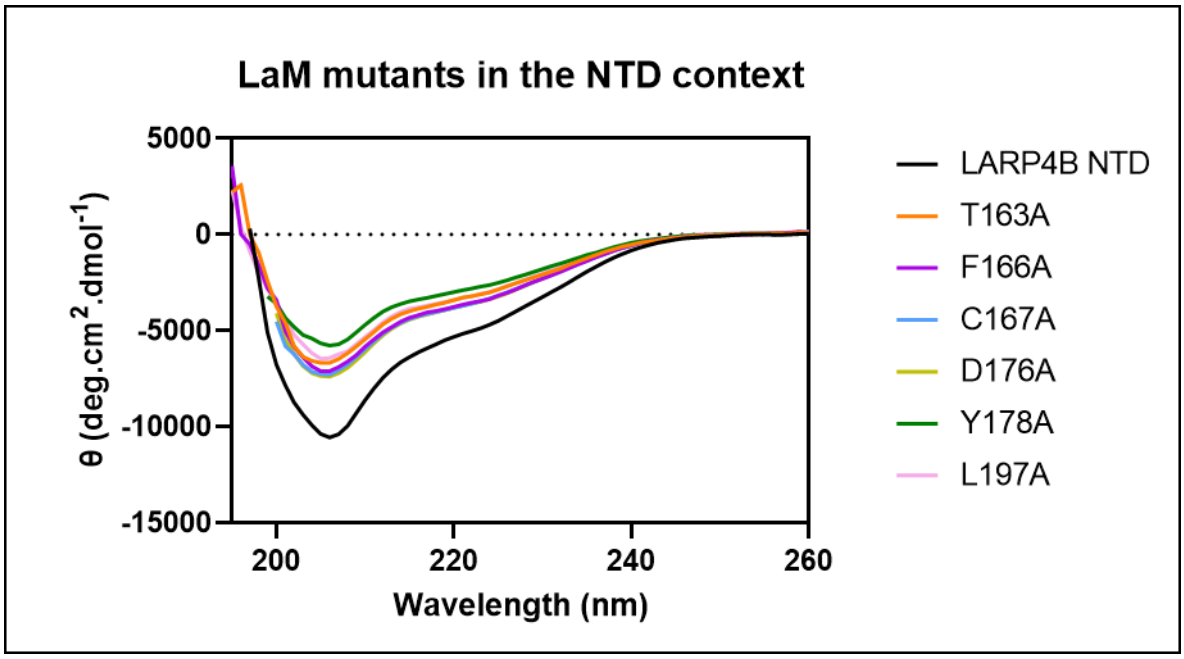


Figure 60. Far-UV CD spectra of LaM point mutants T163A, F166A, C167A, D176A, Y178A and L197A in the context of NTD (1-328) compared to NTD. *The baseline has been subtracted and data have been reported in mean residue ellipticity (MRE) units.*

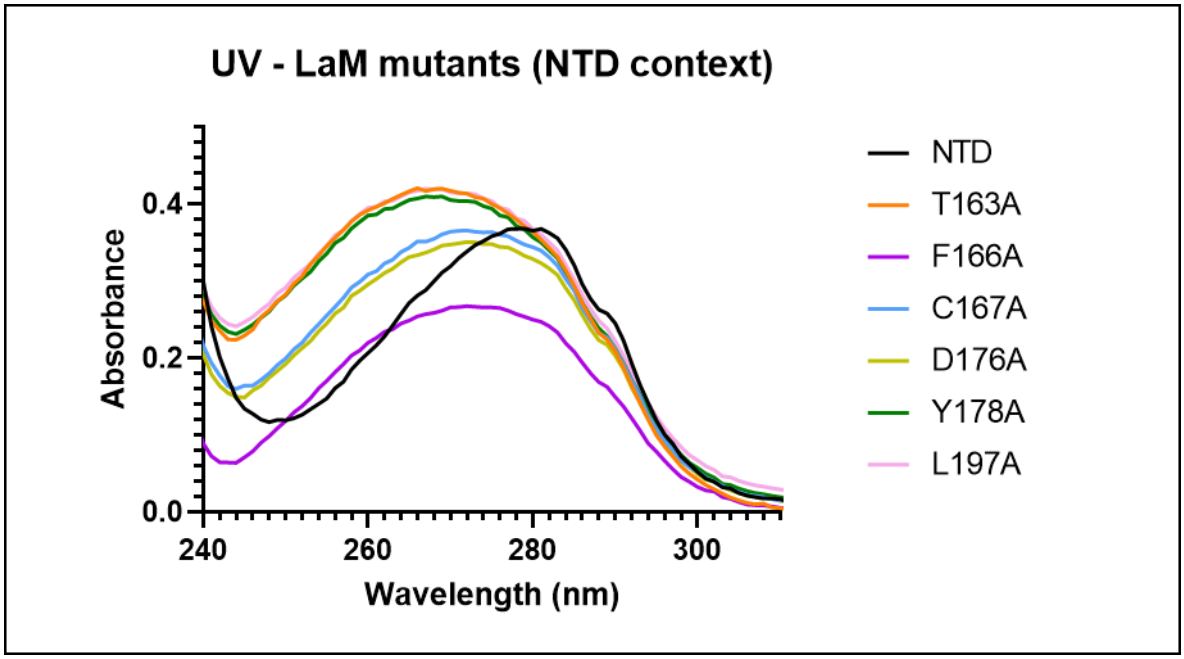
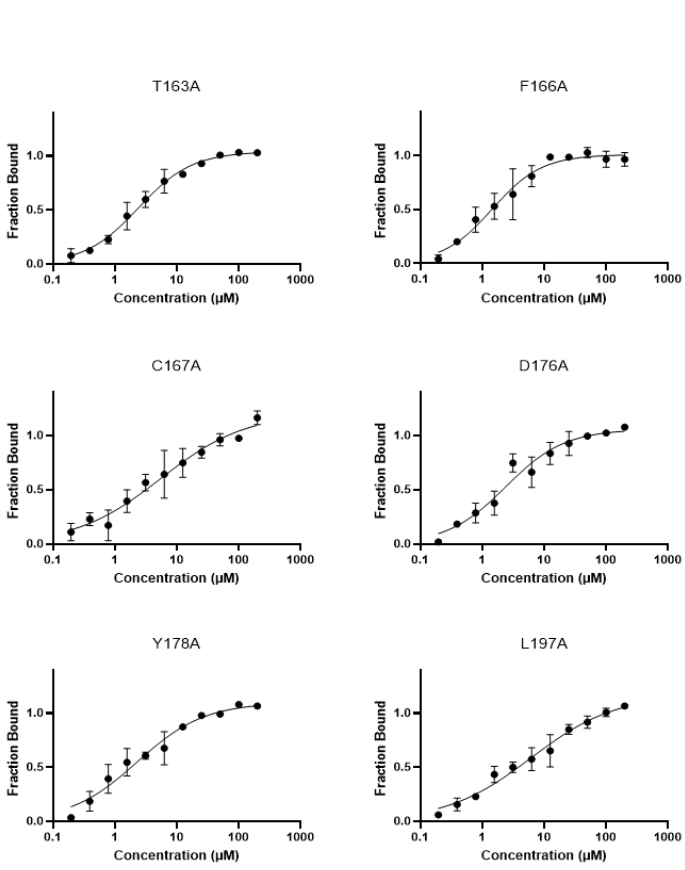


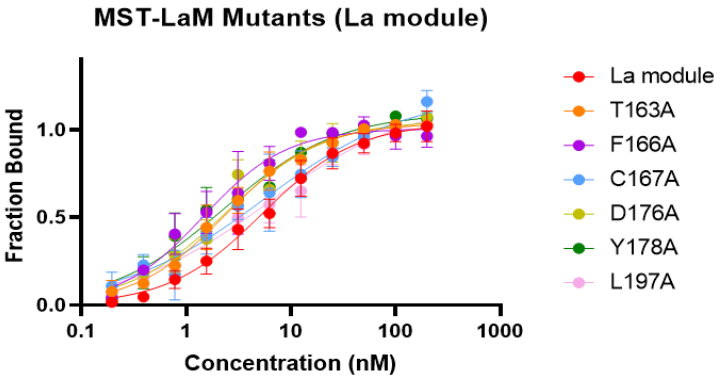
Figure 61. Near-UV spectra of LaM point mutants T163A, F166A, C167A, D176A, Y178A and L197A in the context of NTD (1-328) compared to NTD.

As shown in Figure 61 the LaM mutants mostly have a peak at around 270 nm, where the NTD seems to peak at 280, this can suggest the LaM mutants might be contaminated with

nucleic acids, which will lead to an underestimation of the protein concentration in the sample. This will result in a less intense CD signal of the mutants compared with pure NTD. These mutants will need to be re-purified in order to achieve a better CD curve.

Since from previous EMSA experiments the NTR (EMSA K_D: 5.5 μM) is still very much involved with RNA binding, to eliminate the contribution of the NTR, LaM mutants in the context of the La-module were also cloned and purified alongside the NTD mutants.





Construct	K_D (μM) - MST	Hill Coefficient
La Module	5.2 ± 1.1	1.0 ± 0.1
T163A	2.4 ± 0.6	1.0 ± 0.2
F166A	1.4 ± 0.4	1.1 ± 0.3
C167A	5.5 ± 9.7	0.6 ± 0.2
D176A	2.4 ± 1.0	0.9 ± 0.3
Y178A	2.3 ± 1.0	0.8 ± 0.2
L197A	6.1 ± 4.9	0.6 ± 0.1

Figure 62. MST binding curves for the interaction of LaM point mutants T163A, F166A, C167A, D176A, Y178A and L197A in the context of La-Module (151-328) with 5'FAM-AU-rich RNA (5'-UGGUGAGUUUAUUUUUUUGA-3'). *Binding plots showing the fractions of protein-bound RNA as a function of protein concentrations for the six constructs tested. Two biological replicates were used for T163A and at least three for the other constructs to calculate average values for K_D. For each concentration, the fraction bound has been reported with error bars reporting standard deviation. K_Ds have been calculated from a non-linear regression curve with Hill fitting, reported with an error range that denotes 95% confidence interval, as determined by the fitting of the data to the indicated equation in Prism 9.*

Figure 62 shows the MST data of the same LaM mutants, in the context of the La-module (151-328). Surprisingly, 4 out of the 6 mutants (T163A, F166A, D176A and Y178A) shows a slightly tighter binding towards CKB as tested in MST, with the highest difference being mutant F166A with a 3.7x fold difference. It seems like these point mutations has allowed the protein to bind tighter to the RNA, which contradicts the results of the same mutants in the context of the NTD.

These samples in the context of the La-module have also been analysed by CD (Figure 63). Looking at the curves we can see that comparing to the mutants in the context of the NTD, these curves seem to have a better overlap and have the same shape, suggesting that there is not much difference in secondary structure, this is further confirmed when looking at the Near-UV signal (Figure 64), where the majority of the proteins have a peak at 280 nm, which indicates that there is no nucleic acid contamination.

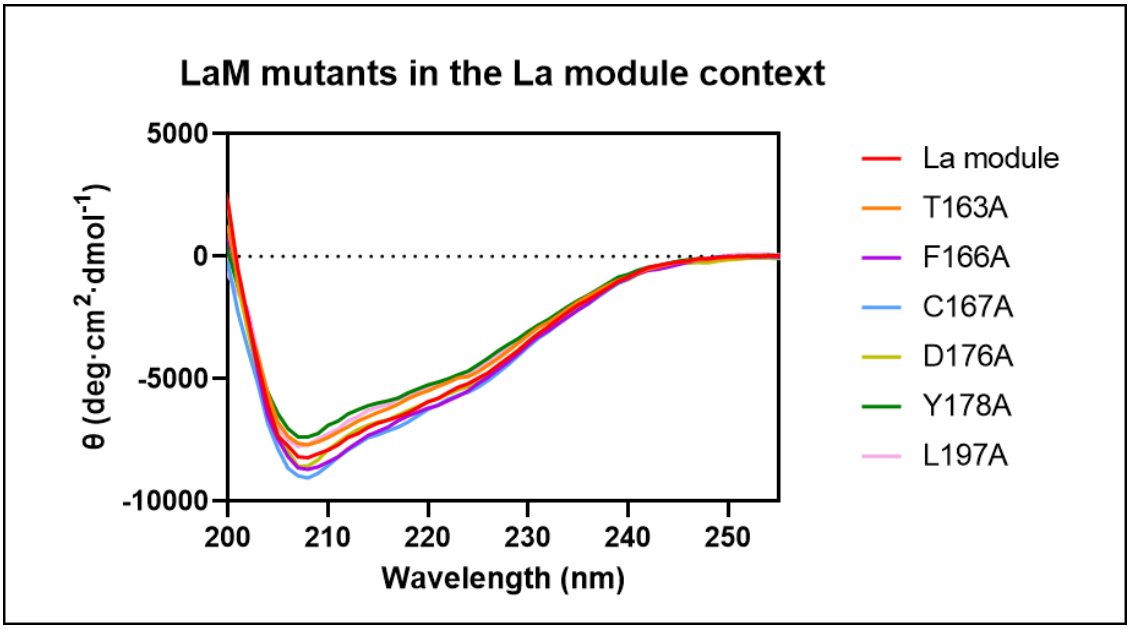


Figure 63. Far-UV CD spectra of LaM point mutants T163A, F166A, C167A, D176A, Y178A and L197A in the context of La-module (151-328) compared to La-module. *The baseline has been subtracted and data have been reported in mean residue ellipticity (MRE) units.*

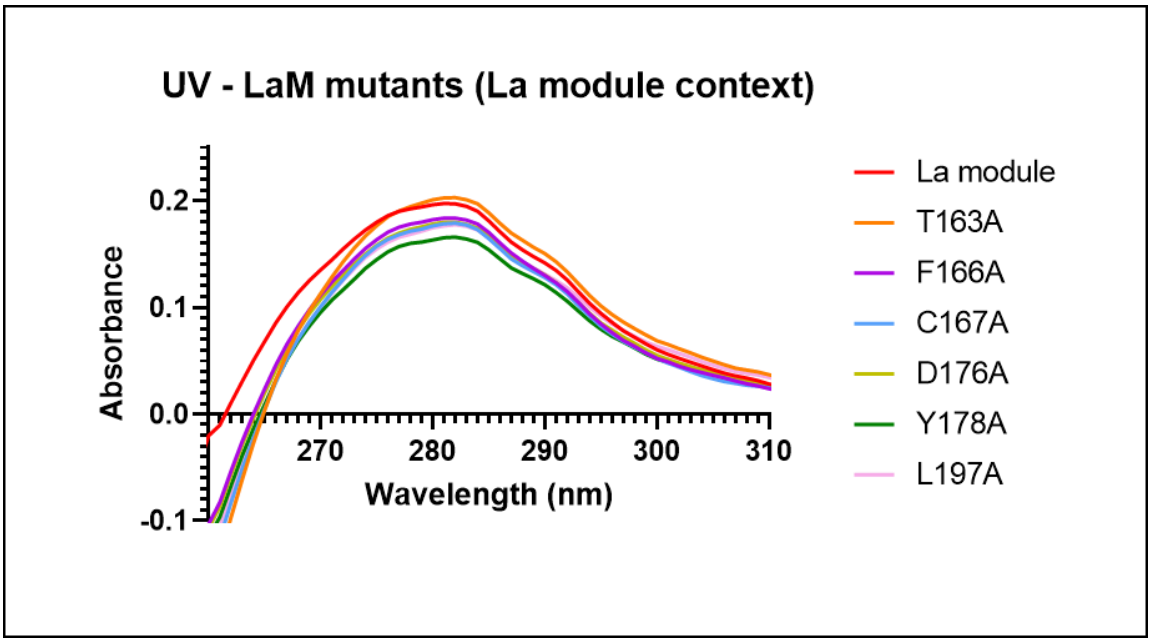


Figure 64. Near-UV spectra of LaM point mutants T163A, F166A, C167A, D176A, Y178A and L197A in the context of La-module (151-328) compared to La-module. *(Negative absorbance values due to an imperfect baseline subtraction)*

These LaM mutagenesis studies suggest that when mutating these 6 conserved residues known to be vital in RNA-interaction in previous LARPs, do not show major contribution in the case of LARP4B, suggesting that other regions in LARP4B La-module, yet to be found, are important for RNA interaction.

4.7 LARP4B PAM2w motif has a role in Protein-Protein interactions, but not a major involvement in Protein-RNA interactions

The PAM2w motif, a PABPC1 binding motif, is present in the N-terminus of both LARP4A and LARP4B, there are two important residues within the PAM2w motif highlighted previously in literature for LARP4A: leucine 15 and the variant tryptophan 22 (Yang *et al.*, 2011). PAM2w is called as such due to having a tryptophan in the otherwise invariantly phenylalanine residue that is conserved in most of the PAM2

containing proteins. Yang et al. has shown that when mutating the leucine and tryptophan to alanine, the ability for LARP4A to bind to PABPC1 is lost. In subsequent studies in the Conte Lab, Cruz-Gallardo et al. have identified that the same mutation L15AW22A prevents the binding to the RNA target of LARP4A (Cruz-Gallardo *et al.*, 2019).

Not only is the PAM2w domain in LARP4A important in protein-protein interactions but also a key player for protein-RNA interactions, therefore the same mutants were designed in LARP4B to provide any insight on the role of the PAM2w motif in LARP4B. There were two mutants tested, one of them was L56AW63A, and the second one was replacing residue W63 back to the invariable F residue, W63F.

EMSAs were performed to analyse the importance of these mutants (Figure 65). The mutant L56AW63A has a binding affinity towards CKB RNA of 0.64 ± 0.25 μM, with a Hill coefficient of 1.0 ± 0.3, and the mutant W63F has a binding affinity of 0.76 ± 0.45 μM, with a Hill coefficient of 1.0 ± 0.4. Comparing them both to the NTD ($K_{D:}$ 0.18 μM), a decrease in binding affinity is seen of 3-4 folds for both mutants, however, the mutants do not seem to drastically reduce the RNA-binding ability of LARP4B, as seen in LARP4A (Cruz-Gallardo *et al.*, 2019). Therefore, it can be concluded that the PAM2w in LARP4B may have some contribution in terms of RNA-binding, but it behaves quite differently from LARP4A where the PAM2w is the key region for RNA-binding.

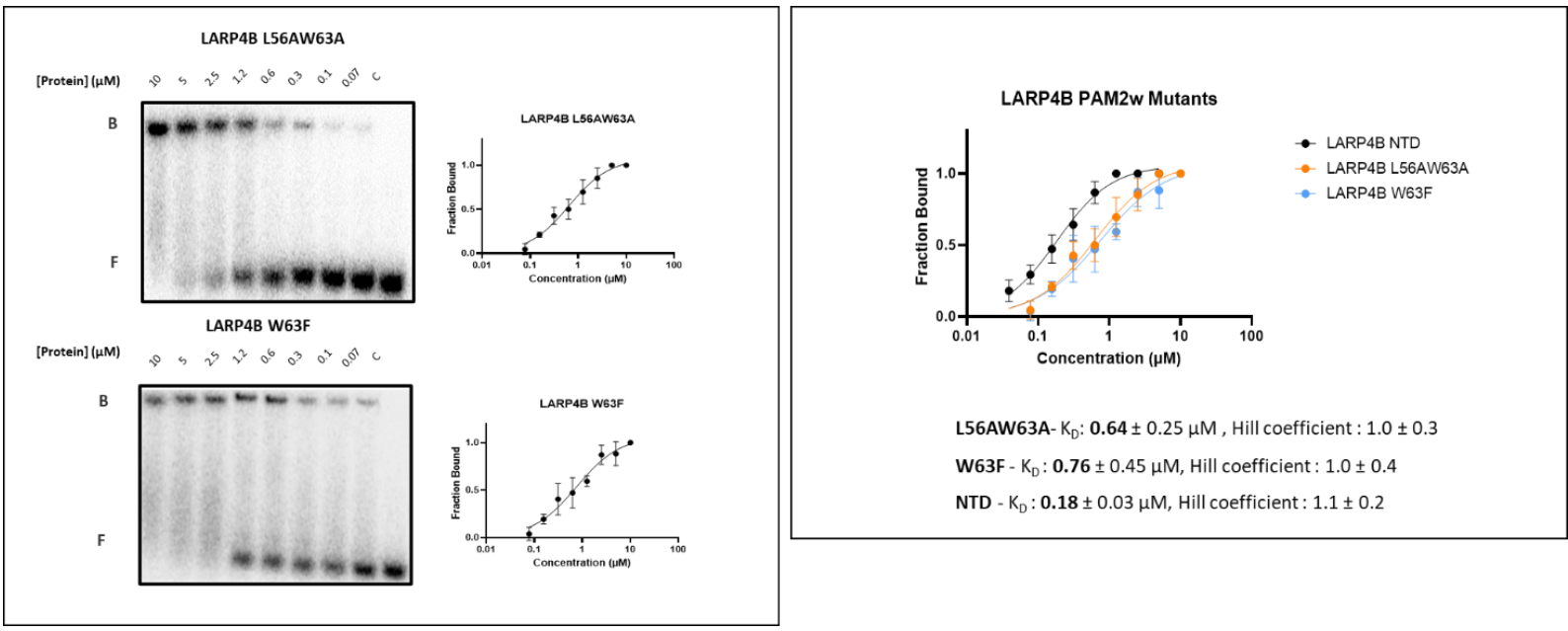


Figure 65. EMSA binding assays of LARP4B PAM2w mutants L56AW63A and W63F with 32P-AU-rich RNA (5'-UGGUGAGUUUAUUUUUUUGA-3'). *Representative autoradiograms are shown for L56AW63A (top) and W63F (bottom) and their binding curves (right). The protein concentrations used were 0, 0.07, 0.1, 0.3, 0.6, 1.2, 2.5, 5 and 10 μM for L56AW63A and W63F. Bound (B) and free (F) RNA populations are labelled. At least three biological replicates were used to calculate average values for KD. For each concentration, the fraction bound has been reported with error bars reporting standard deviation. KDs have been calculated from a non-linear regression curve with Hill fitting, reported with an error range that denotes 95% confidence interval, as determined by the fitting of the data to the indicated equation in Prism 9.*

It has been established that the PAM2w does not considerably affect the RNA binding capabilities in LARP4B, but to test the interaction of the PAM2w motif towards the MLLE domain of PABPC1, pulldown experiments, which is an *in vitro* method used to determine physical interaction between two or more proteins were performed.

Initially, for positive control, it is known that LARP4A NTD binds MLLE using its PAM2w motif (Cruz-Gallardo *et al.*, 2019) (Yang *et al.*, 2011), and the mutation of the PAM2w motif (L15AW22A) abrogates the binding towards MLLE. To replicate the experiment performed in Cruz-Gallardo et al., 2019, LARP4A NTD, L15AW22A and His SUMO tagged MLLE were purified. 200 μg of His-tagged MLLE which acts as the 'bait' is mixed with either 250 μg of untagged LARP4A NTD or L15AW22A which acts as the 'prey.' After addition to a 300 μL nickel resin, the mixture is washed with Nickel wash buffer a couple of times to remove any unbound and excess protein. The mixture is

then eluted with Nickel elution buffer and the fractions are analysed on an SDS-PAGE gel (Figure 66).

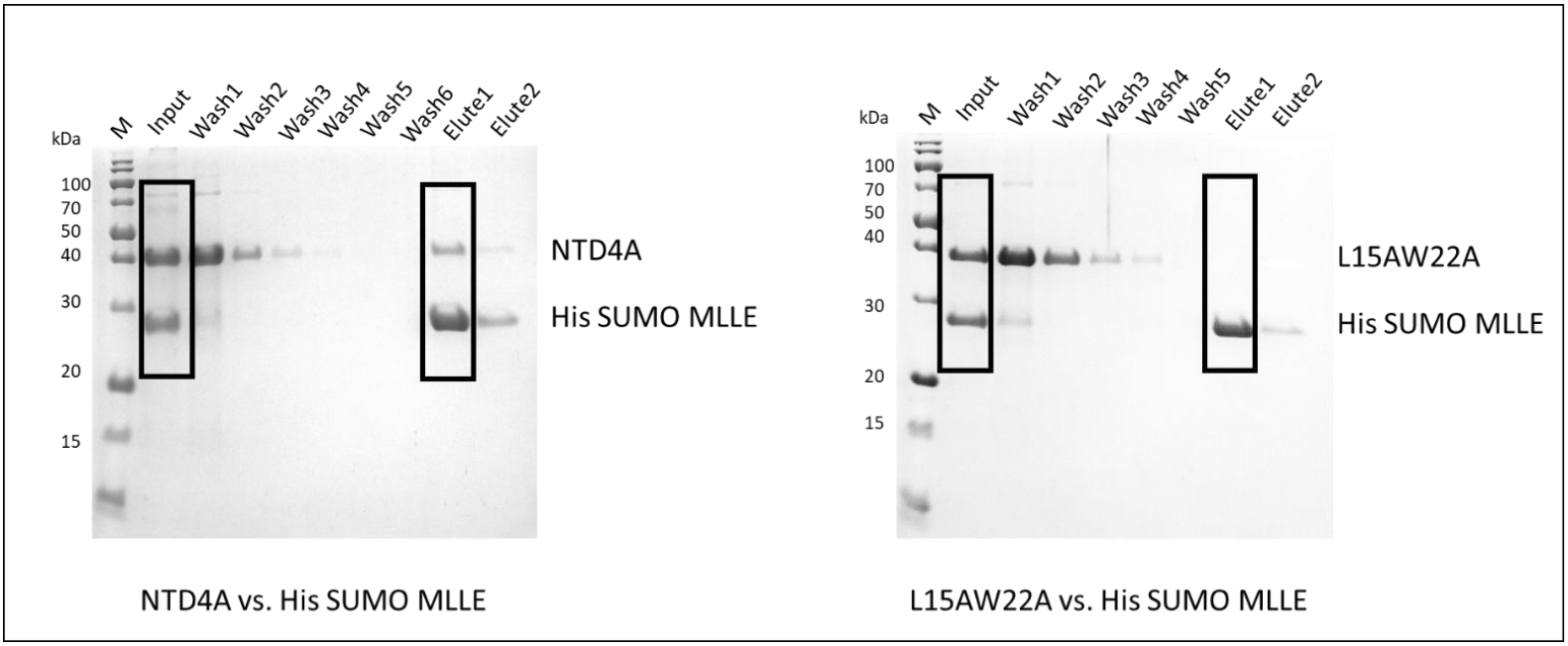


Figure 66. SDS-PAGE analysis of Nickel affinity pull-down assays to analyse the interaction between LARP4A PAM2w mutants to MLLE. *Untagged LARP4A NTD and mutant L15AW22A were purified and mixed with purified His-SUMO-MLLE, used as bait. Pull-down assays were carried out using Ni-NTA affinity beads in a buffer containing 100 mM KCl. The Coomassie-stained 12% SDS-PAGE gel shows the input fractions applied to the resin and the pulled-down proteins eluted with 500 mM imidazole.*

The result of the positive control shows that in the case of the wild type LARP4A NTD, after the elution of the pulldown mixture, two visible bands are present in the elution lane, which indicates that LARP4A NTD binds to MLLE as expected, in the other hand, the mutation L15AW22A has disrupted the binding to MLLE, as expected.

With the positive control completed, LARP4B proteins were then tested to see if the PAM2w motif was important in MLLE interaction. The experiment was conducted in the same way, using 200 μg of His-tagged MLLE which acts as the 'bait' is mixed with either 250 μg of untagged LARP4B NTD, L56AW63A or W63F which acts as the 'prey.' Figure 67 shows the SDS-PAGE pulldown analysis of the proteins LARP4B NTD.

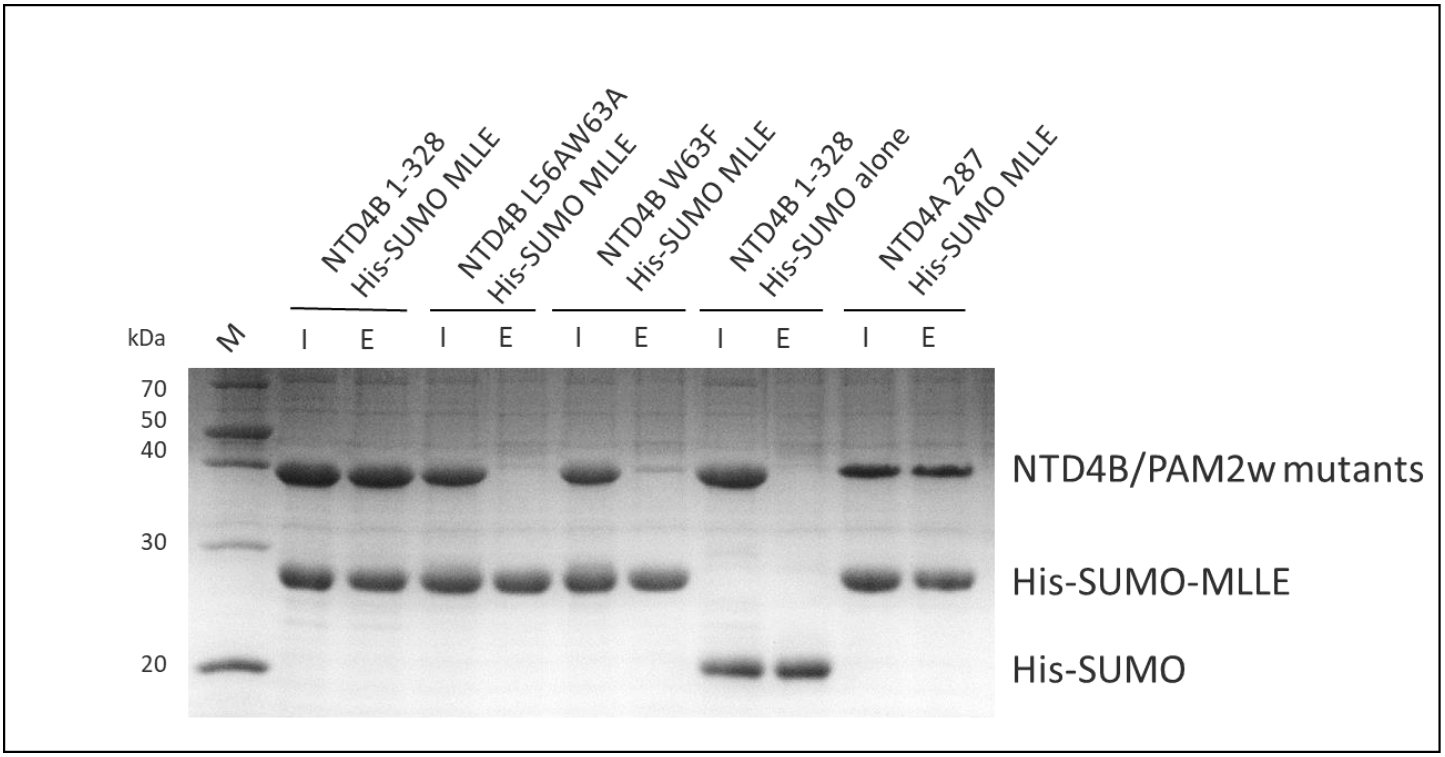


Figure 67. SDS-PAGE analysis of Nickel affinity pull-down assays to analyse the interaction between LARP4B PAM2w mutants to MLLE. *Untagged LARP4B NTD and mutants L56AW63A and W63F were purified and mixed with purified His-SUMO-MLLE, used as bait. Pull-down assays were carried out using Ni-NTA affinity beads in a buffer containing 100 mM KCl. The Coomassie-stained 12% SDS-PAGE gel shows the input fractions applied to the resin and the pulled-down proteins eluted with 500 mM imidazole. Control experiments with His-SUMO alone exclude unspecific binding to the tag, and a representative control experiment performed with LARP4A NTD is shown.*

LARP4B NTD shows an interaction to MLLE, as shown by the presence of the NTD domain band in the elution fraction together with the MLLE. The PAM2w double mutant L56AW63A shows no binding towards MLLE, which suggests that PAM2w in LARP4B is responsible for interaction with PABPC1 MLLE and that these two residues are important at mediating the interaction. The W63F mutant, which mutates the variant amino acid residue tryptophan back to the otherwise conserved phenylalanine present in other PAM2 motif containing proteins, shows a slight recovery of binding towards MLLE, however when comparing to the wild type NTD, the binding is weaker. This presence of the tryptophan therefore is important for the binding to MLLE in LARP4B. In order to show that the His-SUMO tag is not affecting the experiments in any way, in a control experiment LARP4B NTD pulldown is performed with the tag alone, and this showed no

interaction with NTD. Together, these experiments show that the PAM2w motif in LARP4B retains its ability to bind to the MLLE domain in PABPC1.

LARP4A NTD engages with both oligoA RNA and the MLLE domain of PABPC1 at the PAM2w motif. A competition experiment performed by (Cruz-Gallardo *et al.*, 2019) has shown that the oligoA binding in LARP4A is mutually exclusive when using an EMSA competition assay titrating increasing concentrations of MLLE to a preformed LARP4A NTD–oligoA complex, resulting the MLLE dissociating the preformed complex.

To further deduce whether the binding of LARP4B NTD with the engagement of PABPC1 MLLE and CKB RNA is concomitant or mutually exclusive, a competition experiment was carried out (Figure 68) where an increasing amount of MLLE protein titrated to a pre-formed LARP4B NTD-CKB complex was analysed on an EMSA. The left gel shows a maximum concentration of MLLE of 200 μM and the right shows a concentration of 400 μM. With the introduction of MLLE protein, there is no change in the profile of the EMSAs, and the size of the band does not increase or decrease, even with the addition of up to 400 μM MLLE. This suggests that the addition of the MLLE protein cannot bind simultaneously with a preformed protein-RNA complex to form a super-complex.

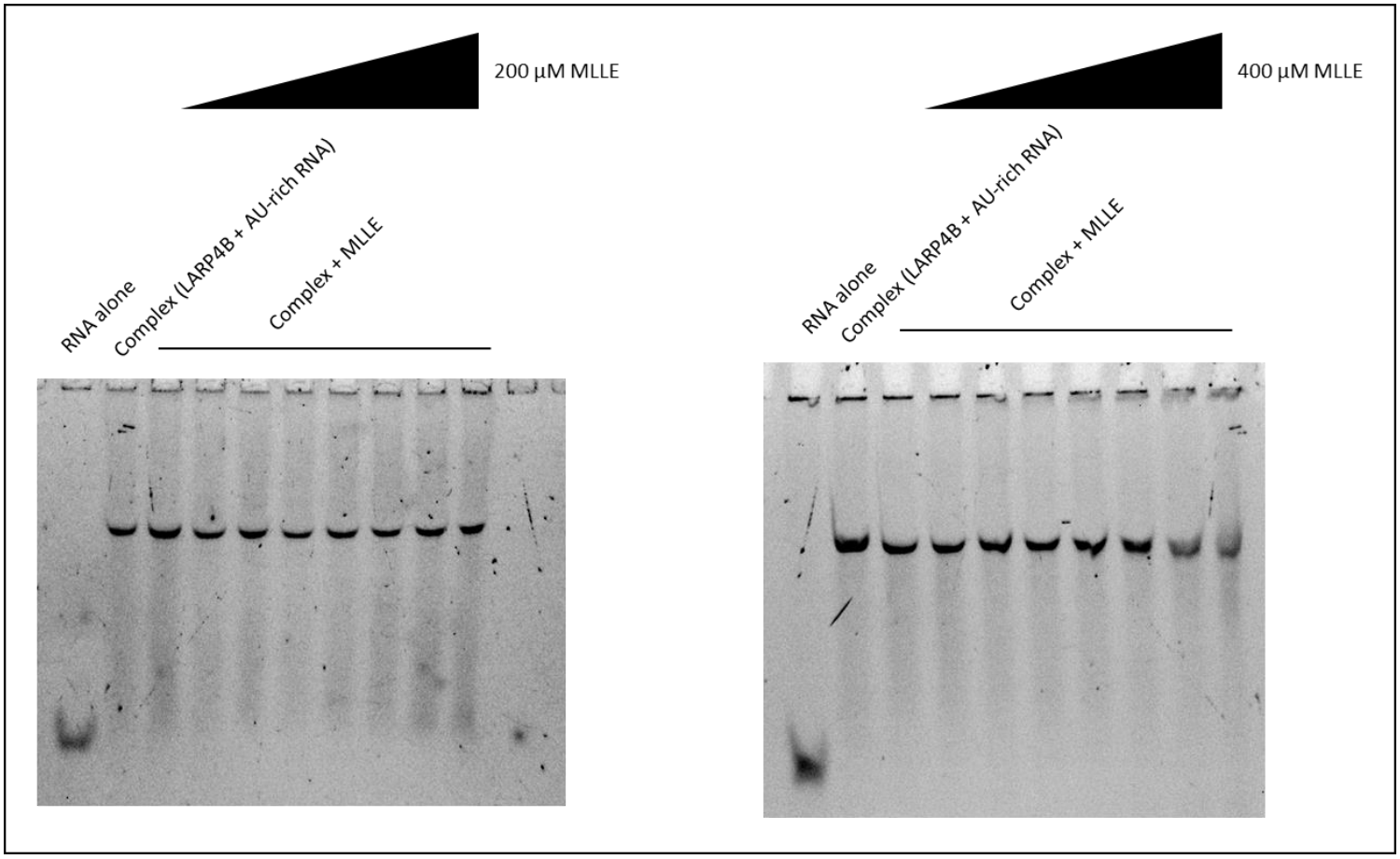


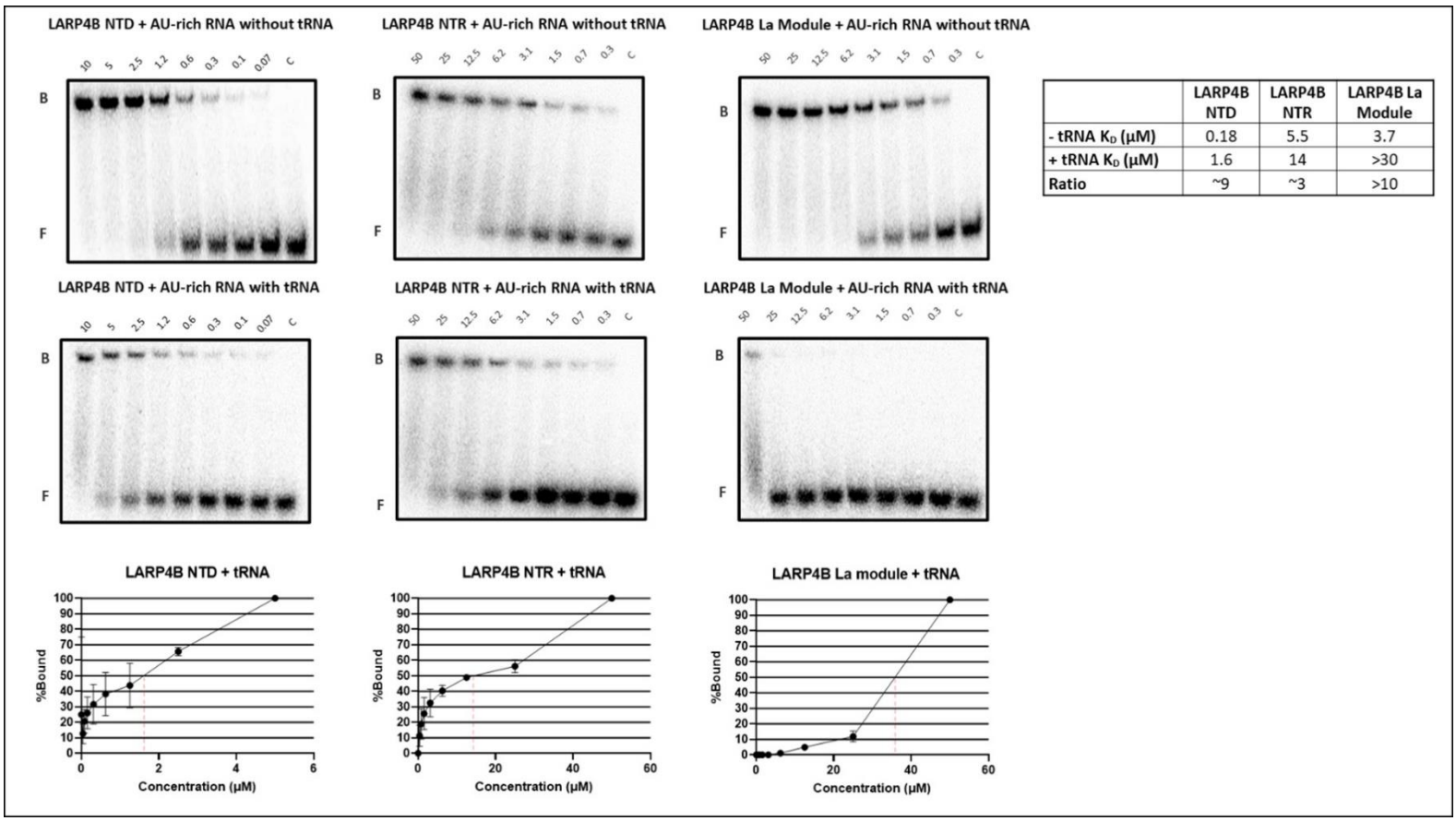
Figure 68. Competition binding experiments between LARP4B NTD, PABPC1 MLLE domain and AU-rich RNA (5'-UGGUGAGUUUAUUUUUUUGA-3'). *Titration of a pre-formed complex of LARP4B NTD-AU-rich RNA with MLLE domain. MLLE concentrations used were 0, 0.7, 1.5, 3.1, 6.2, 12.5, 25, 50, 100 and 200 µM (left) and 0, 1.5, 3.1, 6.2, 12.5, 25, 50, 100, 200 and 400 µM (right).*

4.8 LARP4B NTD can interact with other RNA sequences

All the results so far have been performed without the addition of *E. coli* MRE 600 tRNA mix as a nonspecific competitor. This section shows EMSA results with the addition of tRNA, compared to the results without tRNA. The tRNA is used as a blocking agent in the reaction to minimize the binding of nonspecific proteins to the labelled RNA. The repetitive and random fragments in the tRNA mix provides an excess of nonspecific sites for the proteins that can bind to any general RNA sequence. In the case of LARP4A, the

EMSAs experiments conducted with and without the tRNA mix showed little difference in binding affinity (Isabel Cruz Gallardo's data, not shown), however in the case of LARP4B, it seems like some of the mutants are affected by the presence of the tRNA mix in the experiment.

Figure 69 shows the representative EMSA figure of the constructs LARP4B NTD (left) LARP4B NTR (middle) and LARP4B La-module (right). The K_Ds with the addition of the tRNA is estimated using a curve that plots the band intensity against the protein concentration, the curves are done this way since most of the reactions with tRNA don't reach a plateau, and the resulting K_Ds cannot be measured.

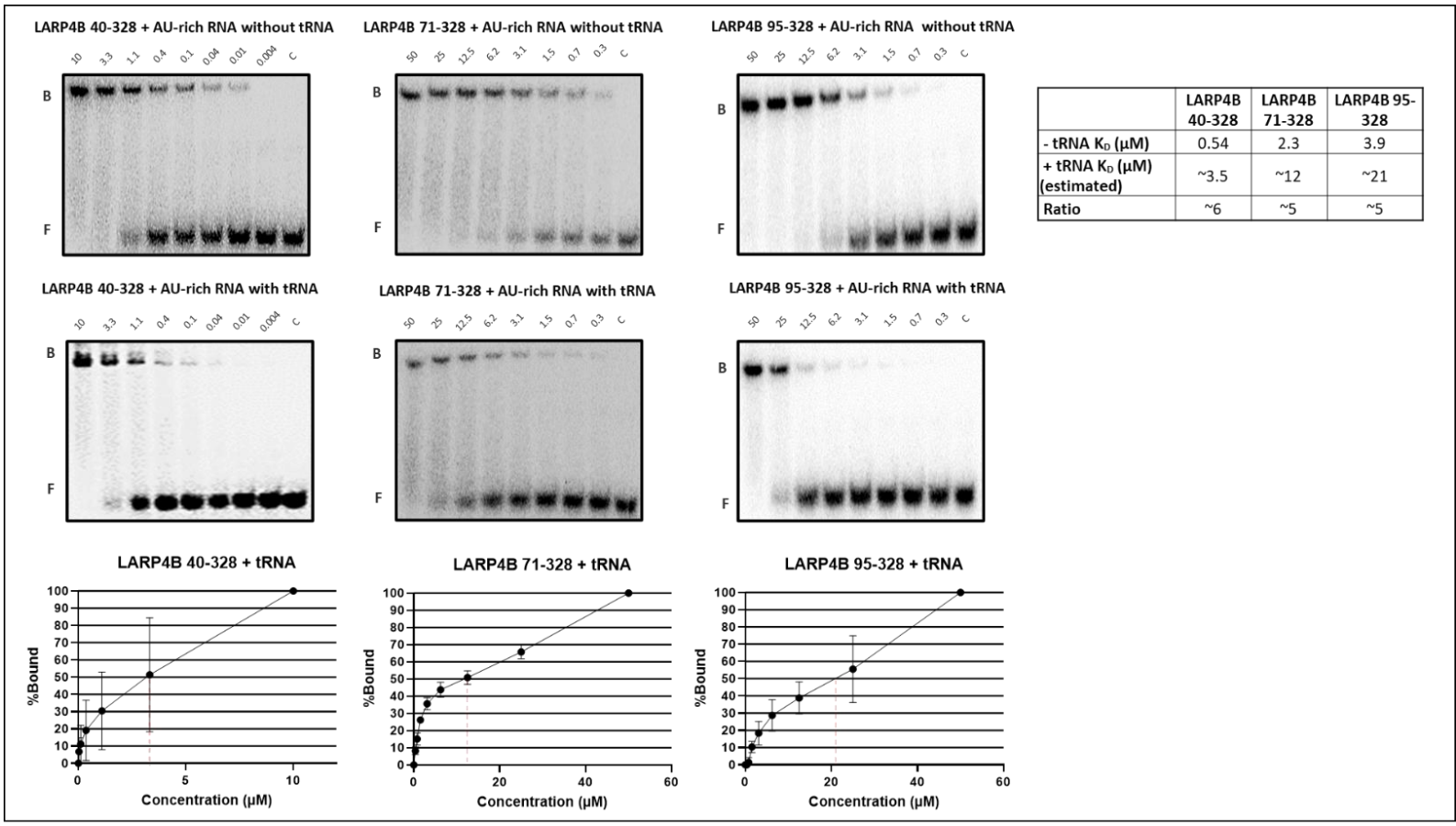


	LARP4B NTD	LARP4B NTR	LARP4B La Module
- tRNA K_D (μM)	0.18	5.5	3.7
+ tRNA K_D (μM)	1.6	14	>30
Ratio	~9	~3	>10

Figure 69. EMSA binding assays of LARP4B N-terminal domain (NTD), N-terminal region (NTR) and La-Module with 32P-AU-rich RNA (5'-UGGUGAGUUUAUUUUUUUGA-3') in the absence of E.coli tRNA (top) and in its presence (bottom). *Representative autoradiograms are shown for LARP4B NTD (left), NTR (centre) and La-Module (right). The protein concentrations used were 0, 0.07, 0.1, 0.3, 0.6, 1.2, 2.5, 5 and 10 μM for NTD and 0, 0.3, 0.7, 1.5, 3.1, 6.2, 12.5, 25 and 50 μM for La-Module and NTR. Bound (B) and free (F) RNA populations are labelled. K_Ds have been estimated on at least three biological replicates as the concentration for which the protein is 50% bound.*

Looking at the difference in binding affinity, LARP4B NTD without the addition of the tRNA competitor has a K_D for AU-rich of 0.18 μM, but in presence of the tRNA mix the K_D changes to approximatively 1.6 μM, which is around 9-fold difference. Similarly, LARP4B NTR without the addition of the tRNA competitor has a K_D for AU-rich of 5.5 μM, but in presence of the tRNA mix the K_D changes to approximatively 14 μM, which is around 3-fold difference. La-module without the addition of the tRNA competitor has a K_D for AU-rich of 3.7 μM, but in presence of the tRNA mix the K_D changes to approximatively >30 μM when fresh tRNA is used which is more than 10-fold difference.

When analysed using the N-terminal deletion mutants (40-328, 71-328 and 95-328) (Figure 70), 40-328 without the addition of the tRNA competitor has a K_D for AU-rich of 0.54 μM, but in presence of the tRNA mix the K_D changes to approximatively 3.5 μM, which is around 6-fold difference. Similarly, 71-328 without the addition of the tRNA competitor has a K_D for AU-rich of 2.3 μM, but in presence of the tRNA mix the K_D changes to approximatively 12 μM, which is around 5-fold difference. 91-328 without the addition of the tRNA competitor has a K_D for AU-rich of 3.9 μM, but in presence of the tRNA mix the K_D changes to approximatively 21 μM when fresh tRNA is used which is around 5-fold difference.



	LARP4B 40-328	LARP4B 71-328	LARP4B 95-328
- tRNA K_D (μM)	0.54	2.3	3.9
+ tRNA K_D (μM) (estimated)	~3.5	~12	~21
Ratio	~6	~5	~5

Figure 70. EMSA binding assays of LARP4B N-terminal deletion mutants 40-328, 71-328 and 95-328 with 32P-AU-rich RNA (5'-UGGUGAGUUUAUUUUUUUGA-3') in the absence of E.coli tRNA (top) and in its presence (bottom)*. Representative autoradiograms are shown for 40-328 (left), 71-328 (centre) and 95-328 (right). The protein concentrations used were 0, 0.004, 0.01, 0.04, 0.1, 0.4, 1.1, 3.3 and 10 μM for 40-328 and 0, 0.3, 0.7, 1.5, 3.1, 6.2, 12.5, 25 and 50 μM for 71-328 and 95-328. Bound (B) and free (F) RNA populations are labelled. K_Ds have been estimated on at least three biological replicates as the concentration for which the protein is 50% bound.*

However, the La-module still shows the largest change, this suggests that the putative RNA sequence present in the tRNA mix that binds to LARP4B is more likely to bind to the La-module rather than anywhere in the N-terminus.

To characterize this further, LaM and the RRM alone were analysed in EMSAs in the presence of tRNA (Figure 71). However, the single domains do not show any drastic difference in AU-rich binding in the presence of tRNA. A possible hypothesis may be that the combination of the LaM plus the RRM is required for the interaction to the tRNA mix sequence that is interacting to the La-module.

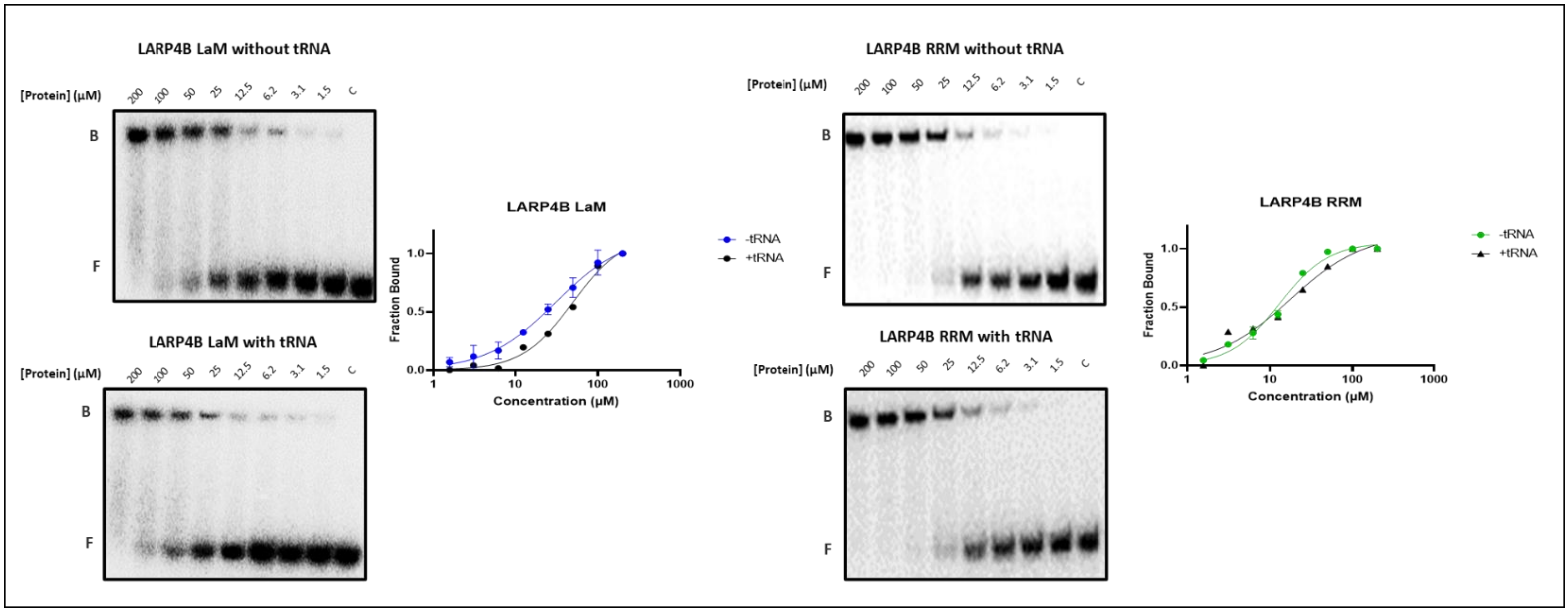


Figure 71. EMSA binding assays of LARP4B La motif (LaM) and RRM with 32P-AU-rich RNA (5'-UGGUGAGUUUAUUUUUUUGA-3') in the absence of E.coli tRNA (top) and in its presence (bottom). *Representative autoradiograms are shown for LaM (left) and RRM (right) and their binding curves have been reported. The protein concentrations used were 0, 1.5, 3.1, 6.2, 12.5, 25, 50, 100 and 200 μM. Bound (B) and free (F) RNA populations are labelled.*

The result of the EMSAs with tRNA raises the question as to whether binding of LARP4B to AU-rich RNA is specific or whether LARP4B NTD can bind to any RNA. To test the RNA binding properties of LARP4B towards different RNAs, several homopolymers were tested against LARP4B NTD, this includes oligo A20, oligo U20 and oligo C20 (Figure 72).

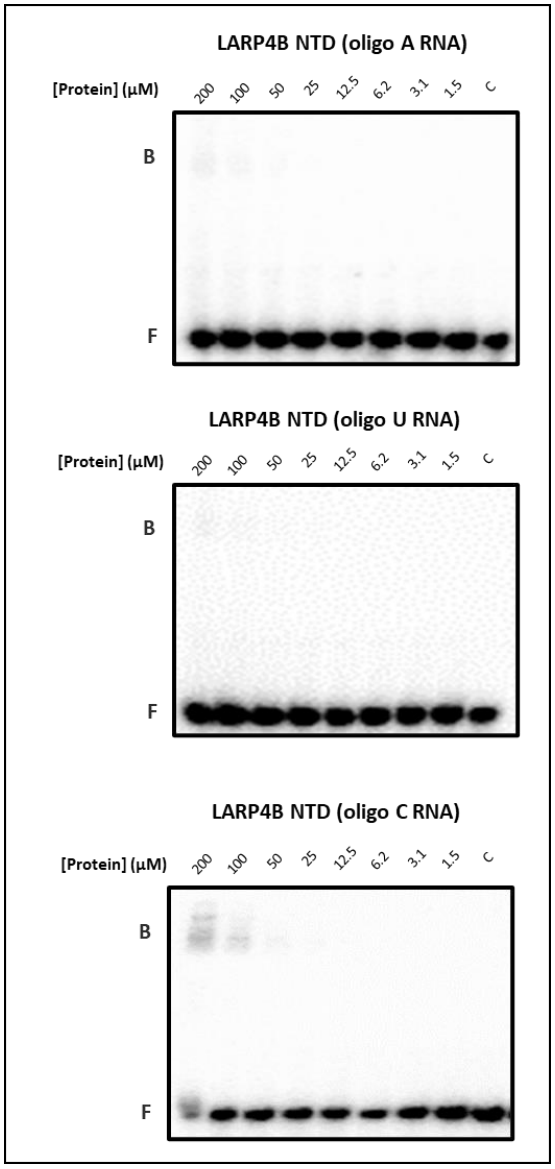


Figure 72. EMSA binding assays of LARP4B NTD with 32P-oligoA20, oligoU20 and oligoC20 RNA. *Representative autoradiograms are shown for oligo A (top), oligo U (centre) and oligo C (bottom). The protein concentrations used were 0, 1.5, 3.1, 6.2, 12.5, 25, 50, 100 and 200 μM. Bound (B) and free (F) RNA populations are labelled.*

The results show that LARP4B does not interact with oligo A20, oligo U20 and oligo C20. This provides evidence that LARP4B does not simply bind to any RNA and confirms the specificity of LARP4B towards AU-rich RNA.

4.9 Conclusions

In this chapter, LARP4B, and its interaction with AU-rich/CKB RNA and PABPC1 MLLE have been characterised. The main methods used to characterise the protein-RNA interaction were EMSAs and MSTs, where MSTs were used to test an array of point mutants, and to validate EMSA results. The advantages of MSTs are that results are extremely quick to obtain so a lot of mutants can be tested, low sample consumption, native buffer conditions and allows for a wide size range for interactants. However, fluorescent labelling of the RNA is required which might affect the way that it interacts with the protein, it can also be sensitive to any changes in molecular properties such as size, charge, hydration shell or conformation (Plach, Grasser and Schubert, 2017). Although an effort was made to check all raw data and remove outliers and anomalies, some of the results obtained via MST had large errors. MST could have a larger variability with results and more technical repeats are required per experiment in order to obtain reliable data. EMSAs that use radioisotope-labelled nucleic acids are highly sensitive, once a good working protocol is obtained, results are extremely consistent, and very low amounts of protein and RNA are needed per run (Hellman and Fried, 2007). The technique itself is simple to learn, but hard to master, there are many variables that needed to be kept consistent like running temperature, time of reaction, run, and staining to the phosphoimager.

LARP4B, despite it having 40% amino acid identity with LARP4A and 74% in their La-modules, does not behave like LARP4A, possibly due to the other regions such as the NTR and the C-terminus which is less similar. The EMSA experiments have shown a tight binding of the NTD to CKB RNA (K_D: 0.18 μM), but the La-module of LARP4B binds relatively tight to CKB RNA as well (K_D: 3.7 μM), compared to the La-module of LARP4A binding to oligoA15 (K_D: >100 μM), LARP4B La-module behaves like a canonical La-module in the sense that it is still required to bind to RNA. The PAM2w

motif of LARP4B also does not seem to drastically decrease the binding of RNA that much (L56AW63A mutant showed a 3-4-fold decrease in affinity compared to the NTD). This is compared to LARP4A where the L15AW22A mutant had a 20x decrease in binding affinity compared to the NTD. The RNA-binding locus of LARP4B does not rely on the PAM2w motif alone, rather the whole NTD is required for the maximum binding towards RNA. However, using pulldown assays, it is concluded that PAM2w in LARP4B still functions to bind PABPC1 MLLE, as a L56AW63A mutation abrogated the binding to MLLE in pulldown assays and the W63F mutant strongly reduced the binding to MLLE, the PAM2w motif is still needed to interact with PABPC1 and could be aided by other regions in LARP4B such as in the C-terminus which have not been looked at in this study. The experiments involving the addition of tRNA in EMSA studies hinted for additional target RNAs that LARP4B could potentially have, as the studies of the PAR-CLIP (Küspert *et al.*, 2015) also had presented other potential targets alongside CKB. We have showed that LARP4B does not bind to any RNA, but remains specific to its targets, and tested it against other homopolymers (poly A, U and C) where it showed no binding.

Chapter 5. Discussion and Future Work

5.1 Project overview

In this study, we have taken an *in vitro* systematic approach to identify and characterise the RNA and protein targets of LARP4B, using LARP4A as a comparison. As LARP4B contains the La-module it was considered a bona fide RNA-binding protein, also studies have shown that it can interact with protein partners such as PABPC1 and RACK1 (Angenstein *et al.*, 2002; Schäffler *et al.*, 2010). Truncation and point mutants of LARP4B were cloned, expressed and purified in order to obtain the RNA-binding signature of LARP4B. EMSA, MST and CD experiments on these mutants provided an insight on the binding affinities of each of the domains in LARP4B versus its target AU-rich RNA.

5.2 The binding mode of LARP4B

Given that LARP4B is a paralog of LARP4A, they have high amino acid similarity in key domains and share the same protein partners. With previous knowledge of LARP4A, where its PAM2w motif plays a key role in oligoA RNA binding, and its unconventional La-module that does not bind RNA (K_D= >100 μM) (Cruz-Gallardo *et al.*, 2019), it is safe to say that LARP4B behaves differently than LARP4A, where the La-module still functions to bind AU-rich RNA, and the PAM2w motif plays its normal role of interaction with PABPC1 and is not the primary RNA-binding locus compared to LARP4A.

Even though LARP4B is similar to LARP4A, they do bind to different RNA-targets, and with different affinity depending on the domain that is involved. LARP4A has a key RNA-binding locus involving the PAM2w which is situated in an intrinsically disordered

region (Cruz-Gallardo *et al.*, 2019). LARP4B requires the whole NTD in order to achieve maximum RNA-binding affinity (K_D= 0.18 μM), but the La-module (K_D= 3.7 μM) and NTR alone (K_D= 5.5 μM), still functions well to bind to RNA. LARP4A and 4B differ in their N-terminus in which LARP4A's NTR is approximately 40 residues shorter than 4B; also a 2 residue difference out of six in the conserved LaM binding pocket; and around 10 residue difference in their RRMs. It would seem that these differences may contribute to differential RNA preferences of LARP4A and 4B.

The EMSA data (also confirmed with MST) suggest that both the La-module and NTR bind AU-rich RNA and this seems to be additive since the full NTD binds 20-30 times stronger than either the La-module or the NTR alone. There is a possibility of the La-module and the NTR binding the same RNA at the same time, however preliminary data by mass spectrometry (not shown) by the Conte Lab has suggested that the stoichiometry of LARP4B NTD:AU-rich RNA is one-to-one. This suggests that the La-module and NTR aids the binding for a single RNA molecule when it is in the context of the NTD, and hence the NTD provides the maximum affinity towards the RNA.

Since neither of the regions listed above dominated RNA binding, the EMSA experiments of the N-terminal truncation mutants 40-328 (K_D= 0.54 μM), 71-328 (K_D= 2.3 μM), and 95-328 (K_D= 3.9 μM), was an attempt to narrow down the region of interest for RNA interaction within the NTR. The 40-328 mutant was designed since in LARP4B the NTR is longer than LARP4A and the initial 40 residues are unique to LARP4B, however, removing the first 40 residues did not significantly alter the affinity, rather the truncation from 40-71 showed the most difference. There could be potential residues that are important within residues 40-71 that are key for the RNA-interaction of LARP4B and should be further investigated by running a RNA-binding motif/secondary structure search and further investigated using *in vitro* mutagenesis studies.

EMSAs of the two domains of the La-module, the LaM (K_D= 30.9 μM), and the RRM (K_D= 13.0 μM) showed weak contribution towards RNA binding. This suggests that the domains alone are not efficient RNA-binders, the full La-module is required to bind to the RNA target. It is interesting to note that the RRM binds tighter than the LaM since historically the LaM which consists of key amino acid residues are responsible to interact with their RNA targets (Alfano *et al.*, 2004; Bousquet-Antonelli and Deragon, 2009; Martino *et al.*, 2015b).

To investigate the importance of the LaM in the RNA binding, our MST results of the LaM mutants mutated the 6 key residues in the LaM to alanine. Surprisingly, the point mutants do not show a significant amount of difference compared to the controls (NTD and La-module, depending on context). In other LARPs, similar mutations in these six residues cause drastic reductions in binding affinity (Teplova *et al.*, 2006; Martino *et al.*, 2015b). The LaM of LARP4s is generally less conserved compared to La and the other LARPs. They have a cysteine at position 24 (hLa numbering) whereas all others have the aromatic Tyr (or Trp, LARP7) which contributes to the hydrophobicity of the interaction pocket (Teplova *et al.*, 2006; Kotik-Kogan *et al.*, 2008). Moreover, LARP4A and LARP4B are less conserved at the Q20 and F55 site, where in LARP4B the Q20 is replaced by T; and M and L respectively in the F55 position of this otherwise invariant residue (Teplova *et al.*, 2006; Kotik-Kogan *et al.*, 2008). The lack of conservation in the LaM could indicate that LARP4s utilises a different mechanism to bind to their target RNA instead of relying on the LaM like other LARPs.

The PAM2w domain of LARP4B is located at the N-terminus, where it is mostly intrinsically disordered, there are no defined structure elements or known RNA binding platforms such as RRMs, K Homology domains, DEAD motif or double-stranded RNA-binding motif (DSRM). However, there is an emerging field of study that looks at the IDRs and their impact with proteins and nucleic acids, there are a lot of protein-protein

and protein-DNA/RNA interactions that happen within these IDRs (Calabretta and Richard, 2015). It is recently accepted that these IDRs could even potentially adopt a defined structure while interacting with their targets, they may even function when disordered (Bondos, Dunker and Uversky, 2021). We can utilise tools like Nucleic Magnetic Resonance (NMR) or Small-angle X-ray scattering (SAXS) to study structures that are flexible and disordered. The NTR of LARP4B still has affinity to bind to AU-rich RNA (K_D= 5.5 μM), however the mutant L56AW63A (K_D= 0.64 μM), and the W63F (K_D= 0.76 μM) in the context of the NTD behaves quite like the NTD (K_D= 0.18 μM), unlike LARP4A. This could suggest that the PAM2w motif in LARP4B has less of a role for RNA binding, and solely focuses on the interaction with the MLLE domain of PABPC1, as shown by the pulldown experiments, where there is an association of NTD-MLLE seen, on the other hand the PAM2w mutants abrogated the protein-protein interaction.

LARP4B could potentially have other RNA targets, the PAR-CLIP experiment (Küspert *et al.*, 2015) resulted in many other promising targets that have not been tested *in vitro*. If a new target RNA of LARP4B is found, it could have another new role that influences the mRNA homeostasis. The EMSAs with the addition of tRNA mix shows us that there is definitely potential for another target that has not been discovered.

Overall, I propose a model of binding for LARP4B towards both of its RNA and protein partners (Figure 73), where in fact, most of the research focused in the N-terminus. We have shown by *in vitro* pulldowns that the PAM2w domain interacts with the MLLE of PABPC1, consistent with literature (Schäffler *et al.*, 2010; Grimm *et al.*, 2020). The NTD of LARP4B is required for maximum AU-rich RNA binding, keeping in mind that the RNA could bind with other domains within the NTD, albeit at a lower affinity. The C-terminus is less prioritised in this study, nevertheless LARP4B is known to bind to PABPC1 in a region called the poly(A)-binding protein interaction protein motif (PBM)

and interact with RACK1 in the RACK1-interaction region (RIR) (Schäffler *et al.*, 2010), in which the domain boundaries are not set and is a good direction for future studies in order to fully understand the function of LARP4B.

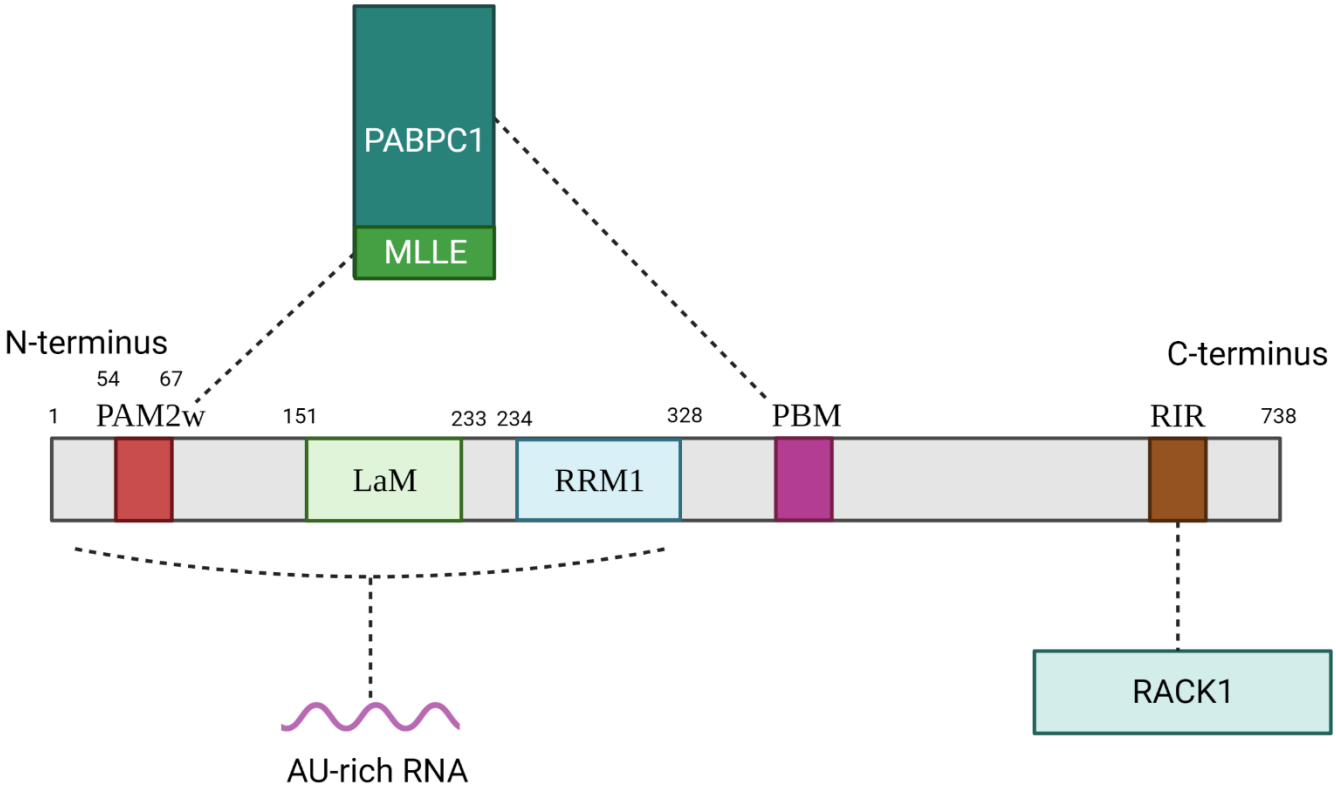


***Figure 73. A model for LARP4B and interaction with its partners**, the dotted lines indicate interaction with the RNA/Protein at that particular region. PAM2w: an atypical Poly(A) Binding Protein Interaction Motif-2, LaM: La Motif, RRM1: RNA Recognition Motif 1, PBM: Poly(A) Binding Protein Interaction Protein Motif, RACK1: Receptor of Activated protein C Kinase 1, RIR: RACK1-interaction region, PABPC1: Poly A Binding Protein Cytoplasmic 1, MLLE: Mademoiselle domain (Image created using BioRender, and the domain organisation is deduced from literature and the database from Uniprot).*

5.3 Project limitations and future work

There are two techniques in this study that allowed us to characterise the binding affinity of protein-RNA. EMSA was shown to be a reliable method that accurately provided reproducible data, MST was used as an alternative method, utilising the principle of thermophoresis that verified the results of the EMSA, and a useful technique for a large

set of mutants due to the speed of the technique. However the errors on a couple of datasets were too large, partially due to nucleic acid contamination and partially due to MST being less precise than EMSAs. The contaminated samples could be re-purified and re-tested via EMSAs or MST.

Most of the literature around LARP4s uses full-length constructs with human cell lines such as HEK293. Since our study uses recombinant protein generated from *E. coli*, attempts were made in order to purify full-length SUMO-tagged proteins of LARP4A and LARP4B, however in every attempt the yield was too low and the purification was unsuccessful due to the sheer size of the protein. The purification could be optimised by altering the protocol such as using a different soluble tag and/or expressing in different vectors and/or different cell lines if needed. Full length proteins are vital in order to study the protein-protein interaction regions in the C-terminus, such as the PBM and the RIR which are both poorly characterised. It is also unknown whether or not the C-terminus has any RNA binding capabilities.

Other regions of LARP4B that could potentially be looked at include the RRM which has shown the capability to bind to RNA. It would be interesting to test mutants in the RRM, in particular mutating the residues that are important for forming the conserved RNP-1 and RNP-2 sequences located in the β-sheets of the RRM and verifying if the RNP sequences are important regions facilitating the interaction to RNA.

Structural studies involving biophysical techniques such as NMR could be done to achieve a structure of LARP4B to further dissect the binding mechanism of the protein by comparing apo protein with RNA-bound protein. NMR is a potentially useful technique since the NTR is intrinsically disordered/flexible, one hypothesis is that the flexible regions could potentially move around or even fold back to the La-module which in turn facilitates the RNA binding. Using NMR, three-dimensional structures in their

natural state can be measured directly in solution and this can provide unique insights into dynamics and intramolecular interactions. Finally, the search for new RNA targets for LARP4B either through computational (Li *et al.*, 2014) or experimental methods (PAR-CLIP) with the aid of *in vitro* assays could be a chance to identify novel targets and roles for LARP4B.

Appendix

1. List of media used

Name	Recipe
Luria Broth (LB)	1% tryptone, 0.5% yeast extract, 1% NaCl
LB/agar	LB medium added with 1.5% agarose
Super Optimal Broth with Catabolite repression (SOC)	2% tryptone, 0.5% yeast extract, 10 mM NaCl, 10 mM MgCl2, 10 mM MgSO4, 20 mM D-glucose

2. List of buffers/solutions used

Name	Recipe
5X Loading Dye	0.25% Bromophenol Blue, 0.5 M DTT, 50% glycerol, 10% SDS, 0.25 M Tris-HCl, pH 6.8
CD buffer	20 mM Tris-HCl, 100 mM KCl,1 mM DTT, pH 7.25
Destaining solution	45% methanol, 45% H2O, 10% acetic acid
EMSA buffer	20 mM Tris-HCl, 100 mM KCl, 5% glycerol, 1 mM DTT, 0.1 mg/mL BSA. 0.01 mg/mL tRNA mix, pH 8.0
Final (protein) buffer	20 mM Tris-HCl, 100 mM KCl, 0.2 mM EDTA and 1 mM DTT, pH 7.25
IEX binding buffer	50 mM Tris-HCl, 100 mM KCl, 0.2 mM EDTA, 1 mM DTT, 10% glycerol, pH 7.25
IEX elution buffer	50 mM Tris-HCl, 2 M KCl, 0.2 mM EDTA, 1 mM DTT, 10% glycerol, pH 7.25

Lysis buffer	Nickel binding buffer with 2 mM PMSF, 0.01 g of lysozyme (Sigma), a tablet of cOmplete™ protease inhibitor cocktail (Roche) per 50 mL.
Nickel binding buffer	50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, 5 % glycerol, pH 8.0
Nickel elution buffer	50 mM Tris-HCl, 300 mM NaCl, 500 mM imidazole, 5 % glycerol, pH 8.0
Nickel dialysis buffer	50 mM Tris-HCl, 0.2 mM EDTA, 200 mM KCl, 1 mM DTT, pH 7.25
Resolving buffer	375 mM Tris-HCl, 0.1% SDS, pH 8.8
Running buffer	25 mM Tris-HCl, 192 mM Glycine, 0.1% SDS
Stacking buffer	125 mM Tris-HCl and 0.1% SDS, pH 6.9
Staining Solution	1% Coomassie Brilliant Blue© (Bio-Rad), 45% methanol, 45% H2O, 10% acetic acid

3. List of rotors used

Name	**Usage**
JLA 8.1000 (Beckman)	Used with 1L bottles up to 12000×g
Allegra X-22R (Beckman)	Used with 50mL tubes up to 3900×g
JA-20 (Beckman)	Used with 20mL tubes up to 48400×g
5415R (Eppendorf)	Used with 1.5mL Eppendorf tubes up to 13100×g

4. Details of proteins used

Construct	Molecular Weight (KDa)	Extinction Coefficient (M^{-1} cm^{-1})	Theoretical pI	Molecular Weight post-cleavage (KDa)	Extinction Coefficient post-cleavage (M^{-1} cm^{-1})	Theoretical pI post-cleavage
LARP4A 1-50	19433.62	12490	5.94	5292.87	11000	4.92
LARP4A 1-79	22595.99	15470	5.50	8455.24	13980	4.39
LARP4A NTD	34753.99	35410	4.98	32346.47	33920	4.78
LARP4A La module	22957.19	18450	6.07	20549.66	16960	5.83
LARP4A NTR	14336.45	18450	4.54	16015.00	12490	4.05
LARP4A L15AW22A	34596.78	29910	4.98	32189.25	28420	4.78
LARP4B NTD 1-328	64651.82	68300	4.97	36382.31	25440	4.52
LARP4B NTD 1-339	65933.26	71280	5.00	37663.74	28420	4.56
LARP4B La module	23370.36	14440	5.39	20587.50	12950	5.00
LARP4B NTR	44284.51	55350	4.92	16015.00	12490	4.05
LAR4PB L56AW63A	64494.61	62800	4.97	36225.10	19940	4.52
LAR4PB W63F	64612.79	62800	4.97	36343.27	19940	4.52
LARP4B 40-328	60454.25	68300	4.98	32184.73	25440	4.50
LARP4B 71-328	57178.50	62800	5.00	28908.98	19440	4.48
LARP4B 95-328	54761.00	57300	5.02	26491.49	14440	4.46
LARP4B NTD T163A	64621.8	68300	4.97	36352.28	25440	4.52
LARP4B NTD F166A	64575.73	68300	4.97	36306.21	25440	4.52
LARP4B NTD C167A	64619.76	68300	4.97	36350.25	25440	4.52
LARP4B NTD D176A	64607.81	68300	5	36338.3	25440	4.55
LARP4B NTD Y178A	64559.73	66810	4.97	36290.21	23950	4.52
LARP4B NTD L197A	64609.74	68300	4.97	36340.23	25440	4.52
LARP4B La module T163A	23184.15	14440	5.27	20529.41	12950	4.9
LARP4B La module F166A	23138.07	14440	5.27	20483.34	12950	4.9
LARP4B La module C167A	23182.11	14440	5.27	20527.38	12950	4.9
LARP4B La module D176A	23170.16	14440	5.39	20515.43	12950	4.99
LARP4B La module Y178A	23122.07	12950	5.27	20467.34	11460	4.9
LARP4B La module L197A	23172.09	14440	5.27	20517.36	12950	4.9
LARP4B LaM	12371.74	4470	5.18	9501.80	2980	4.59
LARP4B RRM	13886.57	11460	6.29	11016.63	9970	6.51
LARP4A FL	94746.97	61770	6.18	80606.22	60280	6.20
LARP4B FL	94692.76	50310	6.31	80552.01	48820	6.48
His SUMO MLLE	23141.23	1490	5.49	n/a	n/a	n/a
SUMO alone	16835.87	1490	6.53	n/a	n/a	n/a

Bibliography

Abeliovich, H. (2005) ‘An Empirical Extremum Principle for the Hill Coefficient in Ligand-Protein Interactions Showing Negative Cooperativity’, *Biophysical Journal*, 89(1), p. 76. Available at: https://doi.org/10.1529/BIOPHYSJ.105.060194.

Acikara, O.B. (2013) ‘Ion-Exchange Chromatography and Its Applications’, *Column Chromatography* [Preprint]. Available at: https://doi.org/10.5772/55744.

Al-Ashtal, H.A. *et al.* (2021) ‘The LARP1 La-Module recognizes both ends of TOP mRNAs’, *RNA Biology*, 18(2), pp. 248–258. Available at: https://doi.org/10.1080/15476286.2019.1669404/SUPPL_FILE/KRNB_A_1669404_SM8752.ZIP.

Alfano, C. *et al.* (2004) ‘Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein’, *NATURE STRUCTURAL & MOLECULAR BIOLOGY*, 11(4). Available at: https://doi.org/10.1038/nsmb747.

Alspaugh, M.A., Talal, N. and Tan, E.M. (1976) ‘Differentiation and characterization of autoantibodies and their antigens in Sjögren’s syndrome’, *Arthritis and rheumatism*, 19(2), pp. 216–222. Available at: https://doi.org/10.1002/ART.1780190214.

Al-Tubuly, A.A. (2000) ‘SDS-PAGE and Western Blotting’, pp. 391–405. Available at: https://doi.org/10.1385/1-59259-076-4:391/TABLES/5.

Angenstein, F. *et al.* (2002) ‘A receptor for activated C kinase is part of messenger ribonucleoprotein complexes associated with polyA-mRNAs in neurons’, *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 22(20), pp. 8827–8837. Available at: https://doi.org/10.1523/JNEUROSCI.22-20-08827.2002.

Bai, S.W. *et al.* (2011) ‘Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration’, *BMC Biology*, 9(1), pp. 1–18. Available at: https://doi.org/10.1186/1741-7007-9-54/FIGURES/8.

Bayfield, M.A., Yang, R. and Maraia, R.J. (2010) ‘Conserved and divergent features of the structure and function of La and La-related proteins (LARPs)’, *Biochimica et biophysica acta*, 1799(5–6), pp. 365–378. Available at: https://doi.org/10.1016/J.BBAGRM.2010.01.011.

Bondos, S.E., Dunker, A.K. and Uversky, V.N. (2021) 'On the roles of intrinsically disordered proteins and regions in cell communication and signaling', *Cell Communication and Signaling*, 19(1), pp. 1–9. Available at: https://doi.org/10.1186/S12964-021-00774-3/TABLES/2.

Bornhorst, J.A. and Falke, J.J. (2010) *Purification of Proteins Using Polyhistidine Affinity Tags*.

Bousquet-Antonelli, C. and Deragon, J.M. (2009) 'A comprehensive analysis of the La-motif protein superfamily', *RNA (New York, N.Y.)*, 15(5), pp. 750–764. Available at: https://doi.org/10.1261/RNA.1478709.

Calabretta, S. and Richard, S. (2015) 'Emerging Roles of Disordered Sequences in RNA-Binding Proteins', *Trends in Biochemical Sciences*, 40(11), pp. 662–672. Available at: https://doi.org/10.1016/J.TIBS.2015.08.012.

Capra, J.A. and Singh, M. (2007) 'Predicting functionally important residues from sequence conservation', *Bioinformatics*, 23(15), pp. 1875–1882. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTM270.

Cardinali, B. *et al.* (2003) 'La Protein Is Associated with Terminal Oligopyrimidine mRNAs in Actively Translating Polysomes *', *Journal of Biological Chemistry*, 278(37), pp. 35145–35151. Available at: https://doi.org/10.1074/JBC.M300722200.

Casamassimi, A. and Ciccodicola, A. (2019) 'Transcriptional Regulation: Molecules, Involved Mechanisms, and Misregulation', *International Journal of Molecular Sciences*, 20(6). Available at: https://doi.org/10.3390/IJMS20061281.

Celie, P.H.N., Parret, A.H.A. and Perrakis, A. (2016) 'Recombinant cloning strategies for protein expression', *Current Opinion in Structural Biology*, 38, pp. 145–154. Available at: https://doi.org/10.1016/J.SBI.2016.06.010.

Černý, J. and Hobza, P. (2007) 'Non-covalent interactions in biomacromolecules', *Physical Chemistry Chemical Physics*, 9(39), pp. 5291–5303. Available at: https://doi.org/10.1039/B704781A.

Choi, B., Rempala, G.A. and Kim, J.K. (2017) 'Beyond the Michaelis-Menten equation: Accurate and efficient estimation of enzyme kinetic parameters', *Scientific Reports 2017 7:1*, 7(1), pp. 1–11. Available at: https://doi.org/10.1038/s41598-017-17072-z.

Cléry, A., Blatter, M. and Allain, F.H.-T. (2008) 'RNA recognition motifs: boring? Not quite', *Current Opinion in Structural Biology*, 18(3), pp. 290–298. Available at: https://doi.org/https://doi.org/10.1016/j.sbi.2008.04.002.

Cléry, A., Blatter, M. and Allain, F.H.T. (2008) 'RNA recognition motifs: boring? Not quite', *Current opinion in structural biology*, 18(3), pp. 290–298. Available at: https://doi.org/10.1016/J.SBI.2008.04.002.

Corradini, C., Cavazza, A. and Bignardi, C. (2012) 'High-Performance Anion-Exchange Chromatography Coupled with Pulsed Electrochemical Detection as a Powerful Tool to Evaluate Carbohydrates of Food Interest: Principles and Applications', *International Journal of Carbohydrate Chemistry*, 2012, pp. 1–13. Available at: https://doi.org/10.1155/2012/487564.

Crick, F. (1970) *Central Dogma of Molecular Biology*, *NATURE*.

Cruz-Gallardo, I. *et al.* (2019) 'LARP4A recognizes polyA RNA via a novel binding mechanism mediated by disordered regions and involving the PAM2w motif, revealing interplay between PABP, LARP4A and mRNA', *Nucleic Acids Research*, 47(8), pp. 4272–4291. Available at: https://doi.org/10.1093/NAR/GKZ144.

Daubner, G.M., Cléry, A. and Allain, F.H.-T. (2013) 'RRM–RNA recognition: NMR or crystallography…and new findings', *Current Opinion in Structural Biology*, 23(1), pp. 100–108. Available at: https://doi.org/https://doi.org/10.1016/j.sbi.2012.11.006.

Dock-Bregeon, A.C., Lewis, K.A. and Conte, M.R. (2021) 'The La-related proteins: structures and interactions of a versatile superfamily of RNA-binding proteins', *RNA biology*, 18(2), pp. 178–193. Available at: https://doi.org/10.1080/15476286.2019.1695712.

Dong, G. *et al.* (2004) 'Structure of the La motif: A winged helix domain mediates RNA binding via a conserved aromatic patch', *EMBO Journal*, 23(5), pp. 1000–1007. Available at: https://doi.org/10.1038/SJ.EMBOJ.7600115.

Dreyfuss, G., Swanson, M.S. and Piñol-Roma, S. (1988) 'Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation', *Trends in Biochemical Sciences*, 13(3), pp. 86–91. Available at: https://doi.org/https://doi.org/10.1016/0968-0004(88)90046-1.

Edelheit, O., Hanukoglu, A. and Hanukoglu, I. (2009) 'Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies', *BMC Biotechnology*, 9(1), pp. 1–8. Available at: https://doi.org/10.1186/1472-6750-9-61/FIGURES/3.

Fred Griffith, B. (1928) 'THE SIGNIFICANCE OF PNEUMOCOCCAL TYPES', (2). Available at: https://doi.org/10.1017/S0022172400031879.

Froger, A. and Hall, J.E. (2007) 'Transformation of Plasmid DNA into E. coli Using the Heat Shock Method', *Journal of Visualized Experiments : JoVE*, 6(6). Available at: https://doi.org/10.3791/253.

Gasteiger, E. *et al.* (2005) 'Protein Identification and Analysis Tools on the ExPASy Server', *The Proteomics Protocols Handbook*, pp. 571–607. Available at: https://doi.org/10.1385/1-59259-890-0:571.

Gebauer, F. *et al.* (2020) 'RNA-binding proteins in human genetic disease', *Nature Reviews Genetics 2020 22:3*, 22(3), pp. 185–198. Available at: https://doi.org/10.1038/s41576-020-00302-y.

Glisovic, T. *et al.* (2008) 'RNA-binding proteins and post-transcriptional gene regulation', *FEBS letters*, 582(14), p. 1977. Available at: https://doi.org/10.1016/J.FEBSLET.2008.03.004.

Gopal, G.J. and Kumar, A. (2013) 'Strategies for the production of recombinant protein in Escherichia coli', *The protein journal*, 32(6), pp. 419–425. Available at: https://doi.org/10.1007/S10930-013-9502-5.

Goss, D.J. and Kleiman, F.E. (2013) 'Poly(A) binding proteins: are they all created equal?', *Wiley interdisciplinary reviews. RNA*, 4(2), pp. 167–179. Available at: https://doi.org/10.1002/WRNA.1151.

Greenfield, N.J. (2007) 'Using circular dichroism spectra to estimate protein secondary structure', *Nature Protocols 2007 1:6*, 1(6), pp. 2876–2890. Available at: https://doi.org/10.1038/nprot.2006.202.

Grimm, C. *et al.* (2020) 'Crystal Structure of a Variant PAM2 Motif of LARP4B Bound to the MLLE Domain of PABPC1', *Biomolecules 2020, Vol. 10, Page 872*, 10(6), p. 872. Available at: https://doi.org/10.3390/BIOM10060872.

Hellman, L.M. and Fried, M.G. (2007) 'Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions'. Available at: https://doi.org/10.1038/nprot.2007.249.

Hentze, M.W. *et al.* (2018) 'A brave new world of RNA-binding proteins', *Nature reviews. Molecular cell biology*, 19(5), pp. 327–341. Available at: https://doi.org/10.1038/NRM.2017.130.

Hickey, C.M., Wilson, N.R. and Hochstrasser, M. (2012) 'Function and Regulation of SUMO Proteases', *Nature reviews. Molecular cell biology*, 13(12), p. 755. Available at: https://doi.org/10.1038/NRM3478.

Jerabek-Willemsen, M. *et al.* (2014) 'MicroScale Thermophoresis: Interaction analysis and beyond', *Journal of Molecular Structure*, 1077, pp. 101–113. Available at: https://doi.org/10.1016/J.MOLSTRUC.2014.03.009.

Jungbauer, A. and Hahn, R. (2009) 'Ion-Exchange Chromatography', *Methods in Enzymology*, 463(C), pp. 349–371. Available at: https://doi.org/10.1016/S0076-6879(09)63022-6.

Kenan, D.J., Query, C.C. and Keene, J.D. (1991) 'RNA recognition: towards identifying determinants of specificity', *Trends in Biochemical Sciences*, 16(C), pp. 214–220. Available at: https://doi.org/10.1016/0968-0004(91)90088-D.

Koso, H. *et al.* (2016) 'Identification of RNA-Binding Protein LARP4B as a Tumor Suppressor in Glioma', *Cancer research*, 76(8), pp. 2254–2264. Available at: https://doi.org/10.1158/0008-5472.CAN-15-2308.

Kotik-Kogan, O. *et al.* (2008) 'Structural analysis reveals conformational plasticity in the recognition of RNA 3' ends by the human La protein', *Structure (London, England : 1993)*, 16(6), pp. 852–862. Available at: https://doi.org/10.1016/J.STR.2008.02.021.

Kozlov, G. *et al.* (2010) 'Structural basis of binding of P-body-associated proteins GW182 and ataxin-2 by the Mlle domain of poly(A)-binding protein', *The Journal of biological chemistry*, 285(18), pp. 13599–13606. Available at: https://doi.org/10.1074/JBC.M109.089540.

Kühn, U. and Pieler, T. (1996) 'XenopusPoly(A) Binding Protein: Functional Domains in RNA Binding and Protein – Protein Interaction', *Journal of Molecular Biology*, 256(1), pp. 20–30. Available at: https://doi.org/https://doi.org/10.1006/jmbi.1996.0065.

Kühn, U. and Wahle, E. (2004) 'Structure and function of poly(A) binding proteins', *Biochimica et biophysica acta*, 1678(2–3), pp. 67–84. Available at: https://doi.org/10.1016/J.BBAEXP.2004.03.008.

Küspert, M. *et al.* (2015) 'LARP4B is an AU-rich sequence associated factor that promotes mRNA accumulation and translation', *RNA*, 21(7), pp. 1294–1305. Available at: https://doi.org/10.1261/RNA.051441.115.

Li, X. *et al.* (2014) 'Finding the target sites of RNA-binding proteins', *Wiley Interdisciplinary Reviews. RNA*, 5(1), p. 111. Available at: https://doi.org/10.1002/WRNA.1201.

Lozano Terol, G. *et al.* (2021) 'Impact of the Expression System on Recombinant Protein Production in Escherichia coli BL21', *Frontiers in Microbiology*, 12, p. 1511. Available at: https://doi.org/10.3389/FMICB.2021.682001/BIBTEX.

Lukong, K.E. *et al.* (2008) 'RNA-binding proteins in human genetic disease', *Trends in genetics : TIG*, 24(8), pp. 416–425. Available at: https://doi.org/10.1016/J.TIG.2008.05.004.

Lunde, B.M., Moore, C. and Varani, G. (2007) 'RNA-binding proteins: modular design for efficient function', *Nature reviews. Molecular cell biology*, 8(6), pp. 479–490. Available at: https://doi.org/10.1038/NRM2178.

Luo, J., Ying, K. and Bai, J. (2005) 'Savitzky-Golay smoothing and differentiation filter for even number data Photoacoustic View project Carotid Elastography View project Savitzky-Golay smoothing and differentiation filter for even number data', *Signal Processing*, 85, pp. 1429–1434. Available at: https://doi.org/10.1016/j.sigpro.2005.02.002.

Mangus, D.A., Evans, M.C. and Jacobson, A. (2003) 'Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression', *Genome biology*, 4(7). Available at: https://doi.org/10.1186/GB-2003-4-7-223.

Maraia, R.J. *et al.* (2017) 'The LARPs, La and related RNA-binding proteins: Structures, functions and evolving perspectives HHS Public Access Introduction and Background', *Wiley Interdiscip Rev RNA*, 8(6). Available at: https://doi.org/10.1002/wrna.1430.

Maris, C., Dominguez, C. and Allain, F.H.T. (2005) 'The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression', *The FEBS Journal*, 272(9), pp. 2118–2131. Available at: https://doi.org/10.1111/J.1742-4658.2005.04653.X.

Martino, L. *et al.* (2015a) 'Synergic interplay of the La motif, RRM1 and the interdomain linker of LARP6 in the recognition of collagen mRNA expands the RNA binding repertoire of the La module', *Nucleic acids research*, 43(1), pp. 645–660. Available at: https://doi.org/10.1093/NAR/GKU1287.

Martino, L. *et al.* (2015b) 'Synergic interplay of the La motif, RRM1 and the interdomain linker of LARP6 in the recognition of collagen mRNA expands the RNA binding repertoire of the La module', *Nucleic Acids Research*, 43(1), p. 645. Available at: https://doi.org/10.1093/NAR/GKU1287.

Martino, L. *et al.* (2015c) 'Synergic interplay of the La motif, RRM1 and the interdomain linker of LARP6 in the recognition of collagen mRNA expands the RNA binding repertoire of the La module', *Nucleic acids research*, 43(1), pp. 645–660. Available at: https://doi.org/10.1093/NAR/GKU1287.

Mattijssen, S. *et al.* (2017) 'LARP4 mRNA codon-tRNA match contributes to LARP4 activity for ribosomal protein mRNA poly(A) tail length protection', *eLife*, 6. Available at: https://doi.org/10.7554/ELIFE.28889.

Mattijssen, S. *et al.* (2021) 'LARP1 and LARP4: up close with PABP for mRNA 3' poly(A) protection and stabilization', *RNA Biology*, 18(2), p. 259. Available at: https://doi.org/10.1080/15476286.2020.1868753.

Merret, R. *et al.* (2013) 'The association of a La module with the PABP-interacting motif PAM2 is a recurrent evolutionary process that led to the neofunctionalization of La-related proteins', *RNA (New York, N.Y.)*, 19(1), pp. 36–50. Available at: https://doi.org/10.1261/RNA.035469.112.

Micsonai, A. *et al.* (2015) 'Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy', *Proceedings of the National Academy of Sciences of the United States of America*, 112(24), pp. E3095–E3103. Available at: https://doi.org/10.1073/PNAS.1500851112/-/DCSUPPLEMENTAL.

Morrison, K.L. and Weiss, G.A. (2001) 'Combinatorial alanine-scanning', *Current opinion in chemical biology*, 5(3), pp. 302–307. Available at: https://doi.org/10.1016/S1367-5931(00)00206-4.

Müller, S. *et al.* (2001) 'SUMO, ubiquitin's mysterious cousin', *Nature reviews. Molecular cell biology*, 2(3), pp. 202–210. Available at: https://doi.org/10.1038/35056591.

Nagai, K. (1996) 'RNA—protein complexes', *Current Opinion in Structural Biology*, 6(1), pp. 53–61. Available at: https://doi.org/10.1016/S0959-440X(96)80095-9.

Nakata, Y., Tang, X. and Yokoyama, K.K. (1997) 'Preparation of Competent Cells for High-Efficiency Plasmid Transformation of Escherichia coli', *Methods in molecular biology (Clifton, N.J.)*, 69, pp. 129–137. Available at: https://doi.org/10.1385/0-89603-383-X:129.

Nicholson, A.L. and Pasquinelli, A.E. (2019) 'Tales of Detailed Poly(A) Tails', *Trends in cell biology*, 29(3), pp. 191–200. Available at: https://doi.org/10.1016/J.TCB.2018.11.002.

Palazzo, A.F. and Lee, E.S. (2015) 'Non-coding RNA: What is functional and what is junk?', *Frontiers in Genetics*, 5(JAN), p. 2. Available at: https://doi.org/10.3389/FGENE.2015.00002/BIBTEX.

Patel, G.P. and Bag, J. (2006) 'IMP1 interacts with poly(A)-binding protein (PABP) and the autoregulatory translational control element of PABP-mRNA through the KH III-IV domain', *The FEBS Journal*, 273(24), pp. 5678–5690. Available at: https://doi.org/10.1111/J.1742-4658.2006.05556.X.

Plach, M.G., Grasser, K. and Schubert, T. (2017) 'MicroScale Thermophoresis as a Tool to Study Protein-peptide Interactions in the Context of Large Eukaryotic Protein Complexes', *Bio-protocol*, 7(23). Available at: https://doi.org/10.21769/BIOPROTOC.2632.

Pushkaran, A.C. *et al.* (2015) 'Understanding the structure-function relationship of lysozyme resistance in Staphylococcus aureus by peptidoglycan o-acetylation using molecular docking, dynamics, and lysis assay', *Journal of Chemical Information and Modeling*, 55(4), pp. 760–770. Available at: https://doi.org/10.1021/CI500734K/SUPPL_FILE/CI500734K_SI_001.PDF.

Re, A. *et al.* (2014) 'RNA–Protein Interactions: An Overview', pp. 491–521. Available at: https://doi.org/10.1007/978-1-62703-709-9_23.

Rial, D. v. and Ceccarelli, E.A. (2002) 'Removal of DnaK contamination during fusion protein purifications', *Protein expression and purification*, 25(3), pp. 503–507. Available at: https://doi.org/10.1016/S1046-5928(02)00024-4.

Rodger, A. and Marshall, D. (2021) 'Beginners guide to circular dichroism', *The Biochemist*, 43(2), pp. 58–64. Available at: https://doi.org/10.1042/BIO_2020_105.

Roux, K.H. (2009) 'Optimization and troubleshooting in PCR', *Cold Spring Harbor protocols*, 2009(4). Available at: https://doi.org/10.1101/PDB.IP66.

Schäffler, K. *et al.* (2010) 'A stimulatory role for the La-related protein 4B in translation', *RNA (New York, N.Y.)*, 16(8), pp. 1488–1499. Available at: https://doi.org/10.1261/RNA.2146910.

Seetharaman, S. *et al.* (2016) 'The RNA-binding protein LARP4 regulates cancer cell migration and invasion', *Cytoskeleton (Hoboken, N.j.)*, 73(11), p. 680. Available at: https://doi.org/10.1002/CM.21336.

Shapiro, A.L., Viñuela, E. and v. Maizel Jr., J. (1967) 'Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels', *Biochemical and Biophysical Research Communications*, 28(5), pp. 815–820. Available at: https://doi.org/10.1016/0006-291X(67)90391-9.

Sonenberg, N. and Hinnebusch, A.G. (2009) 'Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets', *Cell*, 136(4), p. 731. Available at: https://doi.org/10.1016/J.CELL.2009.01.042.

Stavraka, C. and Blagden, S. (2015) 'The La-Related Proteins, a Family with Connections to Cancer', *Biomolecules*, 5(4), pp. 2701–2722. Available at: https://doi.org/10.3390/BIOM5042701.

Swinehart, D.F. (1962) 'The Beer-Lambert law', *Journal of Chemical Education*, 39(7), pp. 333–335. Available at: https://doi.org/10.1021/ED039P333.

Teplova, M. *et al.* (2006) 'Structural basis for recognition and sequestration of UUU(OH) 3' temini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen', *Molecular cell*, 21(1), pp. 75–85. Available at: https://doi.org/10.1016/J.MOLCEL.2005.10.027.

Thompson, M.K. and Gilbert, W. v. (2017) 'mRNA length-sensing in eukaryotic translation: reconsidering the "closed loop" and its implications for translational control', *Current genetics*, 63(4), pp. 613–620. Available at: https://doi.org/10.1007/S00294-016-0674-3.

Uchikawa, E. *et al.* (2015) 'Structural insight into the mechanism of stabilization of the 7SK small nuclear RNA by LARP7', *Nucleic acids research*, 43(6), pp. 3373–3388. Available at: https://doi.org/10.1093/NAR/GKV173.

Ullah, R. *et al.* (2016) 'Activity of the Human Rhinovirus 3C Protease Studied in Various Buffers, Additives and Detergents Solutions for Recombinant Protein Production', *PLoS ONE*, 11(4). Available at: https://doi.org/10.1371/JOURNAL.PONE.0153436.

Wolin, S.L. and Cedervall, T. (2002) 'The La protein', *Annual review of biochemistry*, 71, pp. 375–403. Available at: https://doi.org/10.1146/ANNUREV.BIOCHEM.71.090501.150003.

Wolozin, B. and Apicco, D. (2015) 'RNA binding proteins and the genesis of neurodegenerative diseases', *Advances in Experimental Medicine and Biology*, 822, pp. 11–15. Available at: https://doi.org/10.1007/978-3-319-08927-0_3/FIGURES/1.

Xiong, S., Zhang, L. and He, Q.Y. (2008) 'Fractionation of proteins by heparin chromatography.', *Methods in molecular biology (Clifton, N.J.)*, 424, pp. 213–221. Available at: https://doi.org/10.1007/978-1-60327-064-9_18/FIGURES/18_2_978-1-60327-064-9.

Yang, R. *et al.* (2011) 'La-Related Protein 4 Binds Poly(A), Interacts with the Poly(A)-Binding Protein MLLE Domain via a Variant PAM2w Motif, and Can Promote mRNA Stability †', *MOLECULAR AND CELLULAR BIOLOGY*, 31(3), pp. 542–556. Available at: https://doi.org/10.1128/MCB.01162-10.

Zhang, Y. *et al.* (2015) 'La-related protein 4B maintains murine MLL-AF9 leukemia stem cell self-renewal by regulating cell cycle progression', *Experimental Hematology*, 43(4), pp. 309-318.e2. Available at: https://doi.org/10.1016/j.exphem.2014.12.003.

Zhou, H.L. *et al.* (2014) 'RNA-binding proteins in neurological diseases', *Science China. Life sciences*, 57(4), pp. 432–444. Available at: https://doi.org/10.1007/S11427-014-4647-9.