NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF CHEMISTRY

CLÁUDIO ALEXANDRE LOPES COLAÇO

BSc in Biochemistry

Understanding the detailed mechanism of interaction between La-related proteins and their RNA targets

MASTER IN BIOCHEMISTRY

NOVA University Lisbon

December, 2021

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENTO DE ⟨NOME DO DEPARTAMENTO⟩

Understanding the detailed mechanism of interaction between La-related proteins and their RNA targets

"The cosmos is within us. We are made of star-stuff. We are a way for the universe to know itself." Carl Sagan

CLÁUDIO ALEXANDRE LOPES COLAÇO
BSc in Biochemistry

Adviser: Maria R. Conte
School Academic Lead, King's College London

Examination Committee:

Chair: Ricardo Franco
Associated Professor with Aggregation, Nova School of Science and Technology

Rapporteurs: Aldino Viegas
Assistant Investigator, Nova School of Science and Technology

Adviser: Maria R. Conte
School Academic Lead, King's College London

MASTER IN BIOCHEMISTRY

NOVA University Lisbon
December, 2021

Understanding the detailed mechanism of interaction between La-related proteins and their RNA targets

To you, the one with which the story never ends and to the one who is still to come

ACKNOWLEDGMENTS

Since the very beginning of the whole adventure that was this dissertation, I have received a great amount of love, support and assistance.
To acknowledge all this help, I would like to thank:

My supervisor, professor Sasi Conte, for taking a gamble on a student coming from far away and that you knew nothing about. You started by accepting me, then helped me settle in the group and finally by being by my side with all the scientific, but also emotional support, I cannot thank you enough. It was a challenging journey, but you were always available to help me with my doubts and setbacks. For all that I thank you.

James Jarvis for the endless meetings, both in person and online, to help me with all the NMR assignment. Thank you for the help and for your friendship.

All the lab members of the Conte Group, especially Yifei Gu for always being there to help me during my laboratory work, Jennifer Coleman for helping me while writing this dissertation, Federica Capraro for all the flash freezing adventures, Jing Sun for all the laughs and deep talks while waiting for our cultures to grow.

I also wish to show my appreciation to Shabir Najmudin, who was a very important factor in my adaptation to London and to King's College.

Lastly, I want to thank King's College London for taking a chance on me and accepting me as an Erasmus student.

(Now in Portuguese)

Queria deixar o meu agradecimento à instituição que durante todos estes anos chamei casa e que me deu tanto: a Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa. Foi aqui que fiz amigos para a vida, que obtive conhecimentos que me ajudaram a ultrapassar obstáculos desta tese e que me permitiu ser a pessoa que sou hoje. Assim, posso com muito orgulho dizer que faço, e sempre farei parte, da Gloriosa FCT.

Ao coordenador do mestrado em Bioquímica, o Professor Ricardo Franco, por me incentivar desde o primeiro momento a desenvolver uma tese fora, bem como durante a tese

quando os obstáculos pareciam demasiado grandes para serem ultrapassados, um grande obrigado.

Agora numa nota mais pessoal, uma vez que seria impossível não agradecer às pessoas que me ajudaram de um ponto de vista mais emocional e psicológico acreditando sempre em mim.

Em primeiro lugar aos meus amigos: obrigado a todos vocês, mas em especial ao Diogo Mendes por partilhares as dores de escrever a tese todas as noites destes últimos meses, à Carlota Garcia e à Patrícia Rodrigues por me fazerem companhia mesmo a longa distância e por último à Mariana Calmão que foi a minha companheira de Erasmus apesar de estar em Espanha. Todas as chamadas, videochamadas e Zoom calls não serão esquecidas: és uma excelente amiga. Uma pequena nota para a minha Rita Benvinda que me ajudou a planear a escrita desta tese, bem como a animar-me quando as coisas estavam mais difíceis.

Aos meus padrinhos de Praxe, Beatriz, Cristiana, Martim e Miguel, obrigado por acreditarem neste miúdo e por me aguentarem estes anos todos. Prometo ser menos chato no futuro.

Aos meus avós, Zulmira e Zé, por estarem sempre preocupados comigo e por todo o amor que têm por mim. Sei que sou um pouco esquecido, mas é apenas na cabeça porque o coração não esquece.

A toda a minha família um grande obrigado, em especial ao meu irmão Bruno e aos meus sobrinhos Matilde, Gonçalo, Vicente e Gabriel.

Para finalizar, agradeço:

Aos meus Pais. Sei que não sou o filho perfeito, mas mesmo assim estão sempre dispostos a ajudar-me e nunca meteram uma única barreira nas minhas escolhas académicas. Obrigado por tudo o que fazem por mim e por darem sempre o melhor que sabem para que eu chegue sempre mais longe. Amo-vos com todo o meu ser.

À minha irmã Sofia. Já o disse antes e continuo a dizer: és o maior modelo que tenho na vida. Aquilo que quero para mim desejo-te a ti em dobro. Obrigado por estares sempre pronta a apoiar-me mesmo que às vezes não me exponha totalmente. Amo-te muito e finalmente o Sporting foi campeão.

À Anabela e ao Ilídio, por acolherem um rapaz na sua primeira aventura fora de casa. Foram como pais para mim e sei que pude sempre contar com o vosso apoio para o que quer que precisasse. Obrigado por todos os sacrifícios que fizeram por mim.

À minha sobrinha Beatriz, obrigado por seres o maior raio de sol que a minha vida podia ter.

À minha afilhada, Beatriz Jesus, obrigado. És um orgulho enorme para este rapaz que teve a sorte de conhecer uma pessoa tão incrível como tu e que faz questão de me apoiar em todas as minhas aventuras. Adoro-te.

E um último obrigado ao meu Sporting Clube de Portugal por me ter dado tantas alegrias neste ano tão desafiante.

A todos os mencionados e a todos os que me esqueci, mas que mereciam estar aqui,
Um grande obrigado e que escolham ser sempre felizes.

"Imprisonment. What a curious principle. We confine the physical body, yet the mind is still free."
Arcane

"The cosmos is within us. We are made of star-stuff. We are a way for the universe to know itself."
Carl Sagan

ABSTRACT

The importance of maturation and processing of all RNA molecules is directly related to RNA-binding proteins (RBPs) and their RNA-binding domains (RBDs). Amongst this class of proteins is the La-related proteins (LaRPs) superfamily, that is characterized by a conserved RNA-binding unit labeled the La-module, which is composed by the La motif (LaM) and the RNA-recognition motif 1 (RRM1). The work presented in this report concerns LaRP4B and its relationship with LaRP4A, since both proteins have identified functions related to cancer and further knowledge about them is crucial in order to find possible treatments. The expression and purification of LaRP4B La-module and RRM1 constructs was successful and allowed to produce ^{15}N RRM for future NMR experiments. While using data from previous experiences, the backbone assignment of the RRM was done with a completion of almost 38% of all residues. This assignment also permitted to identify shifts that belonged to the tag present in the RRM. These findings helped to establish a new purification protocol for the RRM so that the tag can be cleaved by introducing an on-column cleavage step. Finally, by utilizing the chemical shifts of the LaRP4B RRM and La motif, obtained by other group members, a Chemical Shift Index study allowed the comparison between the predicted secondary structure of these domains and the ones from LaRP4A. This comparison showed no major differences between both structures that could indicate why the RNA targets are so different between these two closely related members of the LaRP4 family. Further studies are necessary to unveil the mechanism of interaction between LaRP4B and RNA.

Keywords: *La Related Proteins; La-module; La motif; RNA Recognition Motif; LaRP4B; RNA Binding Proteins*

RESUMO

O processamento e maturação de RNAs está intimamente ligada às proteínas que ligam RNA e aos seus domínios. A superfamília das proteínas relacionadas com o módulo La (LaRPs) faz parte desta classe de proteínas e é normalmente caracterizada por uma unidade ligação a RNAs bastante conservada denominada módulo La, que é formado pelo domínio La (LaM) e pelo domínio de reconhecimento de RNA 1 (RRM1). O trabalho apresentado neste relatório está relacionado com a proteína LaRP4B e a sua relação com a proteína LaRP4A, uma vez que ambas as proteínas foram já identificadas com tendo algum tipo de influência em diferentes tipos de cancro e é necessário aprofundar o conhecimento sobre ambas de forma a poder encontrar possíveis terapias. A expressão e purificação dos mutantes da proteína LaRP4B (módulo La e RRM1) foi bem-sucedida e levou à produção de RRM marcado isotopicamente com ^{15}N para ser utilizado em experiências de RMN futuras. Utilizando dados de experiências feitas de forma prévia ao trabalho aqui apresentado, foi possível fazer a atribuição parcial (aproximadamente 38%) dos átomos da cadeia principal do RRM da proteína LaRP4B. Esta atribuição permitiu ainda identificar certos desvios que pertenciam à tag presente no RRM. Assim foi possível desenvolver um novo protocolo de purificação para o RRM introduzindo um passo de clivagem da tag em coluna. Em último lugar, utilizando os desvios previamente obtidos tanto do RRM como do LaM da proteína LaRP4B, um estudo do Chemical Shift Index (CSI) permitiu comparar a estrutura secundária destes módulos com os mesmos, mas da proteína LaRP4A. Esta comparação demonstrou que não existem diferenças significativas entre as duas estruturas que possam indicar o porquê dos RNAs que cada uma liga serem tão diferentes apesar de as duas proteínas serem membros muito próximos da família LaRP4. Mais estudos serão necessários de forma a entender o mecanismo de interação entre a proteína LaRP4B e o seu RNA alvo.

Palavras-chave: *La Related Proteins; La-module; La motif; RNA Recognition Motif; LaRP4B; RNA Binding Proteins*

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

ABBREVIATION LIST

AML – Acute Myeloid Leukaemia
CBR – Conserved Basic Region
CSI – Chemical Shift Index
dsRBD – double-stranded RNA-binding domain
dsRNA – double-stranded RNA
DTT – Dithiothreitol
EDTA – Ethylenediaminetetraacetic Acid
EMSA – Electrophoretic Mobility Shift Assay
FMRP – Fragile X Mental Retardation Protein
HCV – Hepatitis-C Virus
HsLa – Human La Protein
HsLaRP4A – Human La Related Protein 4A
HsLaRP6 – Human La Related Protein 6
HsLaRP7 – Human La Related Protein 7
IDR – Intrinsically Disordered Regions
IPTG – Isopropyl β-D-1-thiogalactopyranoside
IRES – Internal Ribosome Entry Site
IRP1 – Iron-Regulatory Protein 1
ITAF – IRES-Transacting Factor
KH – K Homology domain
LaM – La Motif
LaRP – La Related Protein
LaRP1 – La Related Protein 1
LaRP3 – La Related Protein 3
LaRP4 – La Related Protein 4
LaRP4A – La Related Protein 4A
LaRP4B – La Related Protein 4B

LaRP6 – La Related Protein 6

LaRP7 – La Related Protein 7

lncRNA – long non-coding RNA

LSA – La- and S1- Associated motif

mRNA – messenger RNA

NES – Nuclear Export Signal

NLS – Nuclear Localization Signal

NMR – Nuclear Magnetic Resonance

NRE – Nuclear Retention Element

Nt – Nucleotide

NTR – N-Terminal Region

PABP – Poly(A)-Binding Protein

PBM – PABP-Binding Motif

PDB – Protein Data Bank

PKR – Protein Kinase RNA-activated

PMSF – Phenylmethylsulphonyl Fluoride

pol III – polymerase III

poly(A) – polyadenylated

poly(U) – poly-uridine

PTM – Post-Translational Modifications

Q-MS – Quantitative Mass Spectrometry

RACK1 – Receptor for Activated protein Kinase C 1

RBD – RNA-Binding Domain

RBP – RNA-Binding Protein

RBR-ID – Proteomic Identification of RNA-Binding Regions

RHA – RNA Helicase A

RIC – RNA Interactome Capture

RIR – RACK Interacting motif

RNA – Ribonucleic Acid

RNP – Ribonucleoprotein

RP-mRNA – Ribosomal Protein mRNA

RRM – RNA Recognition Motif

SBM – Short Basic Motif

SDS-PAGE – Sodium Dodecyl Sulphate–Polyacrylamide Gel Electrophoresis

snoRNA – small-nucleolar RNA

snRNA – small-nuclear RNA

ssRNA – single-stranded RNA

STRAP – Serine/Threonine kinase Receptor Associate Protein

TOP – Terminal Oligopyrimidine motifs

tRNA – transfer RNA

UTR – Untranslated Regions

UV – Ultraviolet

WH – Winged Helix

1. INTRODUCTION

1.1 RNA Binding Proteins

1.1.1. Overview of RNA binding proteins

The composition of a ribonucleoprotein (RNP) complex involves the participation of at least one RNA binding protein (RBP) which interacts with a molecule of ribonucleic acid (RNA) through an RNA binding domain (RBD)[1] (figure 1.1)[2]. This binding can occur due to two types of motifs in RNA: sequential or structural and RBDs can bind to these motifs by only one or both of them. Some examples of RBDs are the RNA recognition motif (RRM), hnRNP K homology domain (KH) or DEAD box helicase domain.[3]

Some recent breakthroughs of structures belonging to vast RNP machines like the ribosome,[4] or the spliceosome[5] have shown that canonical RBDs of these RNPs are not involved in their interactions with RNA. This is proof that other unknown modes of binding exist, broadening the possibilities that were not expected before these findings.

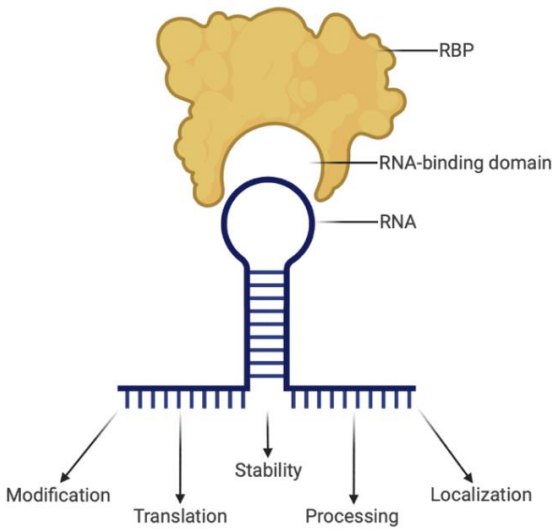

Figure 1.1 Example of an interaction between an RNA-binding protein and RNA, with the RNA outcomes due to RBP regulation influence. Image made in Biorender

The discovery of these RBPs where a canonical RBD is absent or when it is present, is not responsible for RNA-protein interaction, brought another question to the table: it's known that these proteins regulate RNA, but is RNA also capable of regulating RBPs? As mentioned, it is known for some time that RBPs regulate various aspects of RNAs such as its processing, stability, localization, among other activities. They are even referred as "mRNA's clothes" (messenger RNA), since they make sure that specific regions (coding region, 5' and 3' untranslated regions or UTRs) are sheltered or exposed, which eventually aides the mRNA to develop to the subsequent phases of its life.[6] However, research has shown that the contrary may also happen since reports have uncovered the functions of long non-coding RNAs (lncRNAs) like participating in recruiting transcription factors, chromatin-modifying complexes and even organize, scaffold, or inhibit proteins, showing that RNAs can also regulate proteins.[7] These new functions shatter the established convention of only RBPs regulating RNAs (figure 1.2)[2].

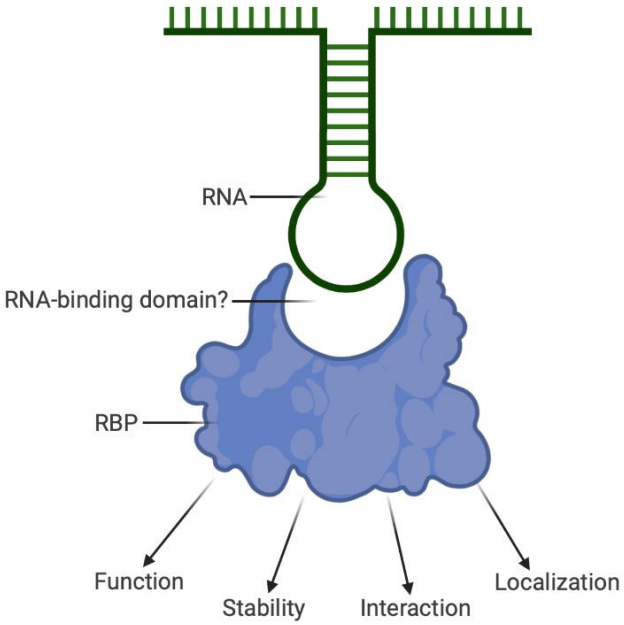

Figure 1.2 Example of an interaction between RNA and an RBP, with RBP outcomes due to RNA regulation. Image made in Biorender

On top of all the advances in RBPs knowledge, one specific discovery has changed the way the scientific community looks at this special group of proteins. It started with the characterization of RNP granules with no membranes that are visible under the microscope like Cajal bodies, stress granules and various others.[8] These granules have shown to be the result of liquid-liquid phase separation propelled by intrinsically disordered regions (IDRs) that belong to the RBPs.[9] IDRs are thought to have impact in the RBPs functions and interactions, specifically the one with RNA. There are three main types of binding in IDRs: one through arginine rich motifs, like RGG and SR repetitions;[10] the other with aromatic residues, such as tyrosine in motifs like [G/S]Y[G/S];[11] and finally a group of heterogeneous and linear motifs containing mostly lysine and, in a lower degree, Arg.[11] IDRs can also promote co-folding

between protein-RNA after interacting with their target RNAs.[12] They may also enable the regulation of RBPs binding to RNA by reversible post transcriptional modifications (PTMs) such as acetylation or phosphorylation.[11] To finalize, IDRs are versatile modules that potentiate RNA-binding by being either highly specific or completely nonselective.

1.1.2. Identification and cataloguing of RNA binding proteins

1.1.2.1. Experimental approaches

The main way to identify RBPs is through experimental approaches. Some of these, are in vitro experiments that use immobilized RNA probes or arrayed proteins and helped identify various new RBPs.[13] More recently, a new in cell technique called RNA interactome capture (RIC) was developed and has been used with great efficiency to identify many new RBPs by focusing on native protein-RNA interactions, since it involves ultraviolet (UV) crosslinking of RBPs to RNA through amino acids in close proximity of nucleoside bases in live cells, succeeded by a collective capture of the RNPs with polyadenylated (poly(A)) RNA on oligo(dT) beads , followed by protein identification being carried out by quantitative mass spectrometry (Q-MS)[14] (figure 1.3)[2]

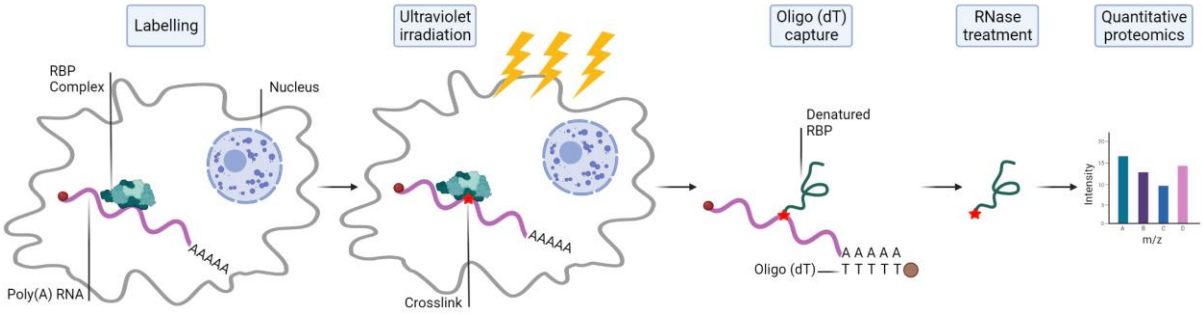

Figure 1.3 Scheme representing the process of identifying RNA binding proteins repertoires by RIC. Image made in BioRender

The implementation of this technique has shed more light onto the fact that there are many non-canonical RBDs, with about half of the proteins in each interactome possessing unknown domains, and hundreds having no previously known interaction with RNA biology. Even more interesting is that studies show that these recently identified proteins possess common biological roles and molecular functions like intermediary metabolism, cell cycle progression, antiviral response, spindle organization and protein metabolism.[15,16]

These important findings help illustrate the importance of the experimental approaches, both in vitro and in vivo, to obtain more knowledge about RNA binding proteins.

1.1.3. Identification of RNA binding domains

The need to map RBDs in proteins is important to clarify how RBPs interact with RNA since a major number of the newly discovered RBPs possess unknown domains. Some reliable methods that use mutagenesis in combination with RNA binding assay, such as electrophoretic mobility shift essay

(EMSA)[17] or the CLIP-coupled PNK assay[18], have been used to map RBDs. However, these are low-throughput and make more sense when trying to discover possible domains in a specific protein. For the identification of RBDs in a large amount of recent RBPs, high-throughput methods are needed. Several techniques were invented, with each one utilizing mass spectrometry in the most variable configurations to identify the protein regions that crosslink to RNA after exposing live cells to UV light radiation. Examples of approaches are the RNP[xl],[19] RBDmap,[11] and proteomic identification of RNA-binding regions (RBR-ID)[20] (figure 1.4)[2].

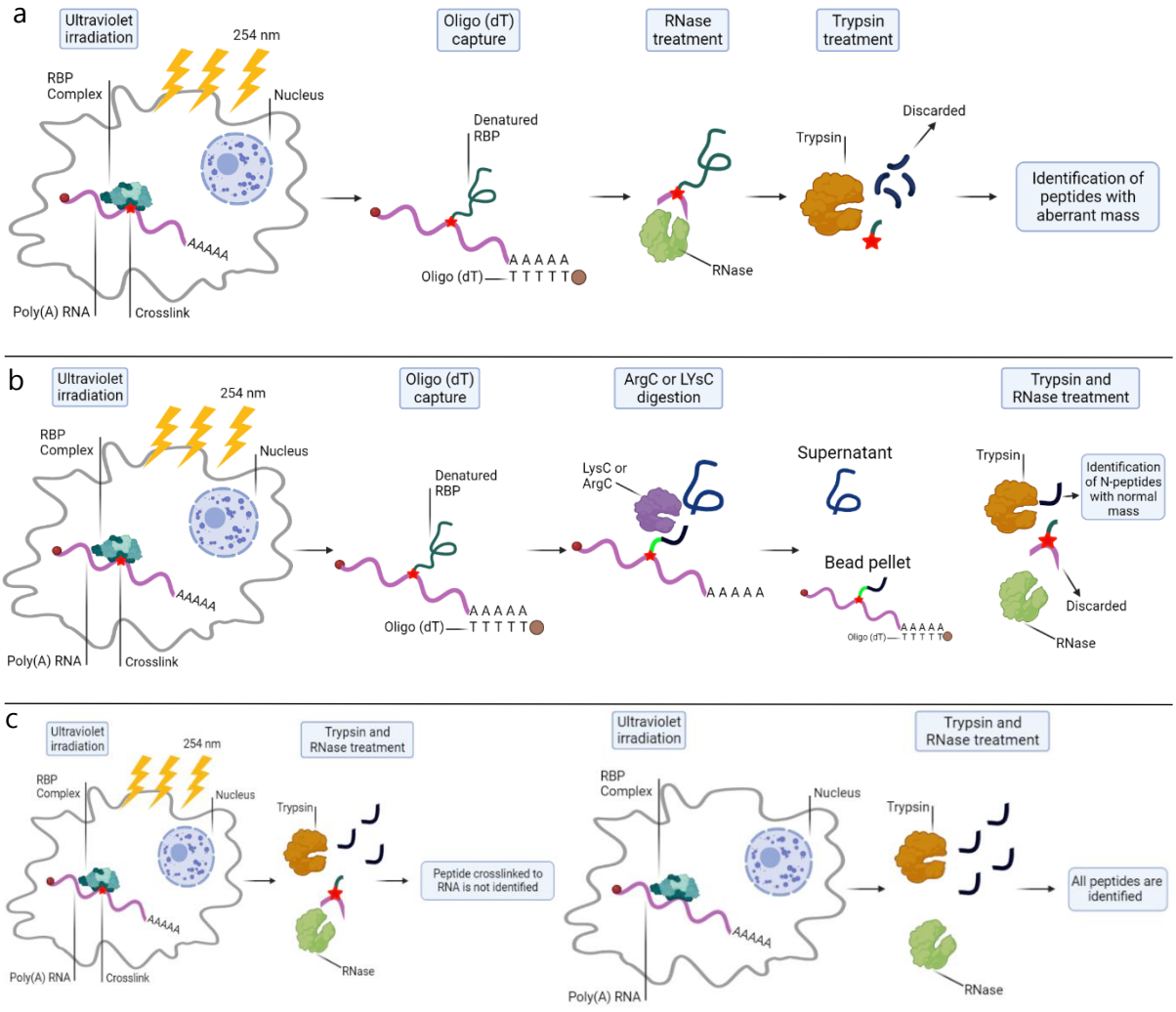

Figure 1.4 Illustration of the methods used to identify RNA binding domains: **(a)** Purification and direct detection of RNA-crosslinked peptides pursued by data analysis with RNP[xl]; **(b)** RBDmap; **(c)** Proteomic identification of RNA binding regions (RBR-ID). Image made in BioRender

Despite that, there is still a lot of information to be known about the RNA targets of these novel RBPs or even about the function of their interactions. If these problems are to be addressed, functional studies, including ones determining the specificity and affinity of these new RBDs for their target sequences, are required.

1.1.4. Modes of RNA binding

There are several types of binding between RNAs and RBPs, with a crescent number of RBDs being discovered in recent years. These can be divided into several groups depending on specificity, other proteins, structure and more.

1.1.4.1. High-specificity RNA binding

This mode of binding is the one mostly connected with canonical RBDs like the RRM. This mechanism is quite specific since RRMs usually interact with 2-8 nucleotides (nt) in single stranded RNA (ssRNA) through sequential stacking interactions. Normally each RRM has preferential sequences. By combining RRMs in a consecutive manner in a RBP, the affinity and specificity tend to increase, because although only one RRM is capable of binding with great specificity, more domains are necessary to define specificity since the number of nucleotides that are recognized by only one RRM is usually too small to define a unique binding sequence.[3] The example shown here in figure 1.5 is of LaRP7 binding to stem-loop 4 of 7SK RNA.[21]

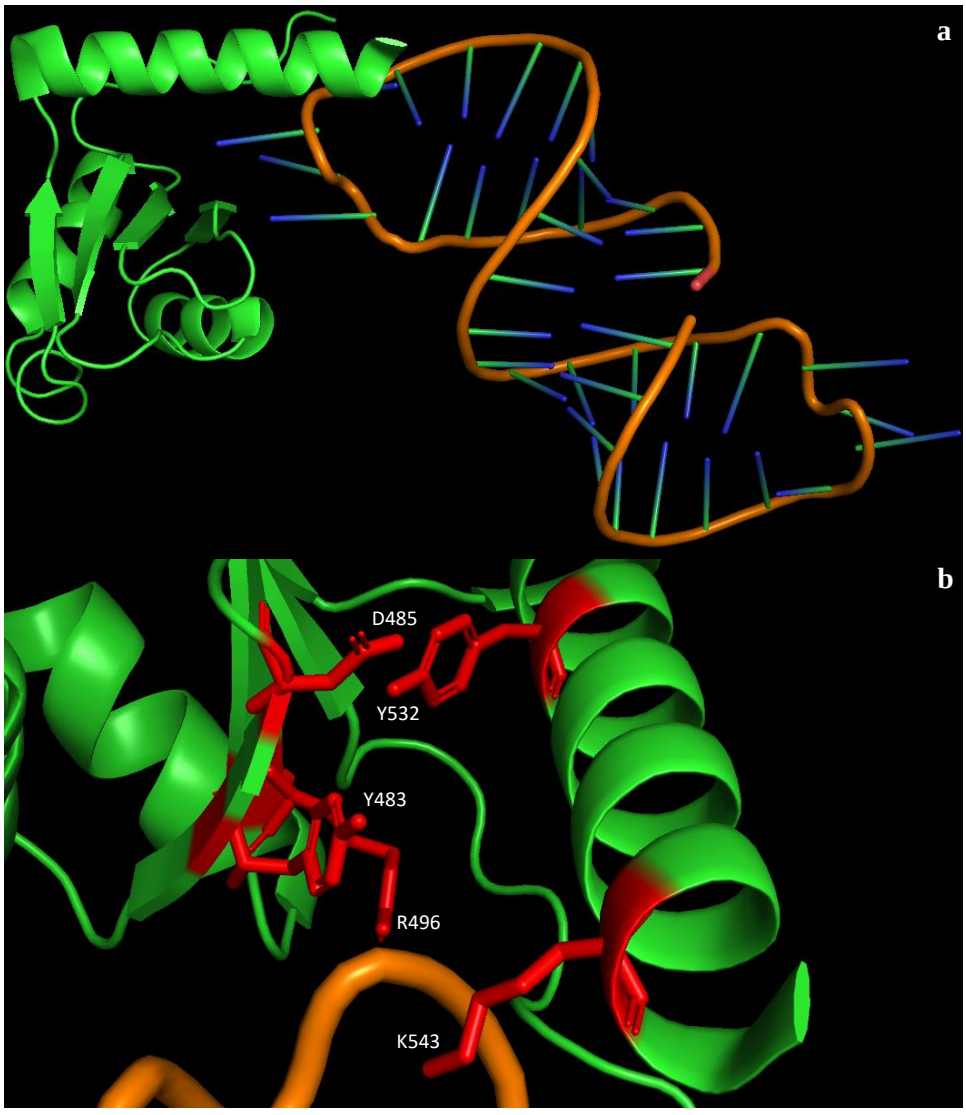

Figure 1.5 Crystal structure of the RRM of LaRP7 binding to the 7SK RNA stem-loop 4 (PDB:6D12) **(a)** Stereo view of the complex **(b)** Close up of the 5 residues in red (Y483, D485, R496, Y532 and K543) responsible for recognition of the 7SK RNA stem-loop 4

1.1.4.2. RNA binding by protein disordered regions

The already mentioned intrinsically disordered regions have some examples of binding interactions already discovered, like the Arg-Gly-Gly (RGG) repeat motif of the fragile X mental retardation protein (FMRP) which co-folds with the target RNA to produce a tight electrostatic and shape-complementation-driven interaction with effects on RNA processing, localization, stability and translation.[22] (figure 1.6)[2].

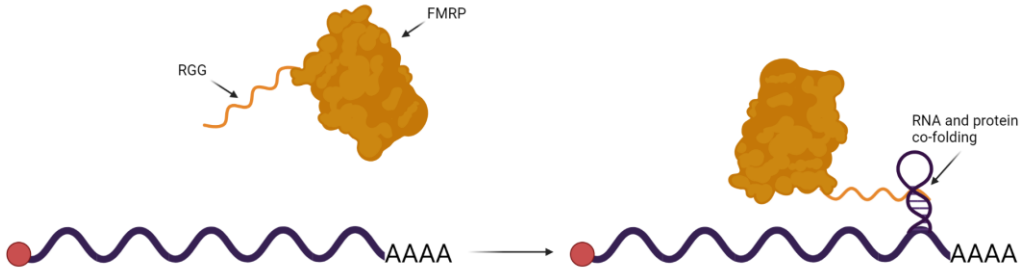

Figure 1.6 Illustration of RNA binding by intrinsically disordered regions with the case of the RGG repeat motif of the FMRP. Image made in BioRender

1.1.4.3. Interactions by complementary shapes

Some proteins do not bind by their sequences or specific regions but for their shapes since they can perfectly complement their target RNAs. An example is the double-stranded RNA-binding domain (dsRBD) from the *Drosophila* protein Staufen (figure 1.7) that possesses five copies of this motif. It recognizes the specific A-shape form of dsRNA by interacting with conserved residues in the loop 2 and the minor grove, and between loop 4 and the phosphodiester backbone in the adjacent major groove. Helix α1 also maintains interaction with the single strand-loop responsible for capping the RNA helix.[23]

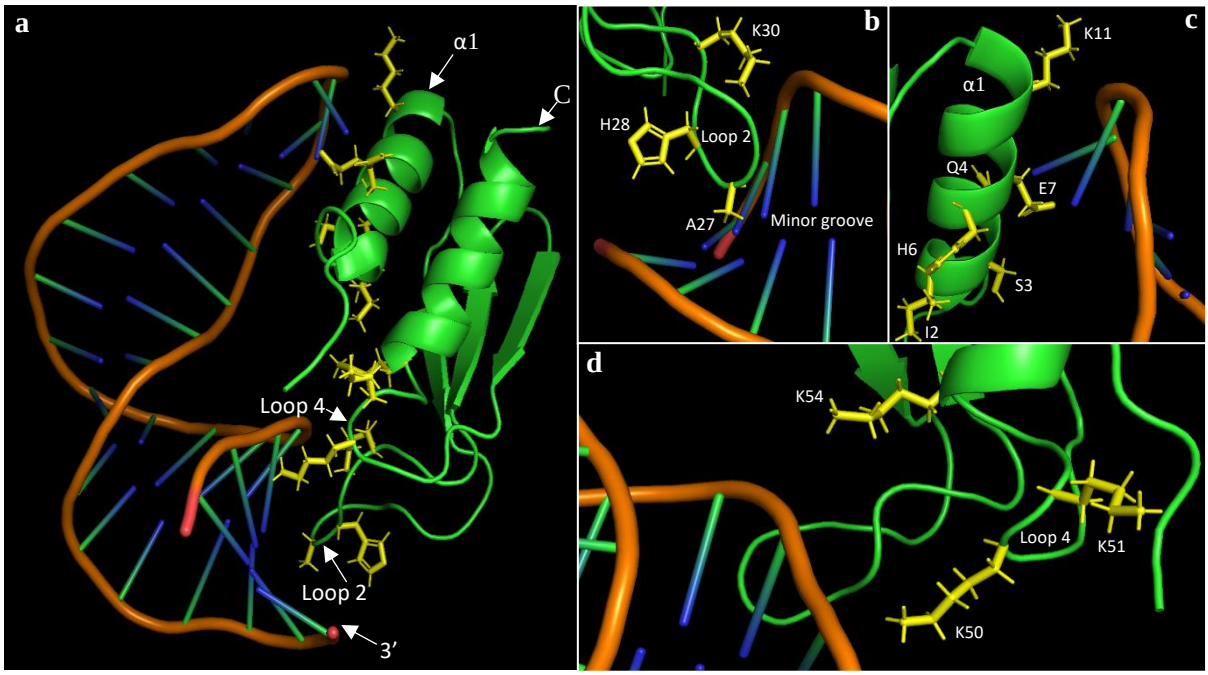

Figure 1.7 NMR structure of the complex between the dsRBD3 from *Drosophila* Staufen protein and a RNA hairpin (PDB: 1EKZ). **(a)** Stereo view of the complex with the residues responsible for interaction with the hairpin highlighted in yellow; **(b)** Close up of residues A27, H28 and K30 (in yellow) from loop 2 that interact with the RNA minor groove; **(c)** Close up of residues I2, S3, Q4, H6, E7 and K11 from helix α1 that bind to the hairpin; **(d)** Close up of the residues K50, K51 and K54 from loop 4 responsible for the interaction with the hairpin major groove. Image made with PyMol

1.1.4.4. Modulation of protein activity

RNA can also bind to proteins by promoting their activation. This is the case for the interferon-induced, double-stranded RNA-activated protein kinase (PKR) when it binds to double-stranded RNA (dsRNA), attained from viral replication, leading to PKR dimerization, phosphorylation and activation (figure 1.8)[2]. The active PKR will phosphorylate eIF2α to interrupt protein synthesis in the infected cells.[24]

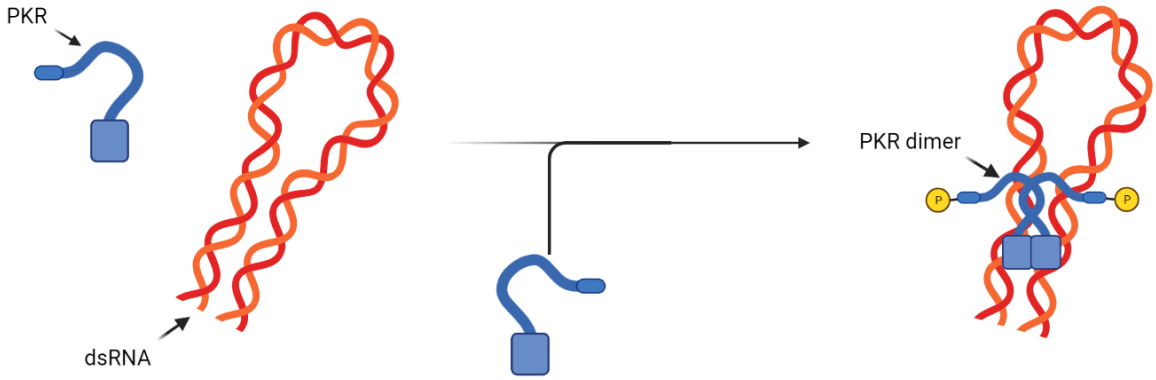

Figure 1.8 Illustration of the modulation of protein activity by RNA binding with PKR as the example. Image made in BioRender

1.1.4.5. Metabolic enzymes moonlighting in RNA binding

Changes in metabolic enzymes can lead to the binding of RNA with the objective of translating specific elements. The iron-regulatory protein 1 (IRP1) is one of these cases since it associates with an iron-sulphur cluster with the objective of catalysing the conversion between citrate and isocitrate. However, when iron levels are down, the iron-sulphur cluster is not synthesized and IRP1 interacts with mRNAs that possess the code for cellular factors involved in iron homeostasis, ending up regulating its fate[25] (figure 1.9)[2].

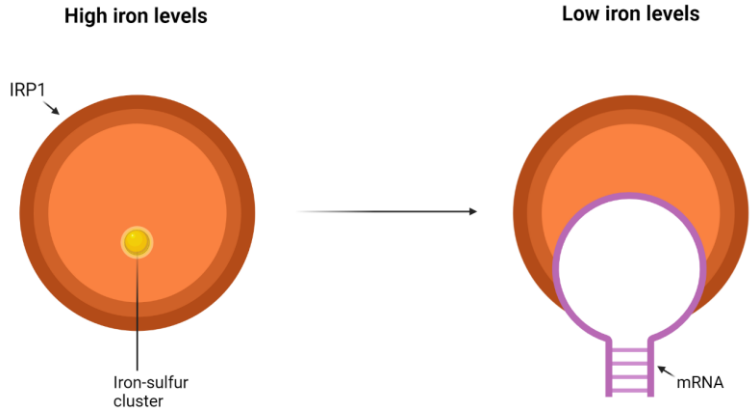

Figure 1.9 Illustration of how changes in metabolic enzymes can force them to bind RNA with the specific case of the IRP1. Image made in BioRender

Despite all the recent finding about numerous RBPs, new domains and various modes of binding, one aspect that still concerns the scientific community is the lack of structures in the Protein Data Bank (PDB) of RNA binding proteins. Still, one superfamily of RBPs has been shifting the focus of the community since it has shown to have various modes of binding, the presence of IDRs, formation of stress granules but the most interesting of all is the interplay between two binding domains that exist in all the constituents of the La-related proteins superfamily.

1.2. The La-Related Proteins Superfamily

1.2.1. La and the La-Related Proteins

La and La-related proteins (LaRPs) constitute an ancient superfamily of proteins that remain conserved in all of Eukarya, except in Plasmodium.[26] Right now, there are 5 families of LaRPs which are: La (aka LaRP3), LaRP1, LaRP4, LaRP6 and LaRP7 (figure 1.10)[27]. All the families possess a conserved bipartite RNA-binding unit called La-module which is quite unique in the world of RBPs. The La-module is constituted by a La-motif (LaM) followed by an RNA-Recognition motif (RRM1). Other than this, each family possesses distinct domains and motifs which contribute to structure and function, with them being specific for their respective families.[28]

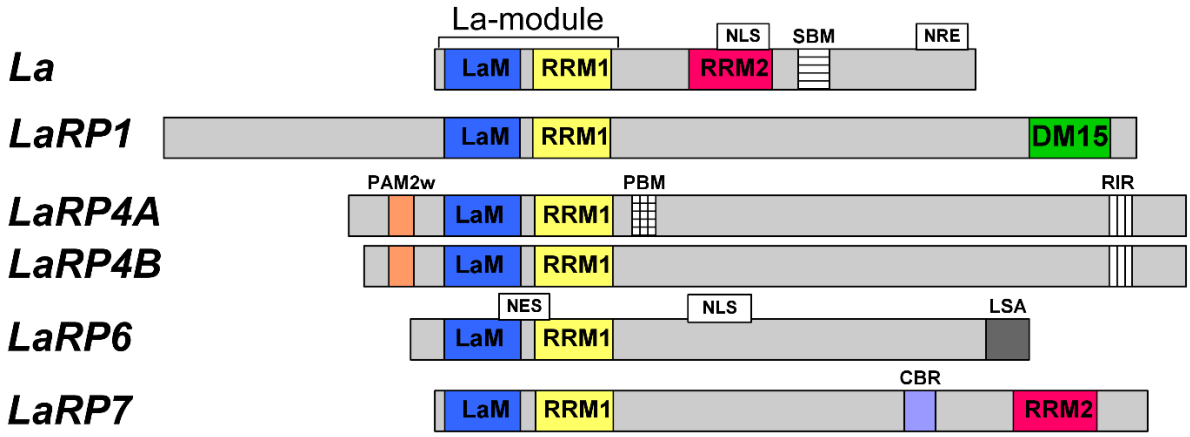

Figure 1.10 Representation of the domains organization across the different LaRP proteins in vertebrates. La motif (LaM, blue), RNA-recognition motif 1 (RRM1, yellow), RNA-recognition motif 2 (RRM2, pink), DM15 domain (green), PAM2w motif (orange), PABP-binding motif (PBM squares), RACK-interacting motif (RIR, black vertical stripes), short basic motif (SBM, thin black horizontal stripes), La- and S1-associated motif (LSA, solid grey), conserved basic region (CBR, purple), nuclear localization signal (NLS, white), nuclear retention element (NRE, white) and nuclear export signal (NES, white). Image made with IBS[29]

To perform their multifaceted and multiple roles in the cell, LaRPs have adapted to bind various RNA ligands and protein partners. The interactions that are best known are for the La protein and LaRP7.

La has functions in two areas of the cell: the nucleus and the cytoplasm. In relation to the nucleus, it is responsible for the association with all newly synthesized RNA polymerase III (pol III) transcripts possessing a short 3' polyuridylate tail. This includes precursors of small nuclear RNAs (snRNAs, like the SRP, 7SK), nucleolar (snoRNAs) and transfer (tRNAs). Its function in pol III-transcribed RNAs biogenesis concerns their processing and maturation, by binding through its La motif.[30] As for tRNAs, La binds the 3'UUU trailer allowing for 3'end maturation to occur through endonucleolytic cleavage by stabilizing the pre-tRNAs in conformations that allow the process. This findings were supported by a mutation in the tRNA of yeast, that perturbs its structure, causing cells to need a homolog of La to process the tRNA, turning La into a chaperone type of protein. [31] This activity also involves a second binding region on La, distinct from the 3'UUUOH interacting surfaces, by binding the RRM1 of La to non-defective tRNAs which affects their processing by increasing the maturation efficiency.[32,33] In higher eukaryotes La can also be found in the cytoplasm. Here it acts as an IRES-transacting factor (ITAF), assisting in the translation of various viral and cellular mRNA.[34,35] There is no 3'-poly(U) trailers in the internal ribosome entry site (IRES), and the interaction of human La (HsLa) with the Hepatitis C Virus (HCV) IRES has shown an intricate interplay of distinct subdomains, since on top of using both the La motif and the RRM1, it also uses the RRM2 which had no previous role in RNA binding. This concert between the three domains is based in the recognition of structural features of the RNA, more objectively a double-stranded stem flanked by single-stranded extensions.[36] Intracellular movement of HsLa across the nucleus, nucleolus and cytoplasm is directed by signalling motifs that are harboured within the C-terminal region, either near or overlapping with a short basic motif

(SBM), which is thought to recognize the 5'-triphosphate end of pre-tRNAs, while also binding and promoting translation driven by IRES.[37,38] The HsLa C-terminal region is also the target for phosphorylation, with a potential role in RNA binding and cellular compartmentalization.[39]

The other LaRP family that has been studied most is LaRP7. It's a mostly nuclear protein with reports of nucleolar localization to a lesser extent. It has some similarities to the La regarding its RNA target, since it also recognizes a terminal UUU_{OH} stretch, but contrary to La, it has a preference to associate with the 7SK snRNA, which is a nuclear non-coding RNA of 331 nt.[40] This specific RNA and the proteins associated with it constitute the 7SK snRNP (small nuclear ribonucleoprotein) complex, that performs the sequestering of the positive transcriptional factor P-TEFb in an inactive condition since the whole complex is composed by the 7SK snRNA which is a scaffold for LaRP7 (responsible for the RNA stability), the P-TEFb inhibitor HEXIM1 that inactivates the P-TEFb and the capping enzyme MePCE (also stabilizes the RNA).[41] The binding to the 3'UUU_{OH} sequence is carried out by the La-module, while the recognition between 7SK RNA and HsLaRP7 is strongly dependent on the RRM2 domain present in the C-terminal half of the protein, by establishing strong connections with a conserved hairpin loop located in the 3' region of the 7SK RNA, namely 7SK-HP4.[42] While both HsLa and HsLaRP7 possess an RRM2, intriguingly the role of this domain in RNA binding and its position in the protein sequence seems to have diverged for the two families, with the RRM2 of LaRP7 being further away from the La module in comparison with the one from La, as seen in figure 1.13.[43]

LaRP1 and LaRP6 are mainly found in the cytoplasm during steady state, which is a state where the cells maintain their ions and molecules at a constant internal level. Here, they interact with mRNAs and take part in translation regulation. Although the binding activities of LaRP1 La-module still need to be clarified, its C-terminal DM15 domain specifically binds the m7GpppC cap of 5' terminal oligopyrimidine (5' TOP) motifs, which is a representative trait of mRNAs that encode for ribosomal proteins and other translation factors. Hence, DM15 may be capable of being a substitute for the cap-binding initiation factor eIF4E, that is a pillar in the central pathway for mRNA recruitment to the ribosome during translation initiation.[44] Moreover, phosphorylation of LaRP1 by mTORC1 seems to decrease mTORC1-dependent induction of translation initiation, although it remains unknown if this has any consequence on the RNA-binding ability of LaRP1. [45]

LaRP6 binds a stem-loop structure comprising of a large internal loop found in the 5' UTR of the collagen α1 and α2 of type I and α1 type III mRNAs.[46] This interaction has been considered as a cornerstone step of the collagen biosynthetic pathway.[47] Further studies identified HsLaRP6 as a regulator of miR-141 and miR-145 biogenesis[48] and of ribosomal protein mRNA (RP-mRNA) localisation and translation.[49] This RP-mRNAs are responsible for the translation of ribosomal proteins that are necessary for ribosome biogenesis, which has been a common hallmark as a driver for cancers when in an hyperactive state. As such, LaRP6 has been shown to be upregulated in breast carcinomas possibly indicating that its inhibition could be potentially used in therapeutic strategies.

An RNA-SELEX analysis even uncovered an intricate binding specificity for HsLaRP6 going from linear motifs, multiple dimeric motifs and internal loops comprised in a double-stranded RNA stem.[50] Some proteins that interact with HsLaRP6 include non-muscle myosin, RNA helicase A (RHA) and STRAP (Serine/Threonine kinase Receptor Associate Protein), the latter recruited via a short, preserved motif at the C-terminus known as LSA (LaM and S1 Associated).[47,51] Unfortunately, the role of these protein associations in biology regarding LaRP6, has yet to be discovered.

However, the family of LaRPs that was the focus of this work is LARP4, that contains LaRP4A and LaRP4B members.

1.2.2. LaRP4A and LaRP4B

1.2.2.1. General overview

LaRP4A and LaRP4B proteins of vertebrates are the evolution product from a single LaRP4 protein from invertebrates.[26,52] Although they can also be named LaRP4 and LaRP5, respectively, the former way is more utilized since it reflects their evolutionary link. Both promote mRNA stability and improve translation, through interactions with the 3' UTR of mRNAs, the Poly(A) binding protein (PABP), RACK1 (receptor for activated protein kinase C 1) and polysomes.[53,54]

LaRP4A plays an important role in poly(A) lengthening of mRNAs and interacts with poly(A) sequences starting at 15 nt.[53,55] It has been reported through a structural analysis, that the N-terminal region (NTR) has an affinity for this RNA target.[56] This region also contains a variant PAM2w motif that works as the binding platform for the PABP MLLE domain, which is highly conserved in eukaryotes and that binds to PAM2 sequence, a motif found in many proteins that bind PABPs with a very important phenylalanine making direct contact with the MLLE but that is substituted by a tryptophan in LaRP4. The study showed that both the PAM2w motif and the PABP play a role in poly(A) recognition, standing at the crossroads of protein-protein and protein-RNA interactions.[53,56] The NTR itself is in a semi-disordered state with no canonical or known RNA-binding domain, which correlates with the recent findings among RBPs that contain RNA binding domains that are in fact intrinsically disordered regions. LaRP4A has also been identified as a regulator for microRNA mir-210 biogenesis.[57] Regarding cancer, it appears that LaRP4A has a role since it's been shown to control cancer cell motility and morphology: gene depletion raises cell migration and invasion in prostate and breast cancer cells, while overexpression diminishes cell elongation and promotes cell circularity.[58] All these findings make the LaRP4A members an interesting object of investigations.

For LaRP4B not much is known since its structure has not been resolved nor its domains structures. As for its binding partners, they are known but the modes of interaction have yet to be explained and even the mechanism by which it interacts with RNA is still a mystery. As is, its preferred RNA targets are a subset of mRNAs containing AU-rich elements at the 3' UTR.[54] Finally, it has also shown influence in cancer both in a positive and negative manner: a genetic screen in mice and human glioma

cells has shown that it can act as a tumour suppressor[59]; on the other hand, a study has identified through LaRP4B knockdown a reduction in leukaemia stem cells on acute myeloid leukaemia (AML)[60], while Li et al has reported the ability to use LaRP4B as a potential biomarker for liver cancer diagnosis and prognosis.[61]

1.2.2.2. The La motif

The LaM or La motif is a elaboration of the winged helix (WH) domain that is usually encountered in transcription factors.[62] In LaRPs it is usually composed by six α-helices that are arranged around a three strands antiparallel β-sheet. Three of these helices (α1', α2 and α4) represent a bolt-on architecture that is a feature to the canonical fold.[63,64] The longer helix α1 and two smaller ones (α1' and α2) are followed by the first strand (β1). The other 3 helices (α3, α4 and α5) can be found between strands β1 and β2 forming a 'C'-like shape. Contrary to other WH domains, the LaM possesses a hydrophobic pocket of helical assembly (α1', α2, α3 and α4) against the apolar side of α1, generating this way the central RNA interaction location. The twisted β-sheet has a solvent-exposed surface on the side opposite to α1.[63] As stated by its name of being a winged helix domain, the LaM contains two wings in its structure: Wing 1 (W1) is 5 to 6 residues long and has the role of linking β2 and β3; Wing 2 or W2 interacts with the base of helix α1', being considered the structured loop at the C-terminus of the domain.[63,65]

This domain is the most conserved region in the LaRP superfamily and by comparing its 3D structure from HsLa and HsLaRP4A, a high degree of similarity was observed. The most notable change between these two proteins is the lack of W2 in LaRP4A[56], showing that even at the LaM level, which is supposed to be the conserved element of this superfamily, important changes may occur. Regarding conserved amino acids, the ones directly related to function in HsLa (Q20, Y23, Y24, D33, F35 and F55) remain, except for two: the hydrophobic Y24 and F55 are replaced by sulphur containing residues C130 and M160, which lead to a change in the cavity's properties.[66] In HsLa these residues are part of the concave surface of the La motif hydrophobic pocket and are responsible for the interaction with the poly(U) sequence of RNA polymerase transcripts. [67] However HsLaRP4A does not recognise these type of RNA sequences, in spite of possessing a similar structural pattern to the one seen in HsLa.[56] Figure 1.11 is the structure of the LaM with the secondary structure elements highlighted.[56]

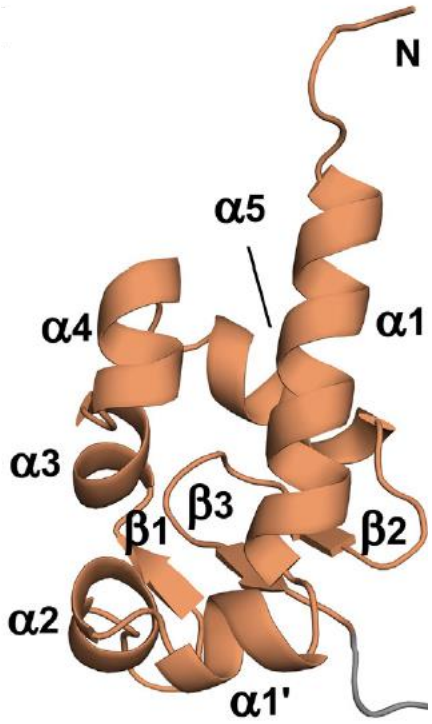

Figure 1.11 Representative structure of LaRP4A LaM (orange)

1.2.2.3. The RRM1

The RNA-Recognition Motif (RRM) is constructed on a four-stranded β-sheet scaffold surrounded by two alpha helices in a β1-α1-β2-β3-α2-β4 topology with five loops (L1-L5). The strands are disposed in a β4-β1-β3-B2 antiparallel sheet, with a packing of helices α1 and α2 on the same side of the sheet.[68] For LaRPs there are two types of RRM: the RRM1 that has evolved with the LaM, and a downstream RRM2, located in the C-terminal region of La and LaRP7 families. Contrary to the LaM, the RRM1 is much more diverse across the LaRPs.[26]

An analysis of the predicted number of β-strands in the β-sheet and the length of loops (L) between secondary structure elements displayed that both HsLa and HsLaRP4A are built around a four-stranded β-sheet, however human La possesses two extra helices: one that precedes the RRM1 core (αN) and a third one which is found at the C-terminal.[63] The HsLaRP4A structure obtained by Nuclear Magnetic Resonance (NMR) showed that the RRM1 does not possess the mentioned helices but its α2 is longer and loop L5, that connects α2 and β4 is elongated (figure 1.12).[56]

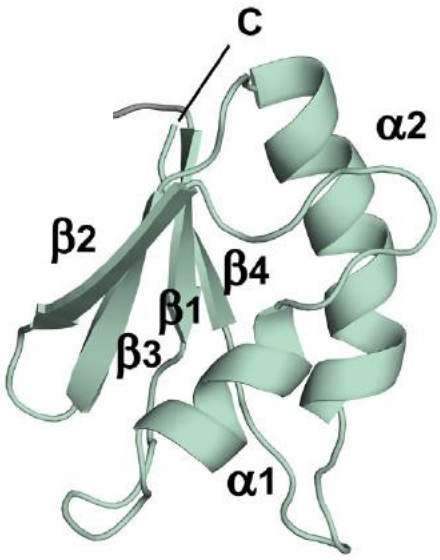

Figure 1.12 Representative structure of the LaRP4A RRM1 (green)

1.2.2.4. The La-module

The La-module is constituted by the LaM and RRM1 being tethered by a short interdomain linker making it a modular structure.[26] It was first observed in human La protein and is a unique RNA-binding unit where the LaM and RRM1 function in synergy to interact with RNA, with both domains folding independently. These properties are shared with LaRP7, human LaRP6, fish LaRP6 and plant LaRP6.[52] Human La protein structure in complex with short poly-uridine (poly(U)) stretches exhibited that the La module (LaM and RRM1) invented its own way of binding RNAs which makes the mechanisms observed in WH domains not viable to explain how LaRPs interact with RNA.[67,69] This is because the LaRPs no longer involve the wings and the recognition helix, which corresponds to helix α5 of LaM,[70] or the canonical surface of the β-sheet in RRM1 in its binding.[68] In fact the process of binding the 3'-terminal uridines by HsLa relies in both LaM and RRM1, with the interdomain linker helping to achieve the proper position for the domains. The target RNA is located in a cleft of the La motif being edged by helices α1, α1' and α2 and lined with the Q20, Y23, Y24, D33, F35 and F55 residues, which are the conserved ones mentioned before. The base U_{-2} is spread out against the RRM1 while U_{-3} stacks on top of U-1 which is stacked on F35. The amino acids located on α1 (Q20, Y23, Y24 in human La), α1-α2 loop (D33), α2 (F35) and α3 (F55) are arranged so that the residues with aromatic side chains form π-interactions with the RNA bases whilst other amino acids dock the RNA with hydrogen bonds and electrostatic connections.[67,71] D33 has the specific role of binding at the same time the 2' and 3' hydroxyls of the last uridine, which is an interaction even more stabilized by N29 and by a dipolar interaction that the N-terminus of helix α2.[52]

Even with most of the contacts between the HsLa La-module and RNA occurring in the LaM, they are not enough for a binding that is considered stable. This is even confirmed by available biochemical and biophysical studies showing that the association of the LaM alone to RNA is too weak to be measured. This means the RRM1 has a crucial contribution to the recognition of RNA by engaging

main-chain atoms at the border of the β-sheet, more specifically strand β2, to pin down uridine U_{-2} in its place, in the deepest cavity of the RNA-binding pocket. The specific recognition in human La is carried out by H-bonds between the O2 and O4 atoms of the U_{-2} with Q20 side-chain amide (LaM) and backbone amide of I140 (RRM1). The recognition of U_{-2} has the role of pulling the LaM and the RRM1 of human La around the RNA, producing a structure with a V-shape, which leads to the formation of a binding crevice that completely protects the 3' end of the RNA.[67]

The La-module of HsLa exhibits a predominant amount of positive charges and an identical electrostatic surface potential in and around the cavity where the terminal uridines are found, including the area in the La motif nearing the N-terminal end of the helix α1 and the interdomain linker part beneath the binding crevice.[67]

As mentioned above, an intriguing discovery was how human LaRP4A (figure 1.13)[56] diverges from what seemed to be a canonical mode of binding by also using its NTR for this interaction. Studies with results from mutagenesis and NMR chemical shift perturbation reveal that the RRM1 participates in the recognition of the 15 nt oligoA together with sequences in the N-terminal, while the LaM is borderline to the mechanism, diverging from other LaRPs in regard to RNA binding mechanisms. This is, then, the first example reported of a La-module that doesn't engage as an individual entity in RNA recognition. NMR perturbation analysis shows some potential RNA interacting surfaces in the helix α2 and the central β-sheet of human LaRP4A RRM1. Additional RNA data has also uncovered that the LaM and the RRM1 do not adopt a fixed orientation.[56,72] Despite that, independent of the existence of a tandem domain configuration, the HsLaRP4A La-module displays a positive strong character at the interface between the LaM and RRM1, that belongs to residues with a positive charge from the β3 strand of the LaM and from the linker (K197, R198). A superposition of the La-modules from both the HsLa and HsLaRP4A over the LaM demonstrates that the positive 3'UUU_{OH} RNA-binding pocket of HsLa superimposes with a strong electronegative area on HsLaRP4A LaM. Two conserved glutamates (E161 and E162, hLa numbering) provide to this rare feature. Interestingly, LaRP4B has these two residues substituted by an aspartate and a histidine, respectively. This variation in electrostatic potential of the hydrophobic pocket of the LaM may have an influence in this unusual behaviour of the HsLaRP4A La-module towards RNA. Figure 1.16 displays the LaRP4A La-module structure.[56]

To conclude, the difference in the RNA target of both LaRP4A and LaRP4B families may be related to a structural difference and that is why it is important to further expand the knowledge on LaRP4B.

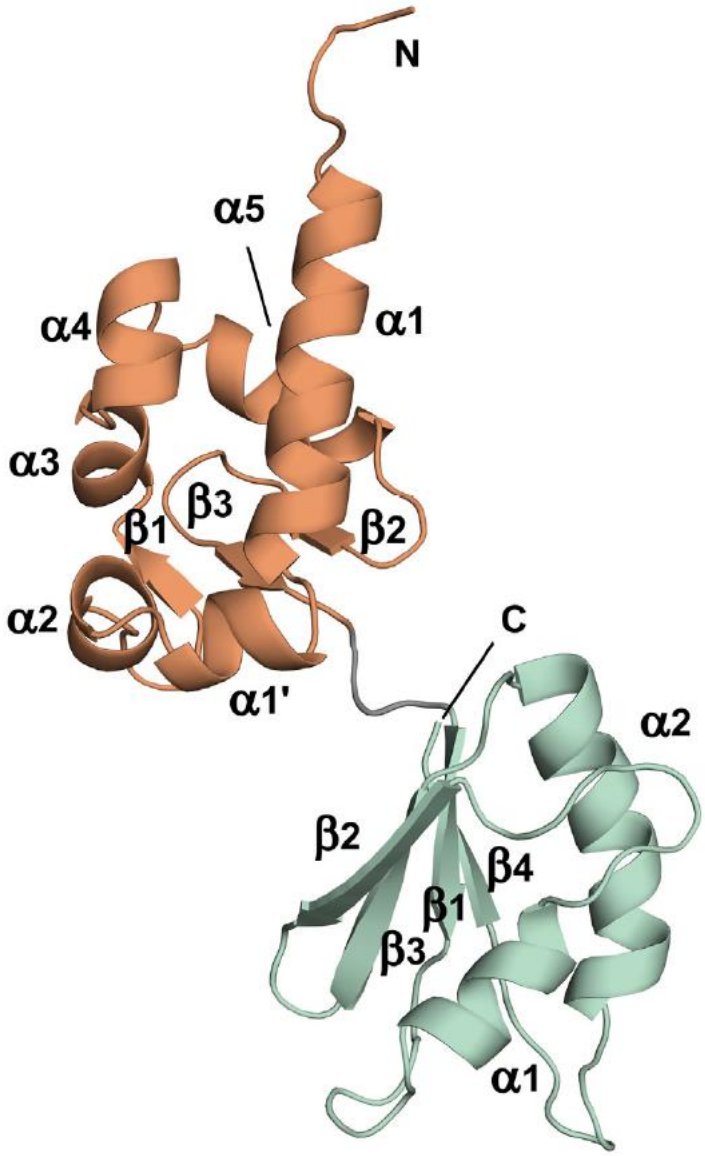

Figure 1.13 Representative structure of the LaRP4A La-module with the LaM in orange, the RRM1 in green and the interdomain linker in grey

1.2.2.5. Interdomain Linker

The interdomain linker of La-modules starts after the end of the LaM, which for the HsLa is the PLP signature of wing 2. These residues have been identified as the last ones that are structurally integrate within the LaM, establishing stable contacts with the rest of the domain. Accordingly, the linker starts at amino acid E99 and consists of two residues (^{99}EV^{100}) attached to the sequence ^{101}TDEYKNDVKNR^{111}.[52]

The dynamic interaction of various elements (LaM, RRM1 and linker) is still an important object of study since it remains to be characterized. For human La, in absence of RNA ligand, the LaM and RRM1 are structurally and dynamically independent of each other, not showing a rigid tandem 3D orientation. However, when complexed with the 3'-UUU_{OH} RNA, the complex is now in a compact V-shaped state, in which the domains plus the linker reorient all together to make contacts with one another

and with the nucleotides. The established domain arrangement maintained by both the LaM and RRM1 in a complex with UUU_{OH} is assisted by the induced ordering of the interdomain linker upon RNA binding but also by some new LaM-RRM1 contacts, like a hydrogen bond between the side chains of Y23 and N139 and a salt-bridge established by R57-D125. [67,73] Interestingly, in a substitution of HsLaRP6 2 residue linker for the eleven residue from HsLa, that the interdomain linker being longer had a negative impact of 500-fold in the interaction of HsLaRP6 with its RNA target proving the importance of the linker to each LaRP.[65]

In LaRP4A, the La-module finds itself in a more elongated architecture in comparison to other LaRPs, which is probably the result of the missing W2 in the LaM and of the short three residue long interdomain linker (^{197}HKR^{199}). This can add to the other possible causes of why the La-module lacks high RNA-binding competence.[56,72]

1.2.2.6. Partners in RNP formation

Both LaRP4A and LaRP4B associate with the poly(A)-binding protein (PABP) in a bipartite way through the variant PAM2w motif localized in the N-terminal region and a PBM (PABP-binding motif) localized downstream of RRM1.[53,74]

The PAM2w motif is unusual, since it accommodates a tryptophan (W) in a position that is normally occupied by a phenylalanine in the conserved PAM2 sequences. Despite this change, HsLaRP4A and HsLaRP4B PAM2w interact with the PABP MLLE domain in a canonical way. The only discrepancy is found in the exit course of the peptide chain following either the F (PAM2) or W (PAM2w), with its significance still to be unveiled.[53,56,74–76]

The cellular localization and functional activites of LaRPs, as well as their binding activities, have all been reported to be controlled by post-translational modifications (PTMs).[39,77] However, the structural ramifications of PTMs on LaRPs and the mechanism by which they influence functional is still largely unexplored.

1.3. Objectives

As demonstrated by the previous chapters, LaRP4B is an important protein to focus on due to the lack of structural knowledge and the fact that even with so many similarities with LaRP4A, their RNA partners are quite different.

Based on this, the main objectives of this dissertation were:

- Express, produce and purify the various LaRP4B domains (La-module, LaM and RRM1) in order to perform structural and biophysical studies to identify their structures and characterize their interaction with mRNA. These studies will be mostly made using Nuclear Magnetic Resonance (NMR).
- Compare the obtained data of LaRP4B with the already published work on LaRP4A to identify possible differences in structure between these proteins.

2. MATERIALS AND METHODS

2.1. DNA purification

All DNA was obtained from Escherichia coli XL-10 cells. Purification was executed using Monarch® Plasmid Miniprep Kit and its respective protocol.[78]

2.2. Protein expression and purification

All proteins were expressed in Escherichia coli Rosetta II cells grown in LB medium. ^{15}N-labelled samples for NMR were grown on minimal media supplemented with 1g/L of $^{15}NH_4Cl$. All cultures were induced with 1 mM of IPTG (isopropyl β-D-1-thiogalactopyranoside) at: OD_{600} 0.6 and left growing at 18 ºC (LaRP4B La Module) and at 37 ºC (LaRP4B RRM) overnight; OD_{600} 0.8 and left growing at 37 ºC (^{15}N-LaRP4B RRM). Cells were harvested and lysed by sonication in a buffer containing 50 mM Tris, pH 8, 300 mM NaCl, 10 mM imidazole, 5% glycerol, protease inhibitor cocktail (Complete Tablets, Roche), 2 mM phenylmethylsulfonyl fluoride (PMSF) and lysozyme. All the proteins were His-tagged and so were purified on a 5 mL His-Trap (GE Healthcare) affinity column with gradients varying from 0 to 300 mM of imidazole. All but two of the purified samples were incubated with His-tagged TEV protease at 4 ºC overnight, whilst they were dialysed in a buffer containing 50 mM Tris pH 7.25, 100 mM KCl, 0.2 mM ethylenediaminetetraacetic acid (EDTA) and 1 mM Dithiothreitol (DTT). The samples were then loaded onto a manual Nickel affinity column to separate the proteins from the cleaved tags, the protease and non-digested products. The proteins were further purified to eliminate any nucleic acid contamination on 5 mL Hi-Trap Heparin (RRM) or DEAE columns (La Module) (GE Healthcare) with gradients of 0 to 1 M KCl. One batch of ^{14}N LaRP4B RRM was incubated with 3x higher TEV concentration in a manual Nickel Affinity column overnight (LaPR4B RRM) with the elution occurring the next day to test a new His-tag cleavage protocol. The other batch was of ^{15}N LaRP4B RRM and didn't go through the cleavage and manual Nickel affinity column process (^{15}N-LaRP4B RRM).

Protein concentrations were calculated using the near-ultraviolet (UV) absorption at 280 nm using theoretical extinction coefficients derived from ExPASY.[79]

2.3. SDS-PAGE

Sodium dodecyl sulphate–polyacrylamide gel electrophoresis was used to confirm the presence of the target proteins during the different stages. For LaRP4B La module the gels used were all with 12% gel percentage while for the RRM 14% gels were utilized.

2.4. Protein sequence alignment

Sequence alignments were performed using Clustal Omega EMBL-EBI portal.[80]

2.5. NMR spectroscopy

The 15N-labelled samples of LaRP4B RRM (234-328) were concentrated to 400-600 μM in a buffer containing 20 mM Tris, pH 7.25, 100 mM KCl, 0.2 mM EDTA and 1 mM DTT.

All the spectra were processed with Topspin 3.5pl7 software (Bruker).[81] All the processing of this experiments was performed previously to my arrival to the lab.

2.5.1. Backbone assignment

The backbone assignment and the analysis of the resonances were performed with CcpNMR Analysis[82] by using triple resonance experiments: HNCACB, HN(CO)CACB, HNCA and HN(CO)CA. The peak list was picked with a HSQC spectrum. All these experiments are part of the previously obtained ones by other lab members.

3. RESULTS AND DISCUSSION

3.1. LaRP4B La Module and RRM Plasmid Purification

With the objective of obtaining more plasmid for future expressions, a Miniprep kit was used after an expression with XL10-Gold cells. The results are shown in table 3.1.

Table 3.1 La-module and RRM plasmids concentration and ratios after purification with the Miniprep kit

Plasmids / Concentration and Ratios	La Module	RRM
Concentration (ng/μL)	187	166
260/230 Ratio	2.13	2.20
260/280 Ratio	1.84	1.85

According to the reported values for nucleic acid purity that are present in table A.1 of the appendix, the purified plasmids are within the stated values.[83]

3.2. LaRP4B La Module

The LaRP4B La Module (151-328) is a 179 amino acids domain with a molecular weight of 20587.5 Da (20.6 kDa), an isoelectric point of 5.0 and an extinction coefficient of 12950 M^{-1} cm^{-1}. With the tag, the molecular weight is of 23370.36 Da (23.4 kDa).

3.2.1. Expression tests

LaRP4B La Module expression was made in LB medium with Escherichia coli cells. The expression was tested with two different post OD_{600}=0.4 (optical density) temperatures: 18 and 37 ºC. The results are displayed in figure 3.1.

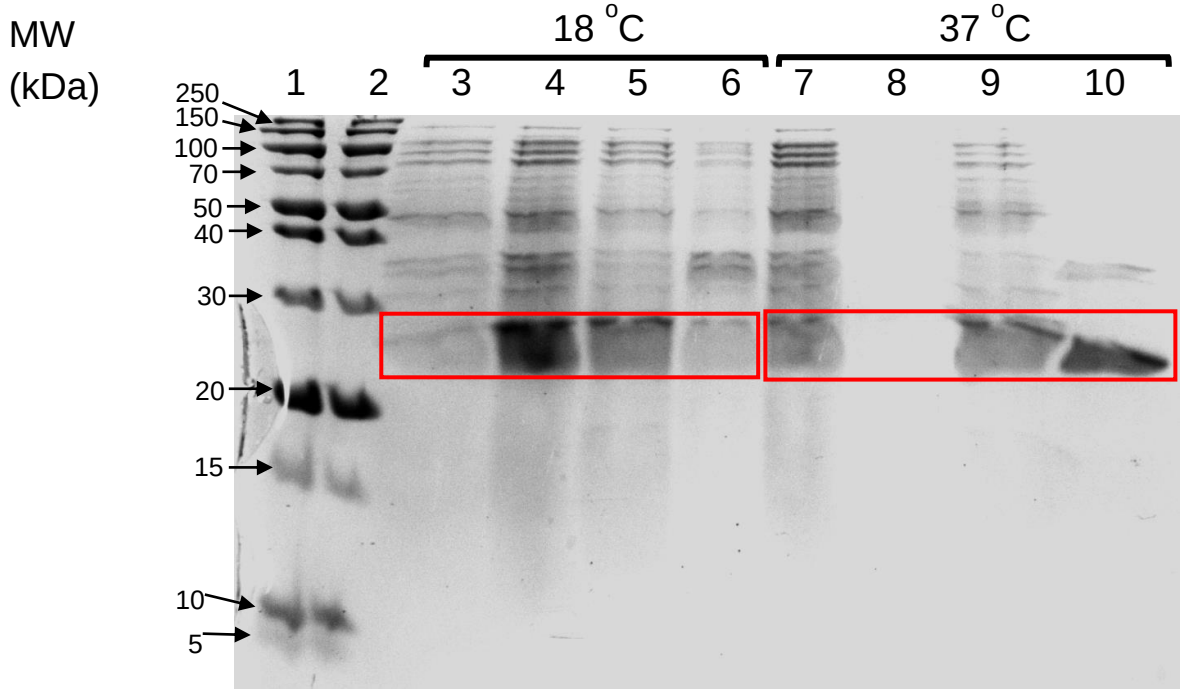

Figure 3.1 SDS-PAGE of the La Module expression tests. Wells: 1,2- Ladder; 3,7- Before induction; 4,8- Total fraction; 5,9- Soluble fraction; 6,10- Insoluble fraction

As seen in the figure above the 18 ºC temperature seems to provide a larger yield in comparison to the 37 ºC where most of the protein was in the insoluble form., There is no evidence of total protein at 37 ºC, However this was probably due to bad laboratory practice and a protein band should have been present since there is a band for both the soluble and insoluble forms.

3.2.2. Protein purification

The protein obtained from the large-scale expression was purified via the protocol detailed in the previous chapter. The first step involved was an affinity chromatography with a His-Trap column. The chromatogram in figure 3.2 (a) and (b) was obtained from this step. Figure 3.3 is the respective SDS-PAGE gel.

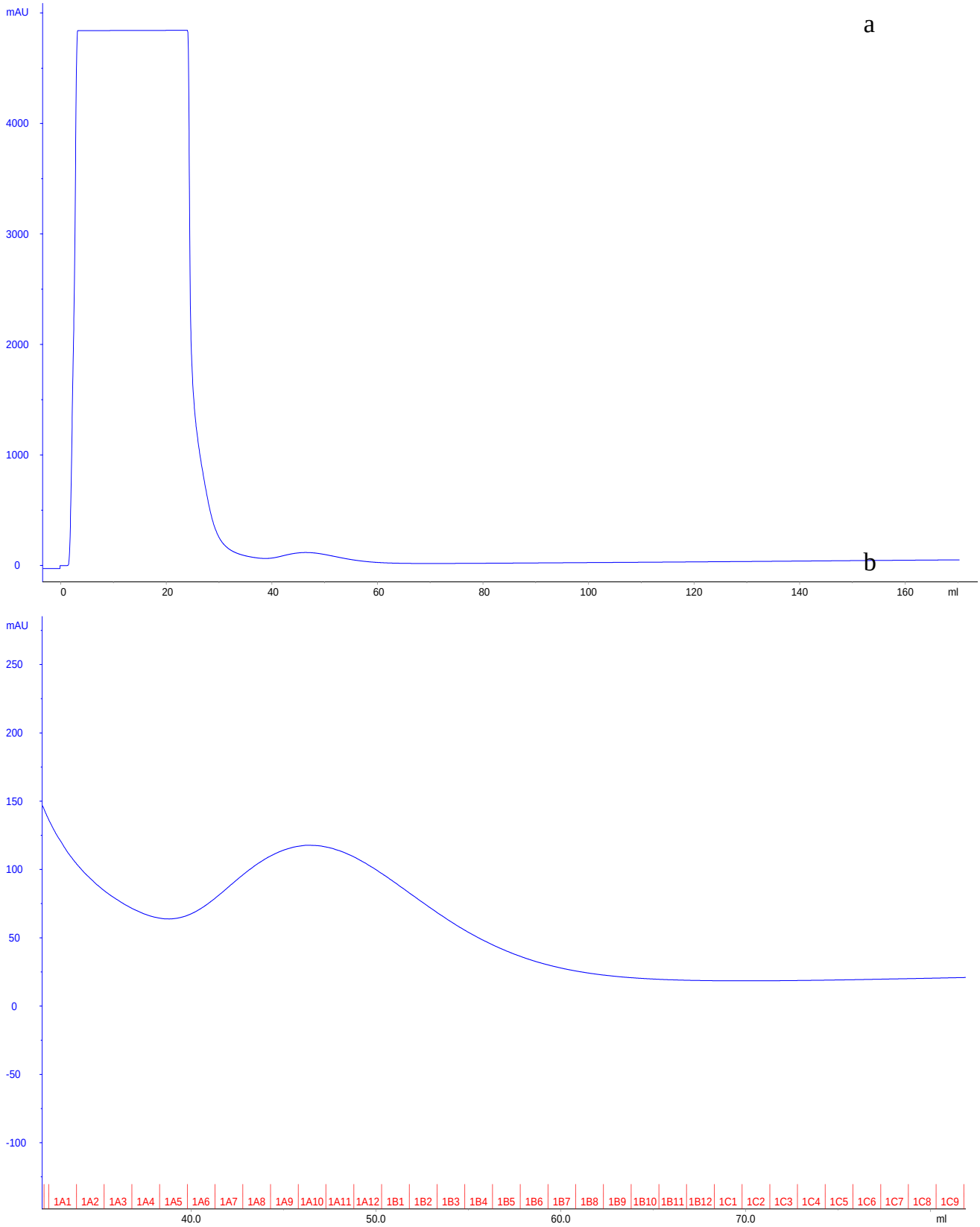

Figure 3.2 (a) Chromatogram of the His-Trap column affinity chromatography purification of the La Module; **(b)** Zoomed in peak of the eluted La-module

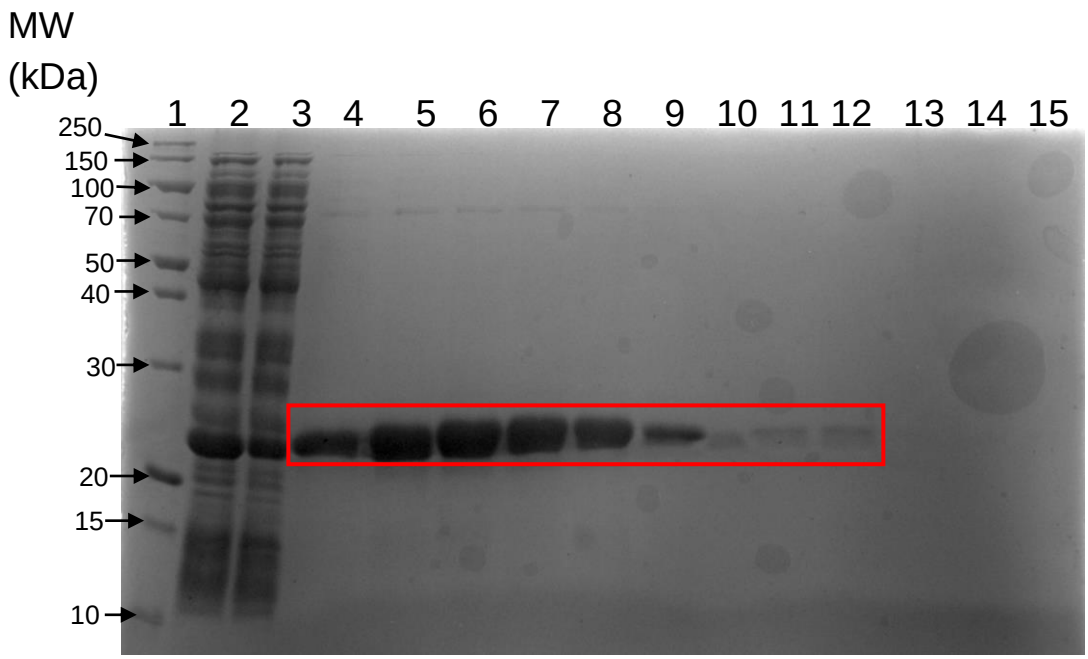

Figure 3.3 SDS-PAGE of the La Module purification by affinity chromatography with a His-Trap column. Label: 1- Ladder; 2- Input; 3- Flowthrough; Fractions 4-A5, 5-A8, 6-A11, 7-B1, 8-B3, 9-B6, 10-B9, 11-B11, 12-C1, 13-C5, 14-C7 and 15-C9

As demonstrated by the chromatogram above, the chromatography appeared to have a good separation along several fractions (A5 to B2). This was partly confirmed by the gel since there was still some protein in the flowthrough and in fractions A5 to C1 which is more fractions than what the chromatogram showed. Since there was still a lot of protein in A5, the fractions from A3 to C1 were collected. The gel also showed the existence of some protein bands, between 70 and 150 kDa.

The next purification step involved once again a His-Trap column to separate the cleaved tag, protease and uncleaved protein from our target protein. Figure 3.4 displays the corresponding SDS-PAGE gel.

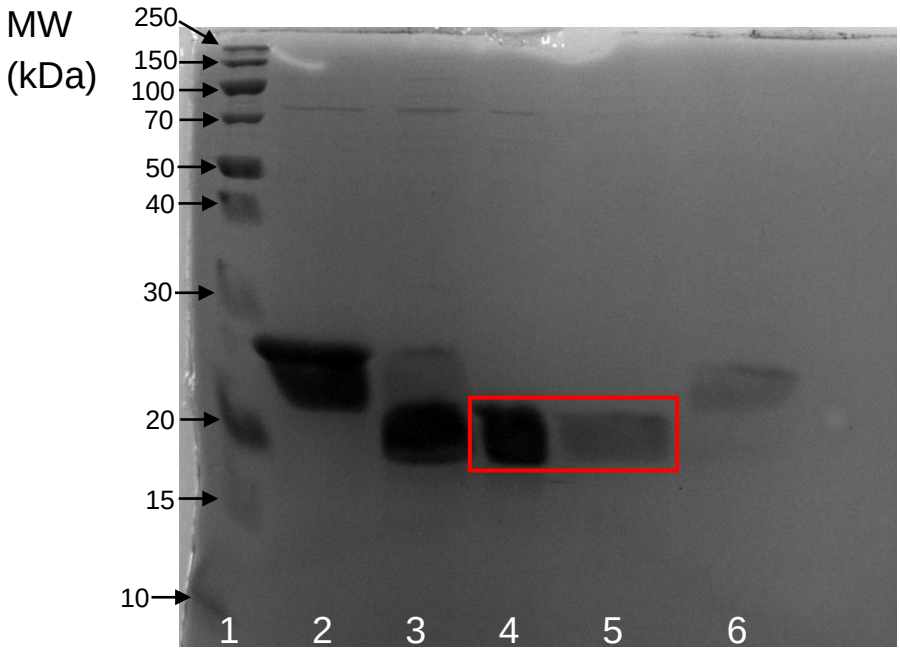

Figure 3.4 SDS-PAGE of the La Module purification with a manual His-Trap column. Label: 1- Ladder; 2-Pre-TEV; 3- After-TEV; 4- Flowthrough; 5- Wash with A; 6- Elution with B

As expected, most of our protein eluted in the flowthrough (lane 4) and a considerable amount in the wash (lane 5). This is because the protein no longer has an His-tag and does not bind to the Nickel in the column, which leads to an elution in the flowthrough, while the uncleaved protein elutes during elution with buffer B as seen in lane 6, since it still binds to the column. Note to the fact that there were still some contaminants between 70 and 150 kDa.

The last purification step was an ion-exchange chromatography with a DEAE column which is a positively charged resin with the objective of binding nuclei acids, which are negatively charged, to it. Figures 3.5 and 3.6 are the chromatogram and the SDS-PAGE gel, respectively.

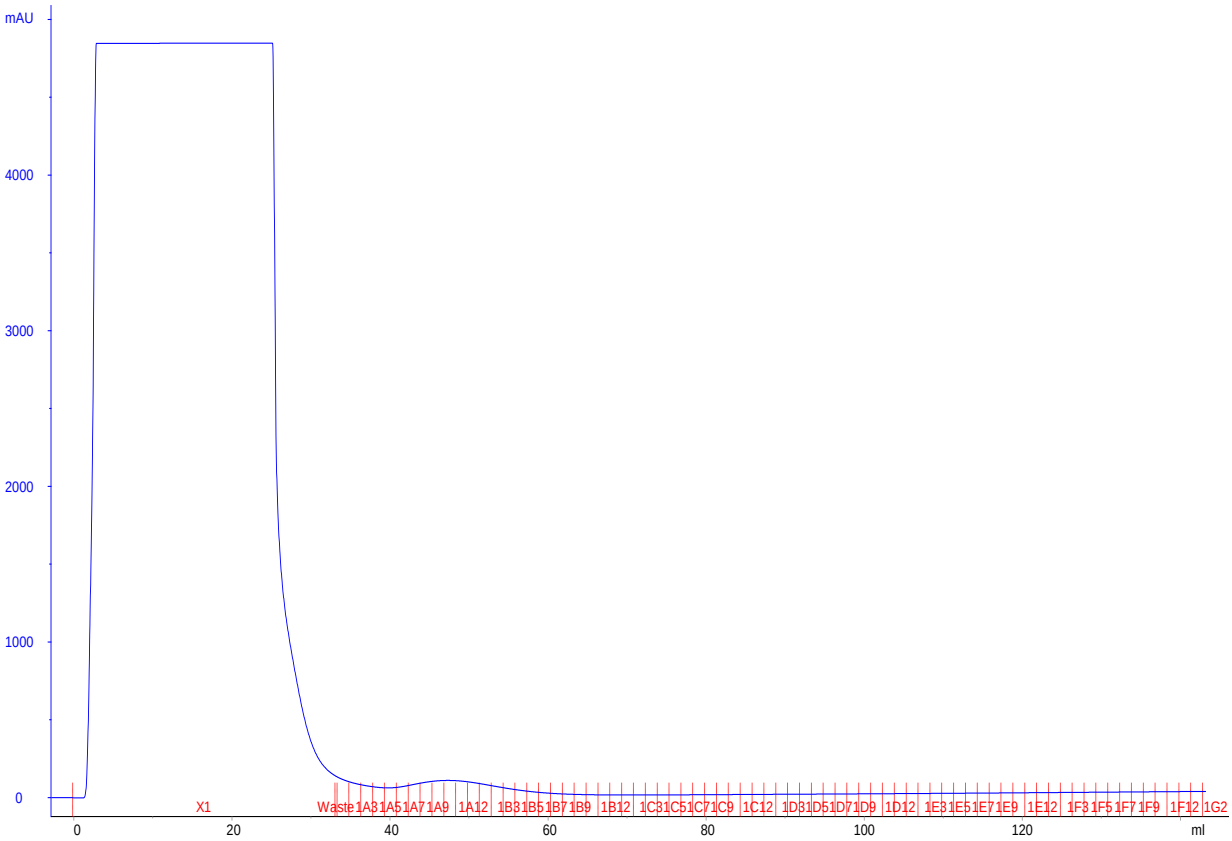

Figure 3.5 Chromatogram of the ion-exchange chromatography with a DEAE column for La Module purification with a small peak that belongs to the eluted nucleic acids

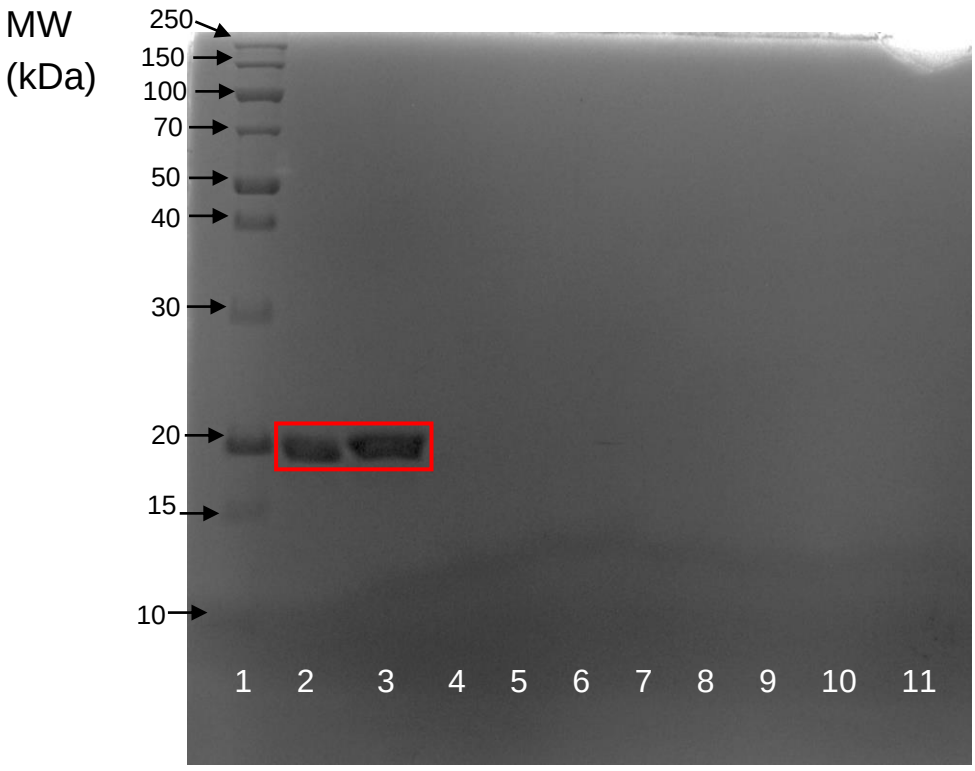

Figure 3.6 SDS-PAGE of the La Module purification using a DEAE column. Label: 1- Ladder; 2- Input; 3- Flowthrough; Fractions 4-A12, 5-B6, 6-B12, 7-C6, 8-C12, 9-D6, 10-D12 and 11-E6

The purification went as expected since the protein eluted in the flowthrough as seen in the gel, while the nucleic acids eluted between fractions A5 and B5. This was expected since the isoelectric point of the La Module of LaRP4B is 5.0 and the working pH was of 7.25, which although in theory makes the protein negative, the interaction with the positive column is quite weak and the amount of salt in the buffer prevents the protein from binding eluting in the flowthrough. The nucleic acids bind strongly and need a greater concentration of KCl to elute. This claim can only be supported by the chromatogram, since the SDS-PAGE gel only shows proteins due to staining with Coomassie Blue. To finish, the gel also showed the impurities between 70 and 150 kDa were purified during this step.

3.2.3. 1D proton NMR

To assess if the purified protein was folded correctly, a 1D ^{1}H-NMR experiment was conducted. The spectrum is shown in figure 3.7.

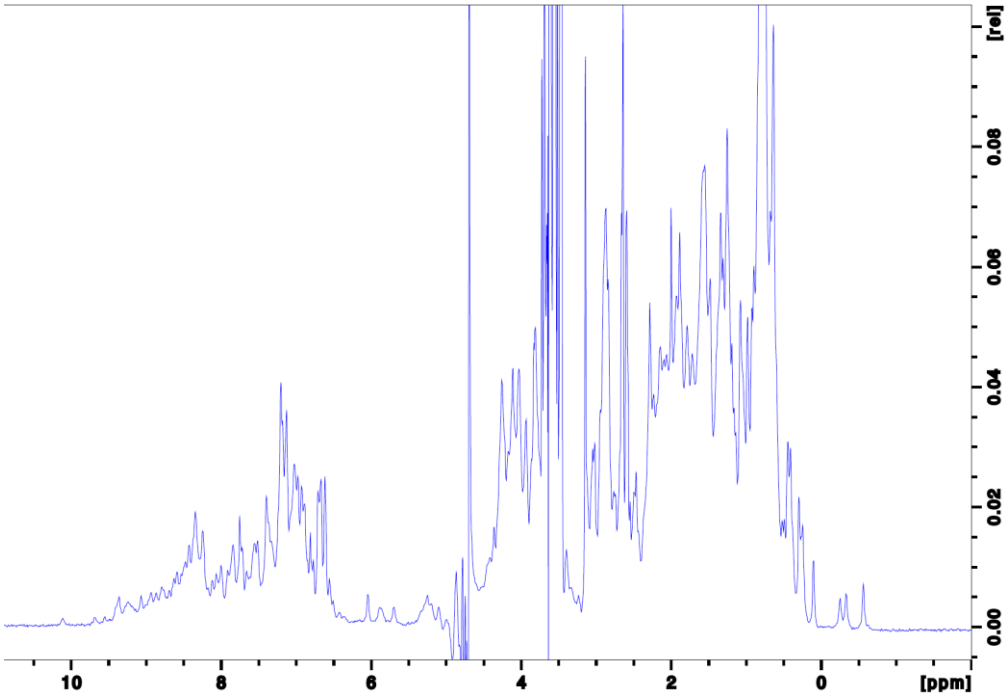

Figure 3.7 1D 1H-NMR spectrum of La Module in buffer with 20 mM Tris, pH 7.25, 100 mM KCl, 0.2 mM EDTA and 1 mM DTT and 10% D2O. Experiment performed on a Bruker Avance III 700 MHz at 25 ºC

The obtained spectrum indicates that the protein is folded. This is backed up mostly by two areas: firstly the range of chemical shifts between 5 ppm and 9.5 ppm which would be narrower if the protein was unfolded due to the fact that amino acids have very similar structures and when the protein is unfolded the chemical shift doesn't vary as much as in a folded form. This happens because the chemical environment is different for each residue resulting in a wider distribution of frequencies; secondly the signals around 0 and -1 ppm belong to the methyl which has no separate signal in a unfolded protein.

For the stronger peaks in the spectrum: the one around 4.5 ppm belongs to water, while the one at 3.5 ppm is suspected to belong to EDTA which is part of the buffer.

3.3. LaRP4B RRM

The LaRP4B RRM (233-328) is a 97 amino acids motif with a molecular weight of 11103.71 Da (11.1 kDa), an isoelectric point of 6.26 and an extinction coefficient of 9970 M^{-1} cm^{-1}. With the tag, the molecular weight is of 13886.57 Da (13.9 kDa), an isoelectric point of 6.29 and an extinction coefficient of 11460 M^{-1} cm^{-1}.

3.3.1. Expression tests

As for the RRM, there is also a reported protocol which was tested at the same temperatures of the La Module expression tests. Figure 3.8 presents the SDS-PAGE gel of said expression test.

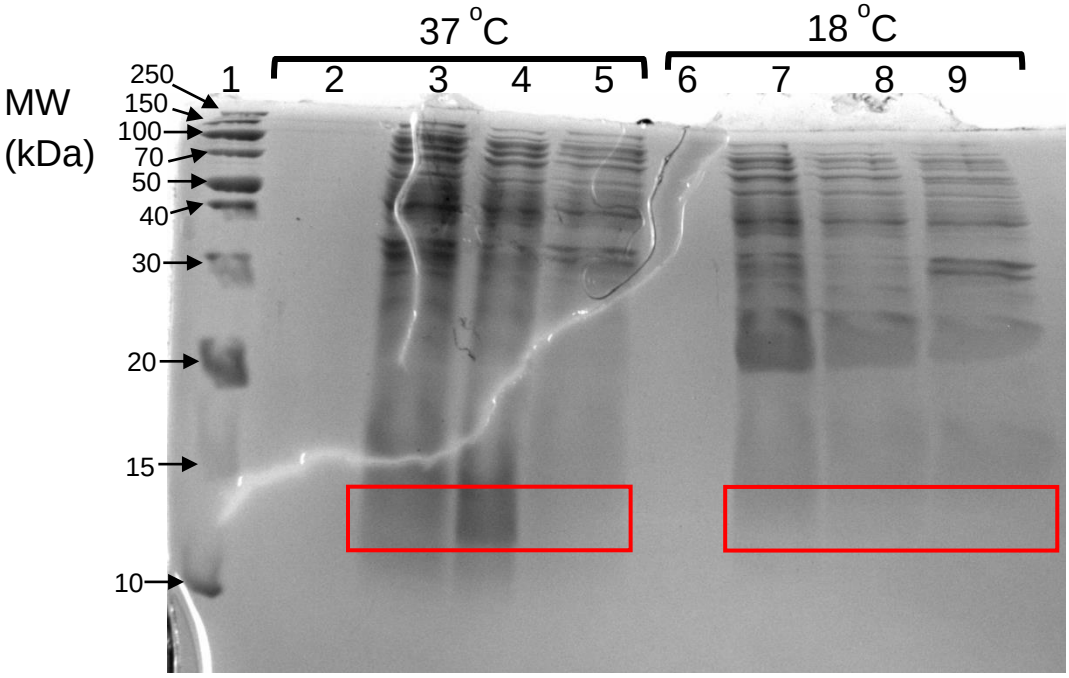

Figure 3.8 SDS-PAGE of the RRM expression tests. Label: 1- Ladder; 2,6- Before Induction; 3,7- Total fraction; 4,8- Soluble fraction; 5,9- Insoluble fraction

As demonstrated by the picture above, the 37 ºC condition has a considerable protein expression in comparison to the 18 ºC one. This is in accordance with the established protocols stating that using 37 ºC across the whole process is better for expression.

3.3.2. Protein purification

The protocol used for the RRM was the same as for the La Module except for using a Heparin column instead of DEAE. The first step was once again an affinity chromatography with a His-Trap column. Figure 3.9 (a) and (b) display the chromatogram while figure 3.10 is the corresponding SDS-PAGE gel.

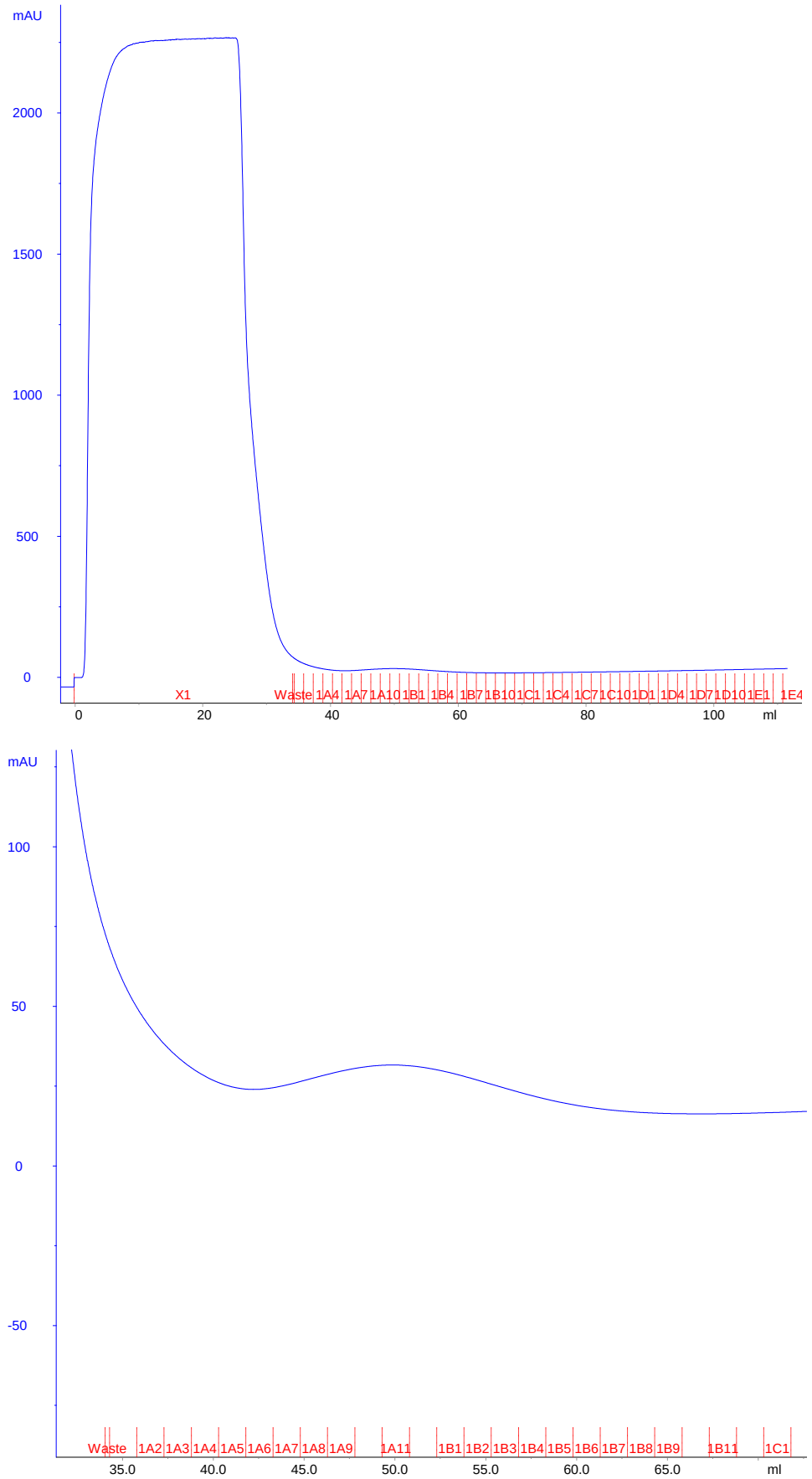

Figure 3.9 (a) Chromatogram of the His-Trap column affinity chromatography purification of the RRM; **(b)** Zoomed in peak of the eluted RRM

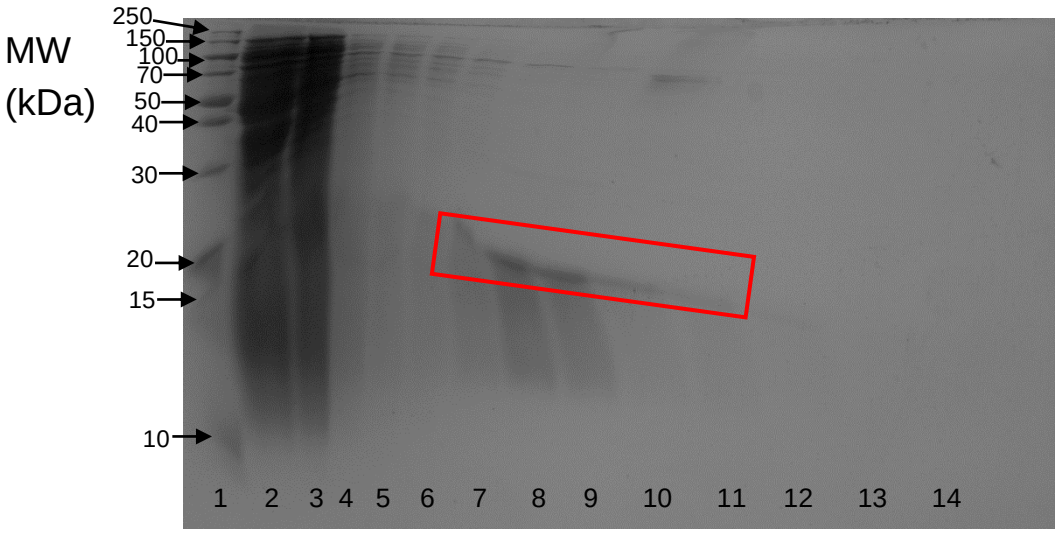

Figure 3.10 SDS-PAGE of the RRM purification with a His-Trap column. Label: 1- Ladder; 2- Input; 3- Flowthrough; Fractions 4-A1, 5-A3, 6-A5, 7-A7, 8-A11, 9-B2, 10-B5, 11-B7, 12-B10, 13-C1 and 14-C3

As depicted by the figures above, the protein eluted between fractions A5 and B7 with the presence of impurities around 70 kDa and around 30 kDa.

The figure below (3.11) presents the gel for the manual His-Trap column run after protease cleavage.

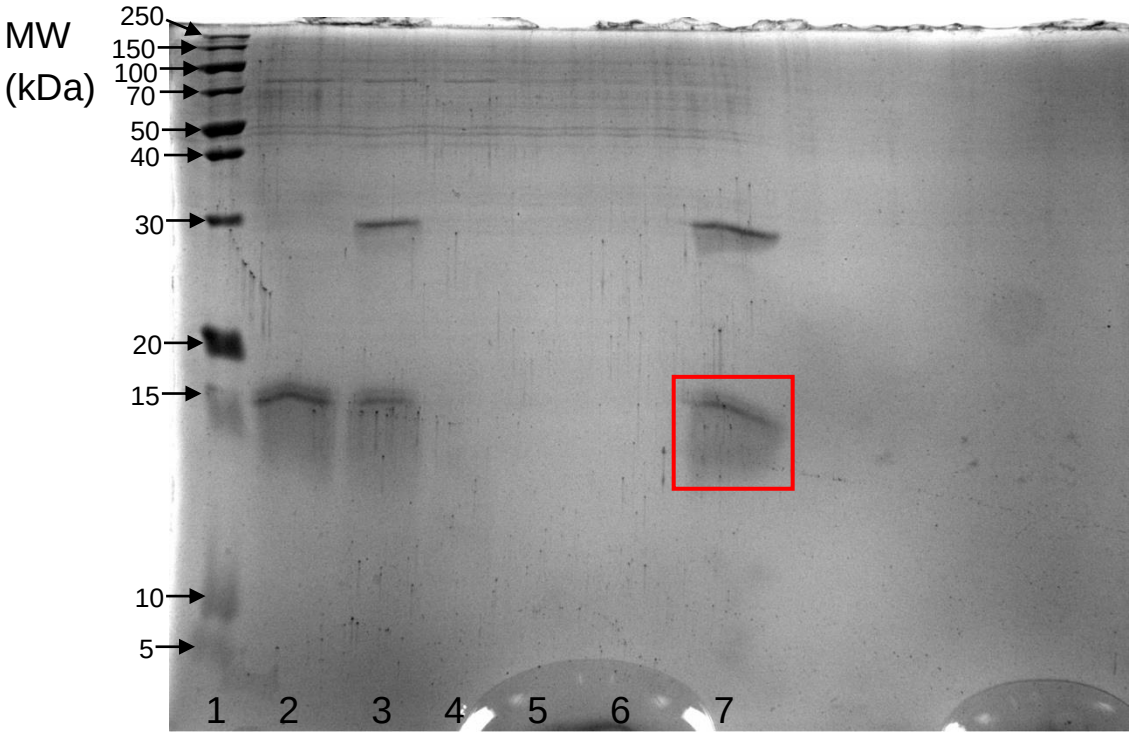

Figure 3.11 SDS-PAGE of the manual His-Trap column after TEV cleavage. Label: 1- Ladder; 2- Pre-TEV; 3- After-TEV; 4- Flowthrough; 5- 1st Wash; 6- 2nd Wash; 7- Elution

Contrary to the La Module, the TEV protease was not able to cleave the His-Tag from the RRM as seen in lane 7 of the gel. This was expected due to previous lab experiments that showed small to no cleavage. However, the purification proceeded as normal meaning that the protein would be purified with the His-tag.

For the last step, as mentioned before, the column used was a Heparin column for an ion-exchange chromatography. The difference here is in the resin charge which is negative. The opposite charge will make the RRM bind to the resin while the nucleic acids are eluted in the flowthrough. Figure 3.12 (a) and (b) represent the chromatograms, while figure 3.13 is the obtained SDS-PAGE gel.

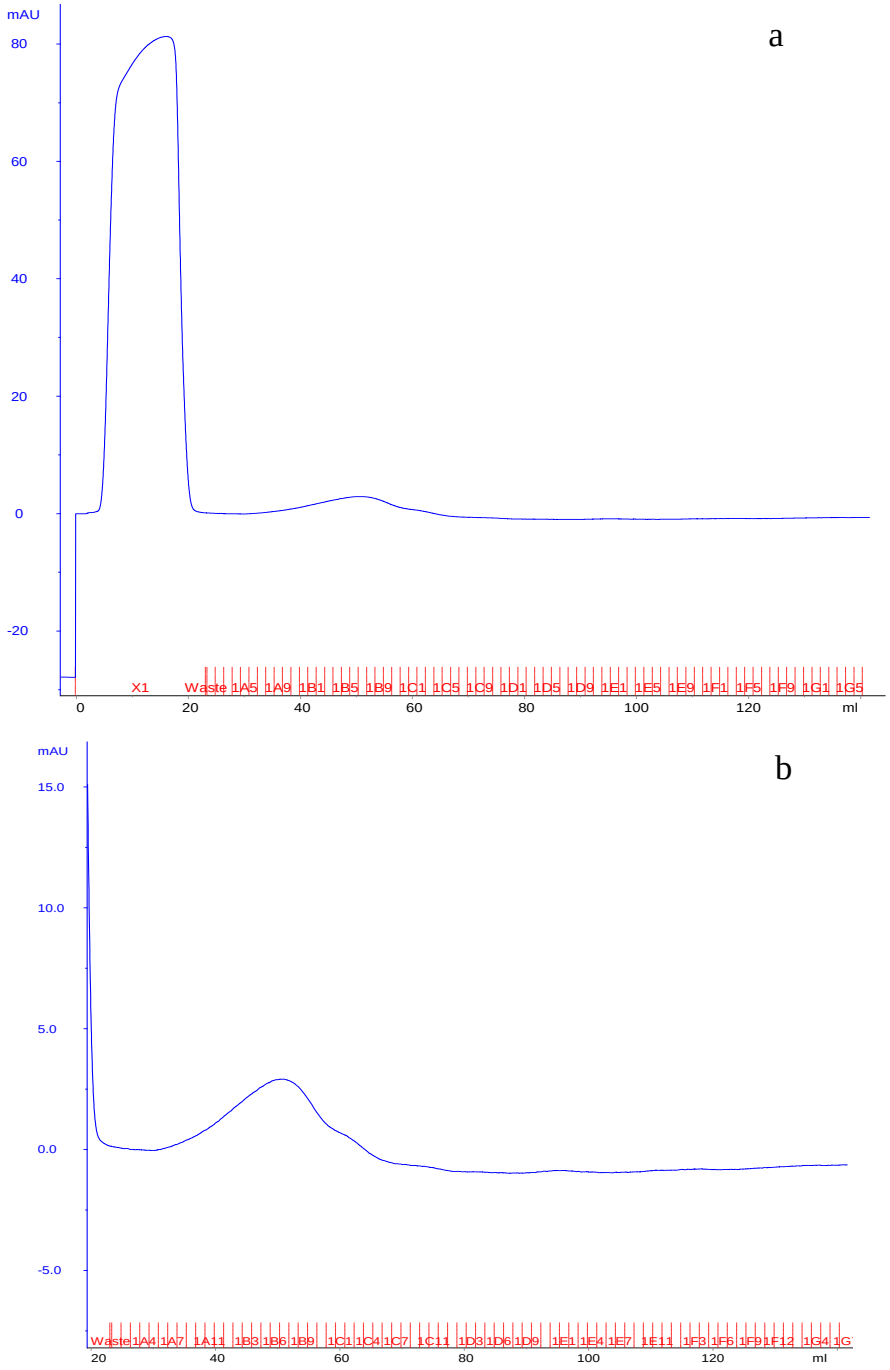

Figure 3.12 (a) Chromatogram of the ion-exchange chromatography with a Heparin column for RRM purification **(b)** Zoomed in peak of the eluted RRM

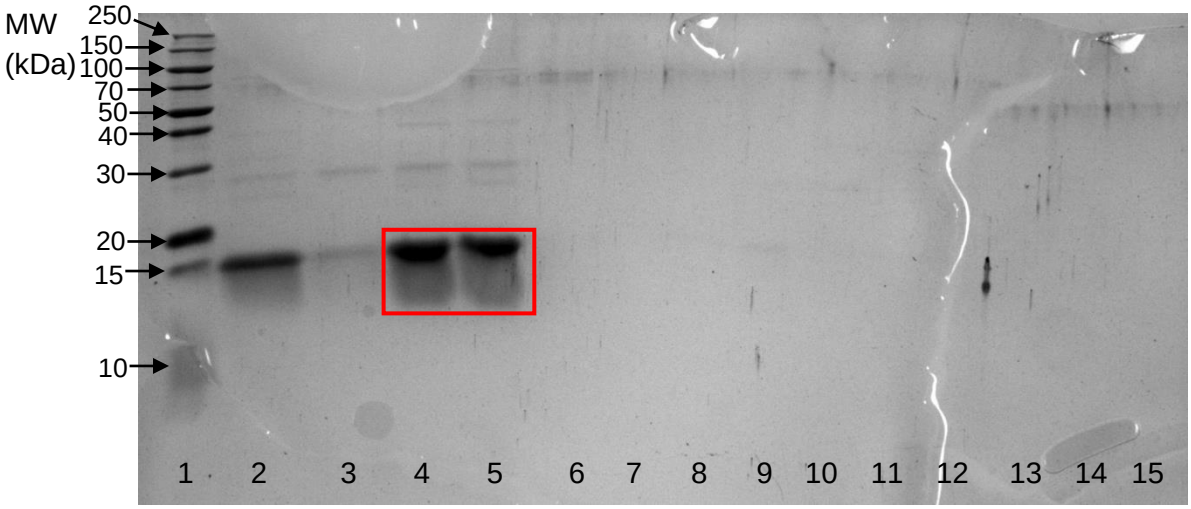

Figure 3.13 SDS-PAGE of the ion-exchange chromatography with Heparin column for RRM purification. Label: 1- Ladder; 2- Input; 3- Flowthrough; Fractions 4-B6, 5-A12, 6-A2, 7-A4, 8-A7, 9-C1, 10-C4, 11-C8, 12-C12, 13-D4, 14-D10 and 15-E2

The gel and the chromatogram indicate that the nucleic acid separation from the RRM was successful, since the chromatograms show the protein eluted around B7 (also backed up by the gel) meaning it bound to the negative column and the nucleic acids eluted in the flowthrough. This purification with the ^{14}N RRM was routine before the ^{15}N expression and purification and due to this the spotted contaminants were ignored.

3.3.3. 15N RRM expression

With the objective of doing relaxation experiments to further expand on the previous NMR experiments made with the RRM of LaRP4B in order to obtain a structure of the RRM, a ^{15}N RRM protein was expressed. Figure 3.14 displays the resulting gel.

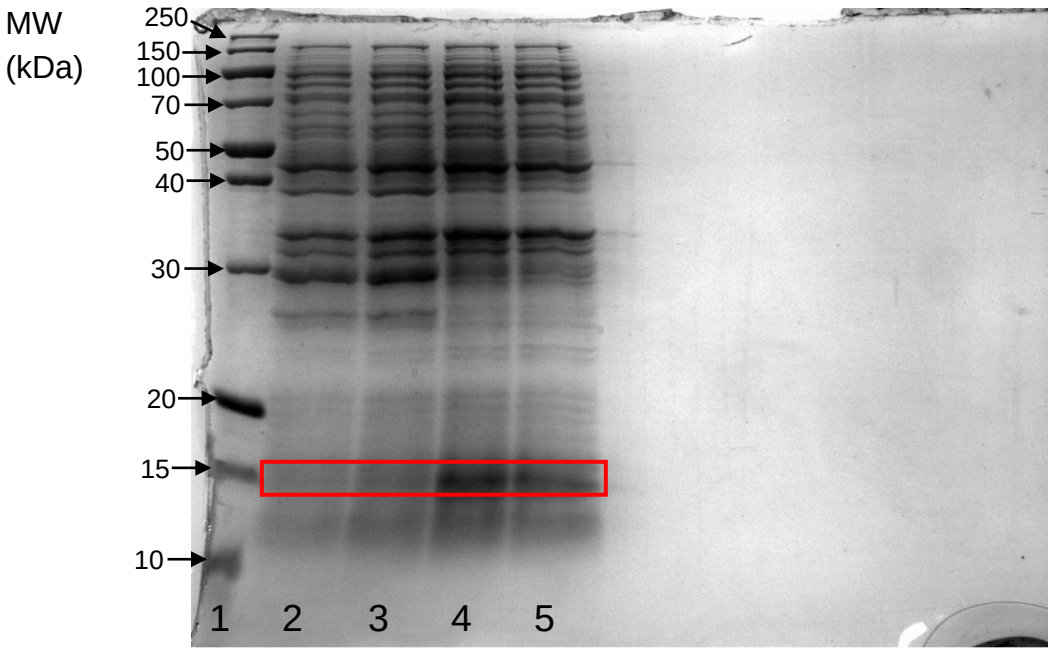

Figure 3.14 SDS-PAGE of the ^{15}N RRM expression. Label: 1-Ladder; 2,3- Before Induction; 4,5- Total fraction

As seen in the figure there was expression of ^{15}N RRM in minimal media at the normal conditions of the ^{14}N RRM.

3.3.4. ^{15}N RRM purification

As observed above, the TEV protease missed its role of cleaving the His-tag, so to lose less protein, this step was not carried out in this purification. As always, the first step was the His-Trap column with its chromatogram and SDS-PAGE gel represented in figures 3.15 (a) and (b) and 3.16, respectively.

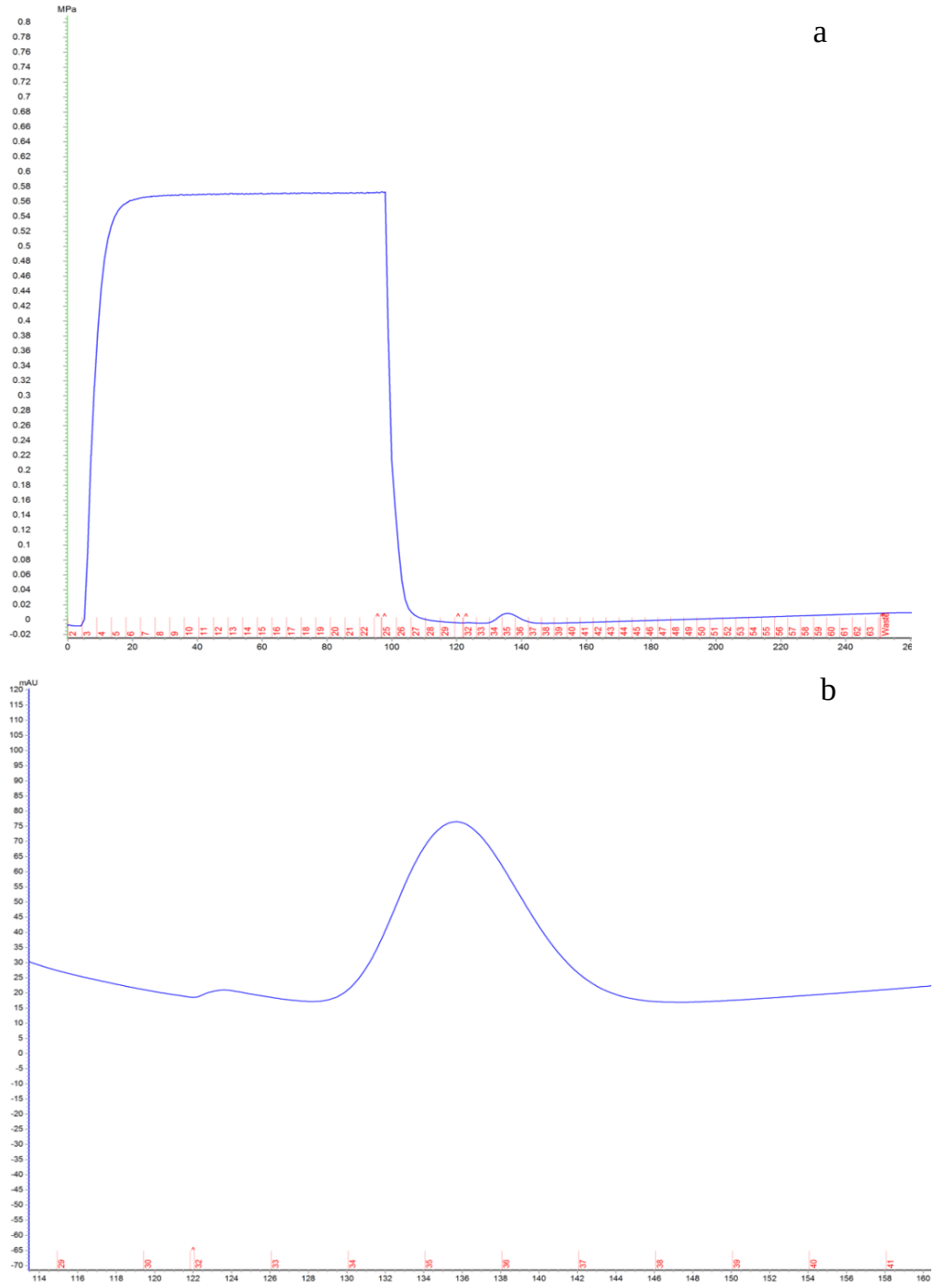

Figure 3.15 (a) Chromatogram of the His-Trap column affinity chromatography purification of the ^{15}N RRM; **(b)** Zoomed in peak of the eluted ^{15}N RRM

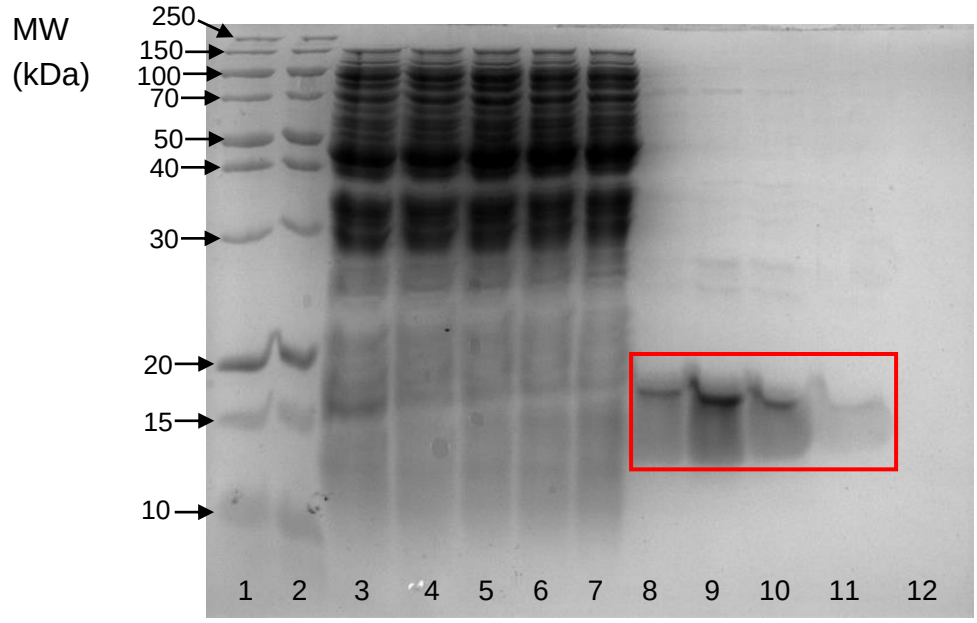

Figure 3.16 SDS-PAGE of the ^{15}N RRM purification with a His-Trap column. Label: 1,2- Ladder; 3- Input; Fractions 4- 6, 5- 11, 6- 16, 7- 21, 8- 34, 9- 35, 10- 36, 11- 37 and 12- 38

As demonstrated by the figures above the first step eluted most of the impurities in the flowthrough (fractions 1 to 22). The target protein eluted between fractions 33 and 37 with the same unwanted proteins from before.

The last step was the heparin column with its chromatograms depicted in figure 3.16 (a) and (b) and the corresponding gel in figure 3.17.

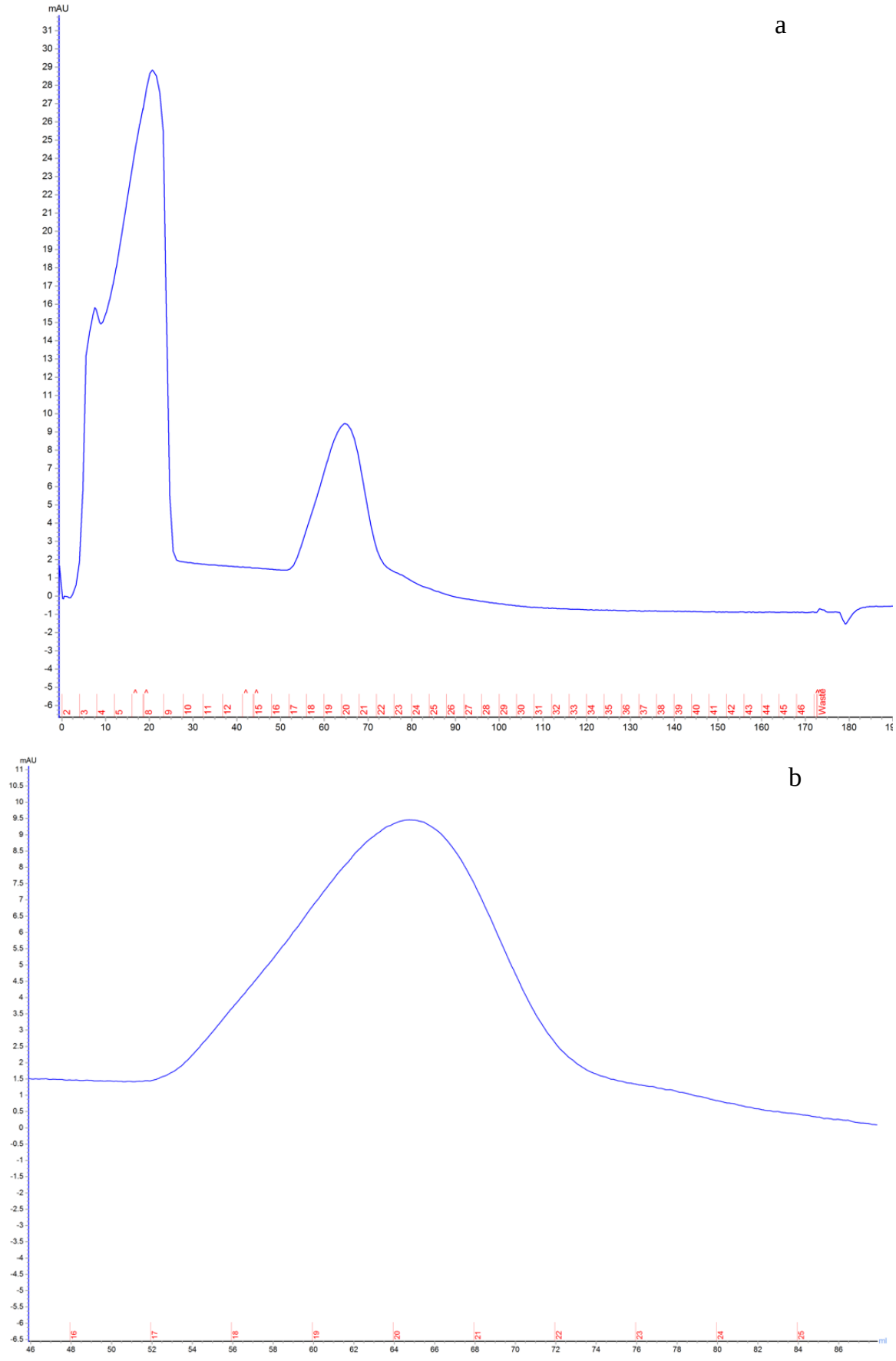

Figure 3.17 (a) Chromatogram of the ion-exchange chromatography with a Heparin column for ^{15}N RRM purification **(b)** Zoomed in peak of the eluted ^{15}N RRM

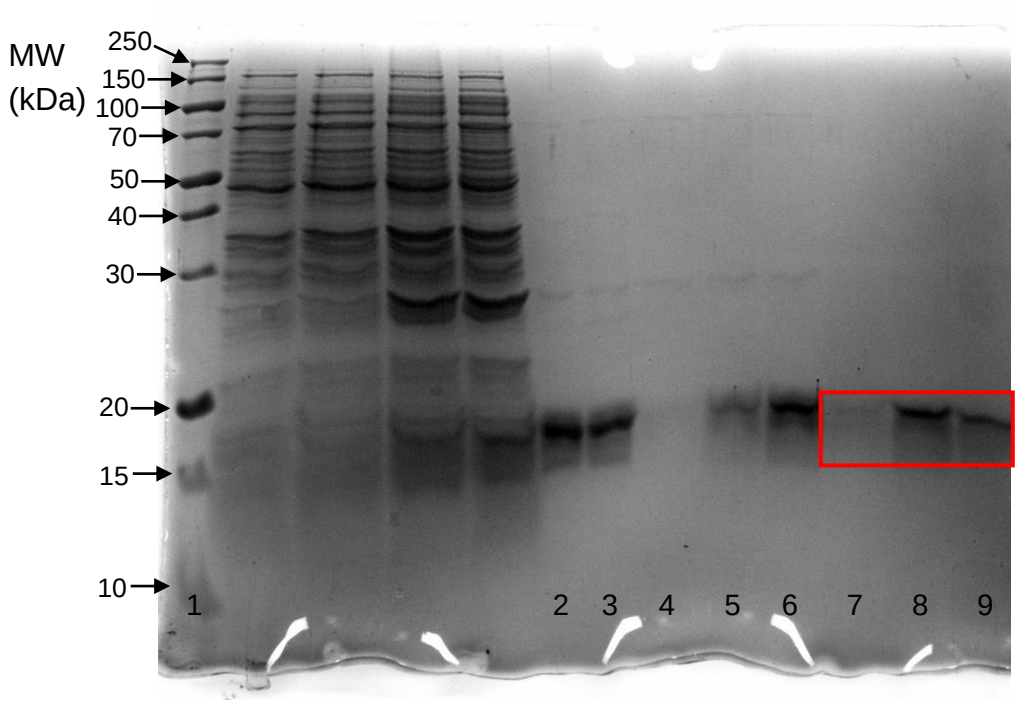

Figure 3.18 SDS-PAGE of the ion-exchange chromatography with Heparin column for ^{15}N RRM purification. Label: 1- Ladder; 2,3- Input; Fractions 4- 3, 5- 5, 6- 8, 7- 17, 8- 19 and 9- 21

Opposite to what happened with the ^{14}N sample, the ^{15}N was completely purified with no traces of other proteins, in fractions 19 to 21. The chromatograms also suggest that the nucleic acids were separated.

3.3.5. Assignment

3.3.5.1. Backbone assignment

The assignment of NMR data from previous experiments with the RRM of ^{13}C and ^{15}N labelled LaRP4B was started using the Analysis program from CCPN. The strategy employed for the assignment of the residues was as follows: first all the peaks were picked and listed in the HSQC spectrum and then assign the backbone. In order to do the backbone assignment, several NMR experiments were used such as the HNCACB, HN(CO)CACB, HNCA and HN(CO)CA spectra. The idea behind the backbone assignment is to first use the HNCACB and the HN(CO)CACB to assign each Cα and Cβ to their respective residues. This is achieved since the HNCACB is an experiment that correlates each NH group with the Cα and Cβ of the same residue which results in a strong signal, but also correlating with the carbons of the previous residue resulting in a weaker signal. As for the HNC(CO)CACB it only correlates the HN to the Cα and Cβ of the preceding amino acid. Figure 3.19 shows the process of using both experiments together to assign the backbone of consecutive residues.

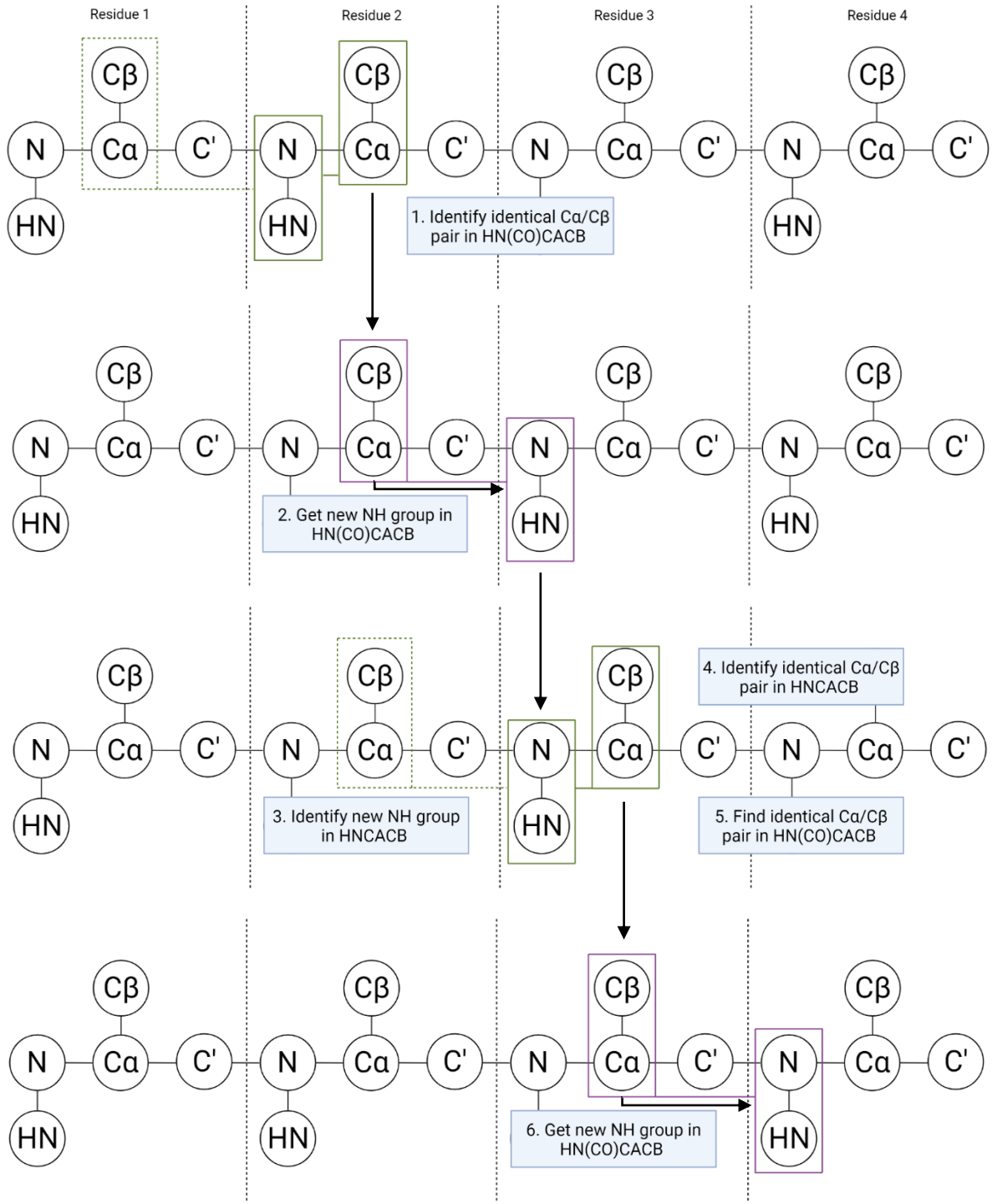

Figure 3.19 Schematic representation of the strategy used for backbone assignment using the HNCACB (green lines) and HN(CO)CACB (purple lines) experiments. Image made in BioRender

In an NMR spectrum, the way this works is through identifying the corresponding signals for the Cα and Cβ chemical shifts, since they adopt values that are characteristic to each amino acid. Figure 5.1 of the Appendix presents these values, and some residues have specific shifts which allow for an easier identification, with them being: Alanine, Serine and Threonine for their Cβ values; Glycine for

not possessing a Cβ and having a shifted Cα; Valine, Isoleucine and Proline for having Cα shifts lower than the rest of the residues. In some cases, the HNCACB and HN(CO)CACB shifts may not possess good quality shifts and in these cases the HNCA/HN(CO)CA and HNCB/HN(CO)CB pairs may be used to complement the analysis with the strategy being the same.

As can be seen in the figure below (3.20), corresponding to the assignment graph, only the NH's, Cα's and Cβ's of 36 amino acids were assigned which corresponds to a completion of 37.89% of the total residues. The assignment was interrupted only to focus on lab work and the comparison between LaRP4A and LaRP4B, with the rest of the assignment being carried out by another lab member (James Jarvis).

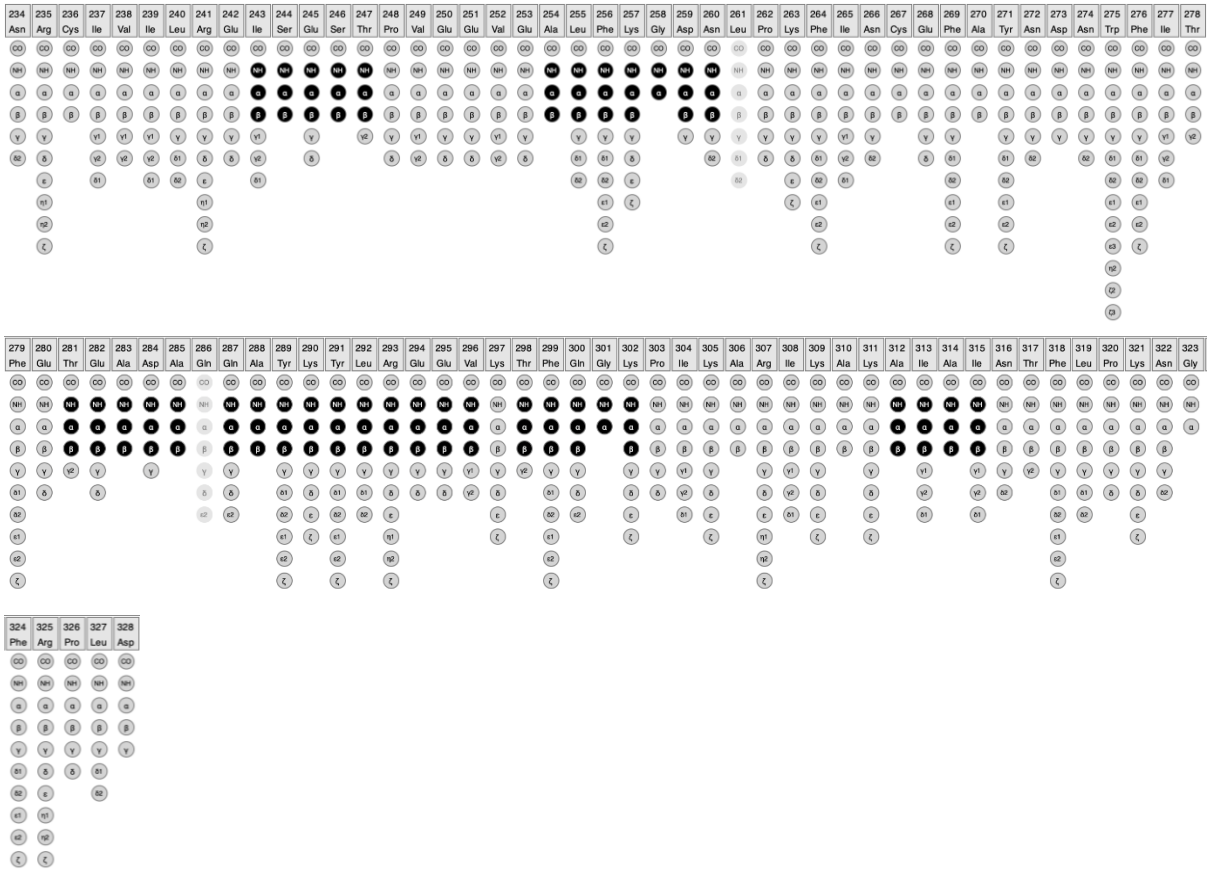

Figure 3.20 Assignment graph of the assigned NH, Cα and Cβ atoms of the residues belonging to the LaRP4B RRM with a completion percentage of 37,89 of the total residues

Figures 3.21 and 3.22 are a multi-panel display of the residues from threonine 281 to alanine 285 as an example of some of the assigned residues.

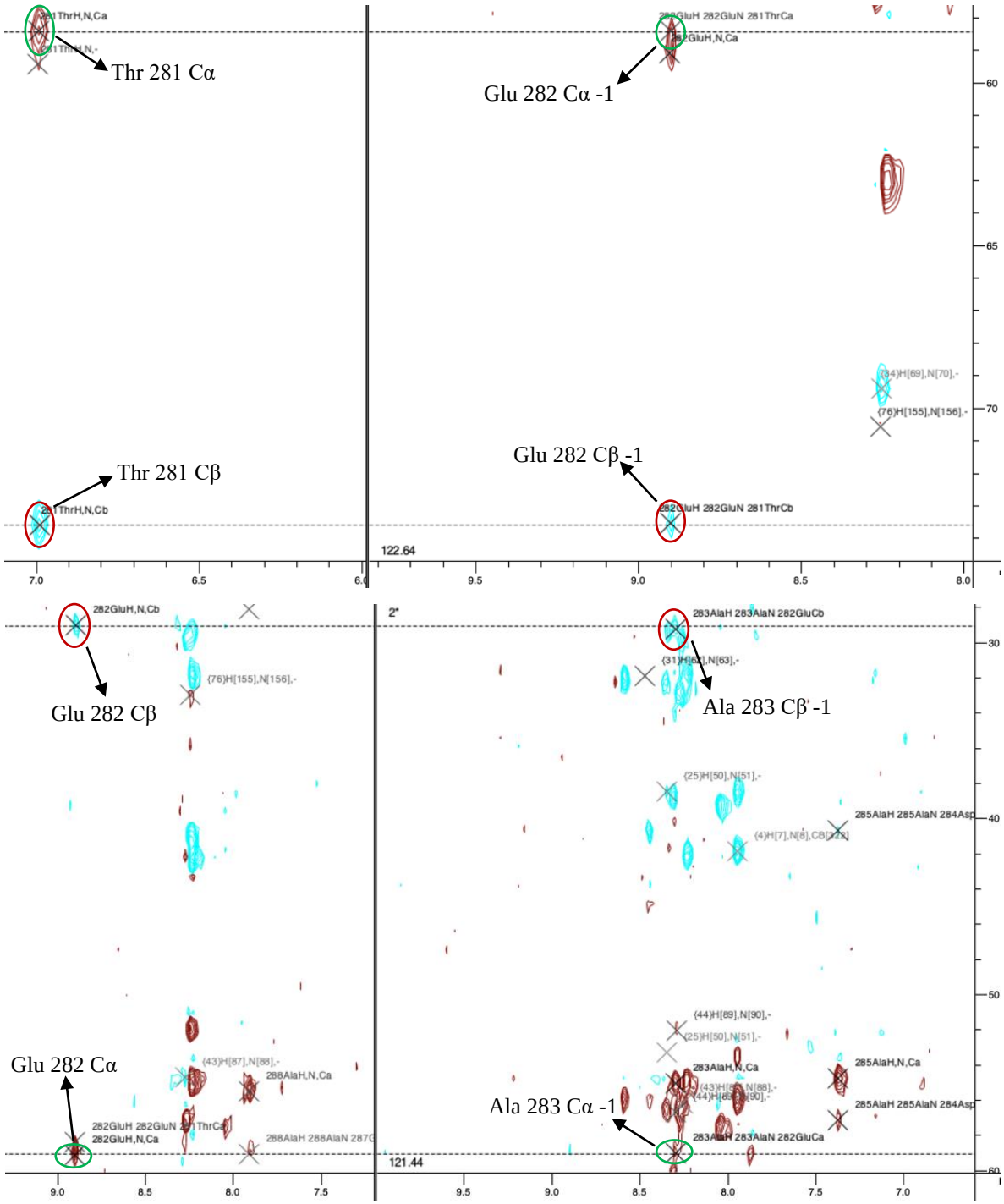

Figure 3.21 Multi-panel display of an HNCACB spectrum with the assigned sequence of residues: Thr281-Glu282-Ala283. Red circles indicate Cβ's and green circles indicate Cα's

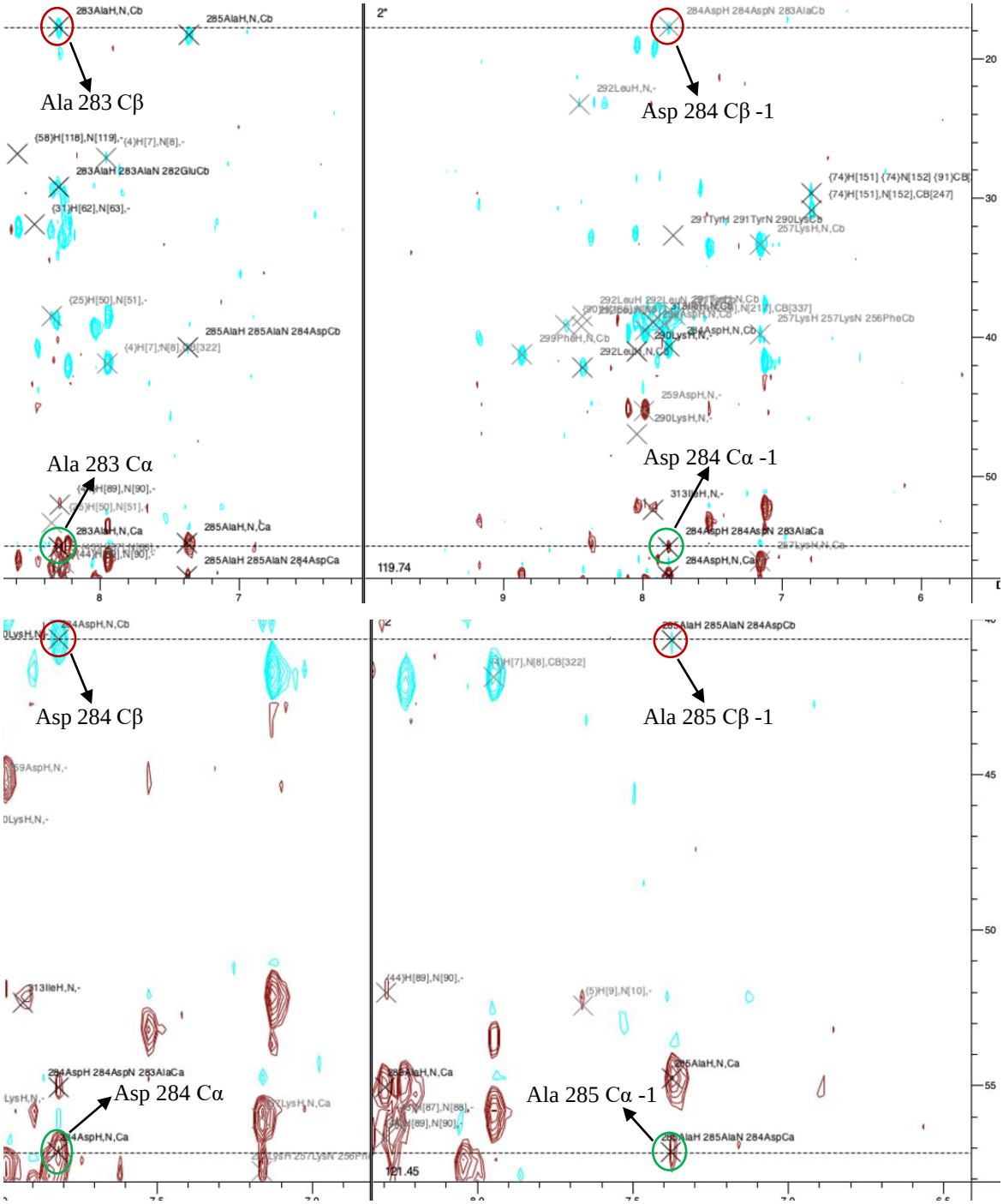

Figure 3.22 Multi-panel display of an HNCACB spectrum with the assigned sequence of residues: Ala 283-Asp284-Ala285. Red circles indicate Cβ's and green circles indicate Cα's

As stated before, the RRM is usually purified with the His-tag since the TEV protease doesn't cleave the tag. With this information, it was expected that some peaks with low to almost no intensity would appear during assignment since the His-tag is in the beginning of the protein and as such should be less structured. However, this was not the case since what was observed were strong, intense peaks, indicating that the tag residues may be more structured than what is normally expected. This can indicate that the tag residues are interacting with the protein amino acids. This statement can be backed up by some other evidence during purification because as seen in the gels in figures 3.10, 3.11 and 3.16,

there is always a large amount of protein that doesn't bind to the His-Trap column and comes out in the flowthrough and this can be explained by the possible fact that the tag is unavailable by the interaction with some of the protein residues. On top of all there is also the lack of cleavage by TEV which further supports the hypothesis of the His-tag not being properly available.

Since this interaction between the tag and some residues can lead to errors in a possible obtainable structure, an optimization of the purification protocol was required and was employed so that the NMR data is not affected by the His-tag.

3.3.5.2. Protocol Optimization

The first attempt at optimizing the protocol was by increasing the TEV concentration in the dialysis step, while the dialysis occurred at two different temperatures: room and 37 ºC. Below is the SDS-PAGE gel after the manual His-Trap column (figure 3.23).

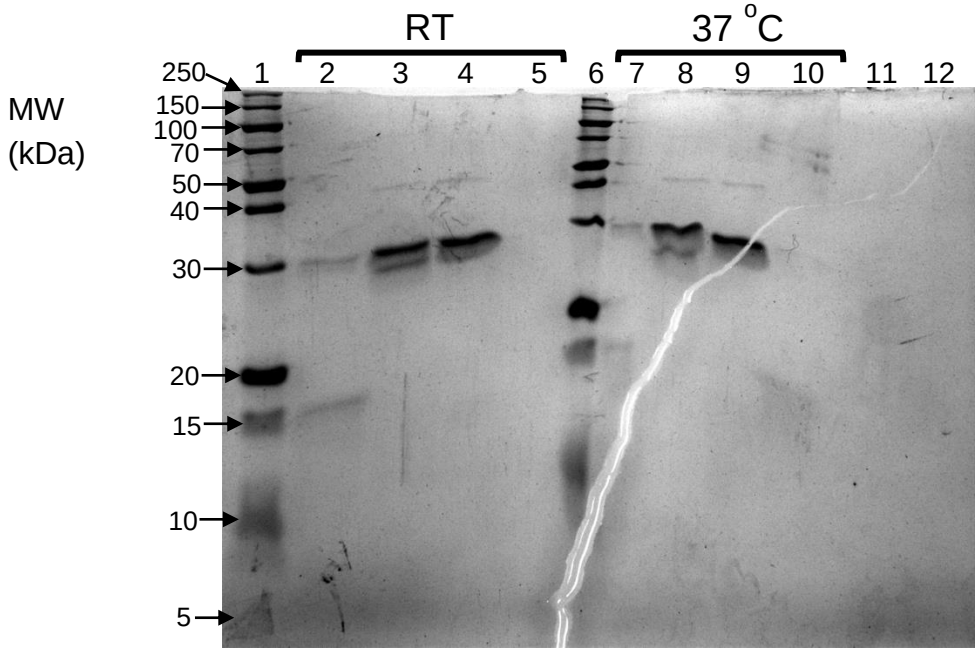

Figure 3.23 SDS-PAGE of the TEV cleavage optimization protocol at both room temperature and 37 ^{0}C. Label: 1,6- Ladder; 2,7- Pre-TEV; 3,8- After-TEV; 4,9- Flowthrough; 5,10- Elution; 11- Room temperature wash; 12- 37 ^{0}C wash,

Before analysing the gel, it is important to note two facts: first, the RRM used already had been through a cleaving attempt with the usual amount of TEV; second, the available protein was divided in two batches: room temperature and 37 ºC which lead to a low amount of protein in each batch. Based on these facts and with the help the gel, we can clearly notice that no protein was cleaved.

After some consideration, although the experiment was not conducted in perfect conditions as stated above, this proved to not be the best strategy to improve cleavage. As such, another protocol was developed, this time with an on-column cleavage process. The figure below is a schematization of this process (figure 3.24).

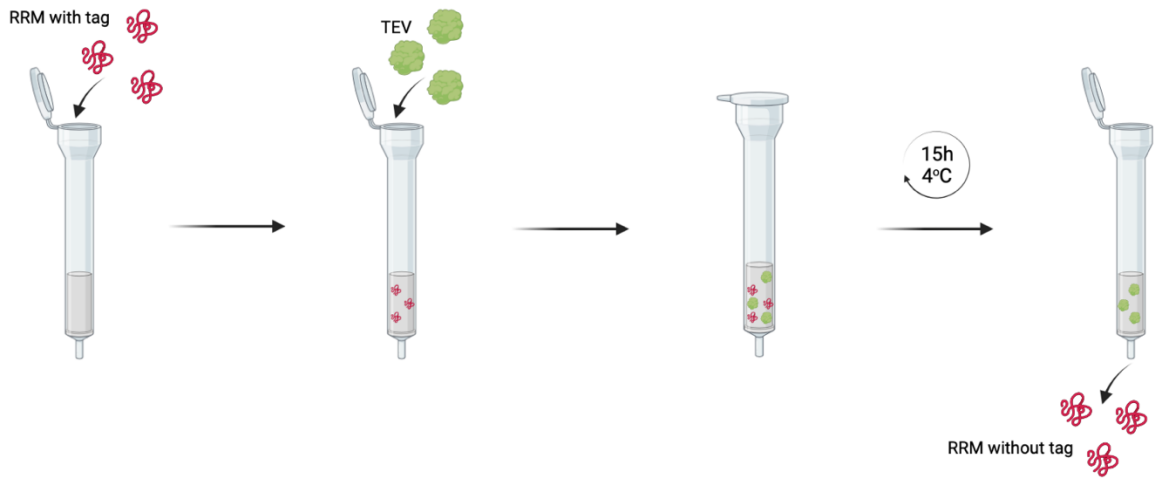

Figure 3.24 Schematization of the on-column cleavage process. Image made in Biorender

The first step represented is the addition of the RRM solution to the His-Trap column. Then the addition of TEV occurs, which also binds to the column since it also has a His-tag, with the column being incubated overnight for fifteen hours at a 4 ºC temperature. Lastly the target protein is eluted without the tag since, theoretically, the TEV would cleave the His-tag present in the RRM. Figure 3.25 is the gel that resulted from this experience.

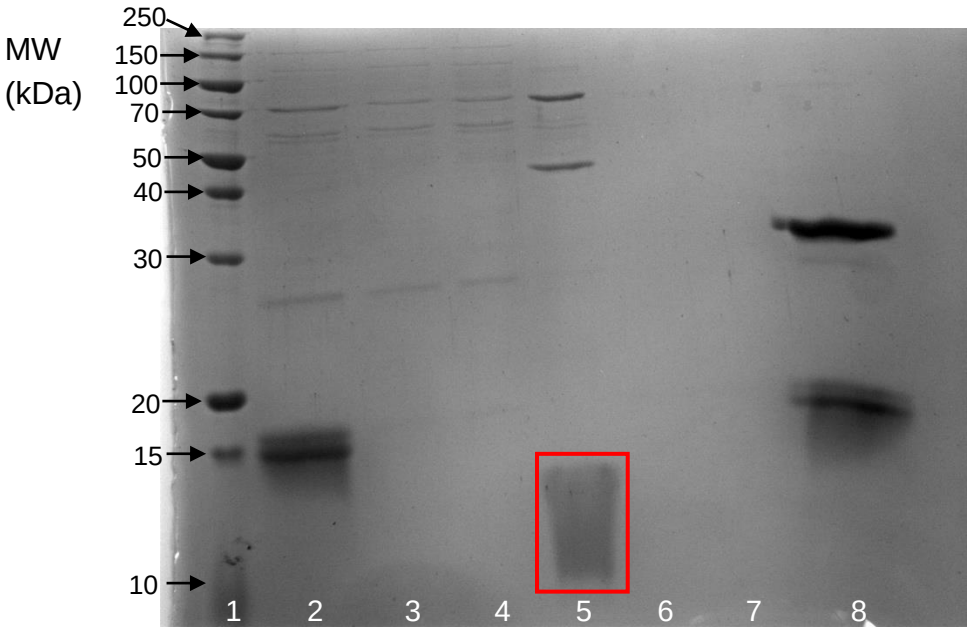

Figure 3.25 SDS-PAGE of the RRM on-column cleavage in a His-Trap column. Label: 1- Ladder; 2- Pre-TEV; 3-Flowthrough; 4- Pre-incubation wash; 5- Post-incubation flowthrough; 6- 1st wash; 7- 2nd wash; 8- Elution

The gel reveals that after the incubation, the protease seems to have cleaved a good amount of protein (lane 5) although a considerable bulk was still left uncleaved (lane 8). There is one hypothesis of why this system works better compared to the previous experiment, in that when the His-tag of the RRM binds to the nickel in the column, this prevents the tag from folding onto the protein, and therefore leaving the sequence exposed to the protease which proceeds to cleave the tag. A visual representation

(without the proteins correct structures and not at scale) of this hypothesis is presented in figure 3.26 for a better understanding.

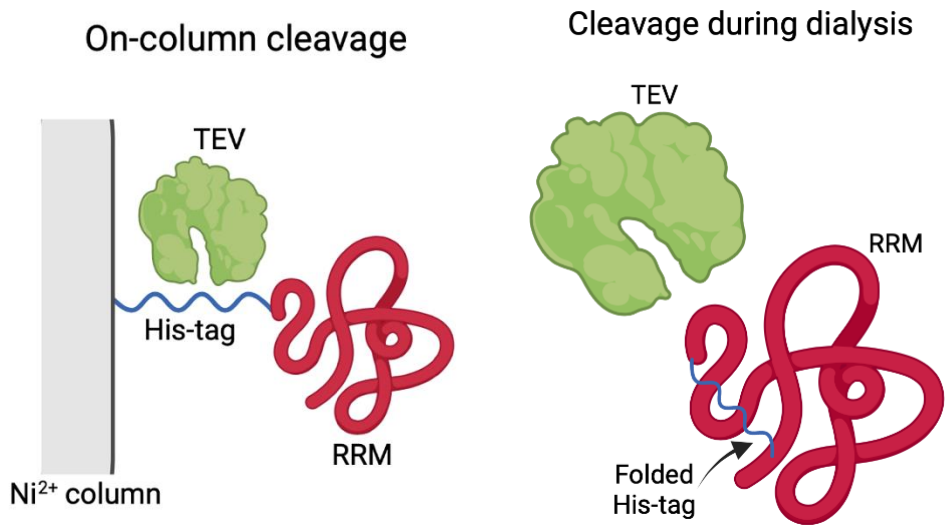

Figure 3.26 Schematic representation of theorized events that occur: on the left during on-column cleavage; on the right in cleavage during dialysis. Image made in Biorender

In order to obtain the highest amount of cleaved protein, other variables need to be fully addressed and optimized such as: TEV concentration; incubation time and temperature; buffer composition since some chemicals may influence the protease activity.

3.4. Comparison between LaRP4A and LaRP4B

With the goal of finding differences between LaRP4A and LaRP4B, since they seem similar at first sight but bind different RNA targets, a structural and data comparison was carried out. To achieve this, NMR data was utilized by comparing peaks that belong to different atoms of each amino acid. The method employed is by using the Chemical Shift Index (CSI) where the random coil shifts are subtracted from the experimental shifts. This enables to obtain an estimate of which secondary structure elements, such as α-helices and β-sheets, our protein possess and compare them with the ones from LARRP4A. This knowledge may help in building a hypothesis of why LaRP4A and LaRP4B bind such different RNAs.

First, an alignment between the LaMs of LaRP4A and LaRP4B, and the RRMs, was made using Clustal Omega. These alignments can be found in figures A.2 and A.3 of the appendix.

Before proceeding to the deviation, a subtraction of the experimental values of the amino acids that are conserved between LaRP4A and LaRP4B was performed. Although this method is not as viable as the CSI since it does not give an estimate of the secondary structure but only differences between conserved residues, it is a good place to start and identify small stretches of residues where a difference exists between both proteins. Figure 3.27 is the bars graphic relative to this variation ($\Delta\delta$) for the RRM, while figure 3.28 is for the La motif.

Figure 3.27 Bars graphic of chemical shift variation (Δδ) between the conserved LaRP4B and LaRP4A residues of the RRM

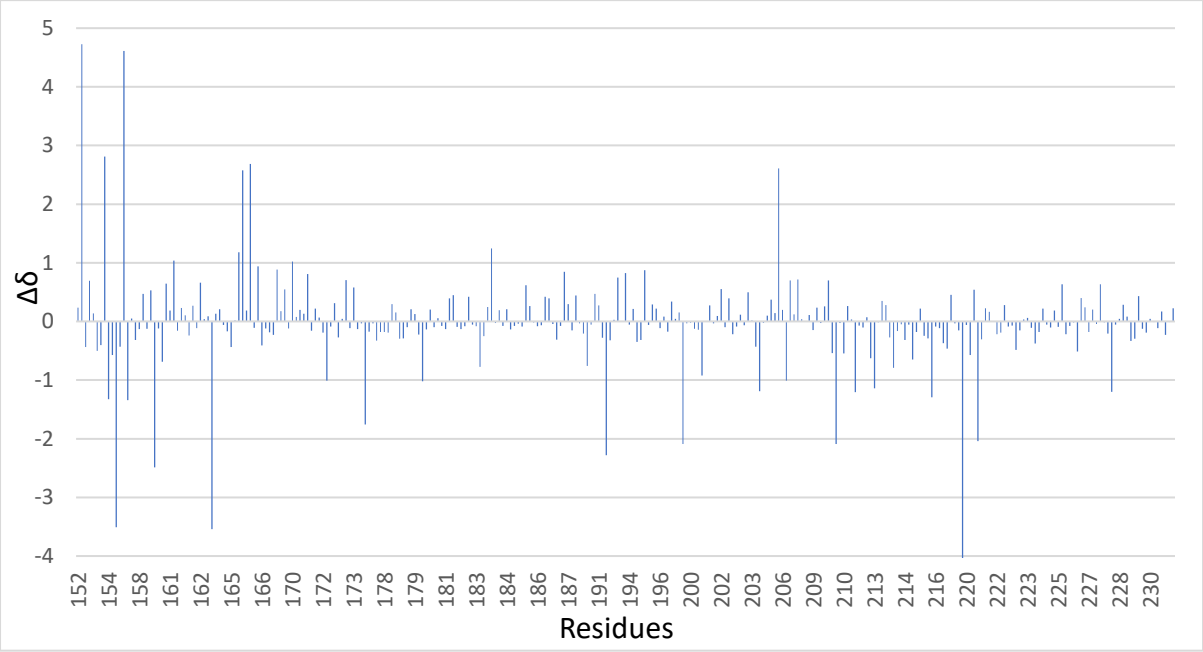

Figure 3.28 Bars graphic of chemical shift variation (Δδ) between the conserved LaRP4B and LaRP4A residues of the La motif

Analysing the RRM first, we can identify three regions where the differences seem more substantial and two where, although they are not as different, it can still carry significance. The first ones are from glutamate 97 to glutamate 100, lysine 118 to cysteine 122 and glycine 174 to arginine 176. The other two are from leucine 110 to lysine 112. It is important to note that between LYS 112 and LYS 118 there is a five amino acids sequence that are not shared by LaRP4A and LaRP4B.

For the La motif there are in total five regions. However, contrary to the RRM, none of these has a large variation which may indicate that the RRM can have a greater importance than the La motif in binding RNA. This could be interesting since the LaRP4A LaM is borderline to the protein-RNA interaction by not playing an important role in comparison with the NTR and the RRM as stated in the introduction and, if these results are to be confirmed by a future structure, then the LaRP4B LaM may also not have a great impact in RNA recognition which would further accentuate its similarities with LaRP4A. The identified regions go from serine 152 to aspartate 155, leucine 160 to phenylalanine 166, threonine 205 and aspartate 206, isoleucine 210 to valine 213 and serine 216 to glutamine 221. The first region might have a bigger difference due to being the N-terminal, which is a flexible stretch of the protein where the shifts variate more.

3.4.1. Chemical Shift Index (CSI)

The chemical shift index is a method to identify which secondary structures exist and where they are in proteins in a quick and accurate way. As stated in previous research, it is of common knowledge that Cα chemical shifts experience a downfield shift when in helices and an up field one when located in β-strands and the opposite happens with the Cβ atoms. The concept of the CSI is basically a two-step filtration method: the first step is based on a chemical shift index, in this case a ternary index with values of -1, 0 and 1, which is given to all the identified residues based on their chemical shifts. The second step is the identification of said secondary structure based of the values and local densities of the indices.[84] In Cα's, positive shifts tend to indicate α-helices, while negative shifts are a sign of β-strands. The opposite happens for Cβ's, with the positive shifts being synonym for β-strands and negative shifts associated with α-helices. The protocol is as follows:[85]

1) With the ^{13}C chemical shift reference values present in figure A.1 of the Appendix, the following process was done: (a) If the obtained Cα chemical shift is above the range given in figure A.1 for the same residue, apply a 1 to it; (b) if the shift is below the range, then a -1 was assigned; (c) lastly, when the measured shift was inside the range then a 0 was given. This procedure allows to define the chemical shift index for each residue in the protein.
2) For the Cα, in every group of four or more 1's that are not halted by a '-1' is a α-helix, while all groups of three or more -1's not stopped by a '1' is a β-strand. Every other region is nominated as a random coil.
3) A local density of different than zero chemical shift indices that goes above 70% is needed when defining helical or strand structures. A minimum of three consecutive -1's is necessary to identify a β-strand, and no less than four 1s are needed for α-helices. The rest of the regions not identified as either α-helix or β-strand, or the ones in which the density goes below 70% are defined as coil.

4) The ends of helices and strands is defined by the opposite chemical shift index (-1 for helix and 1 for strand). However, in some cases this does not happen and when two consecutive zero-valued indices appear that marks the termination point.

The Cβ analysis is different since as stated by Speraf and Bax [86], its chemical shifts for both helices and strands tend to overlap. This overlap has shown that it's almost impossible to consistently identify α-helices with this carbon.[86] So, step 1 remains but the rule stated is that for every group of four or more 1's not interrupted by a '-1' is a strand, while all the other regions are considered a random coil or other element that is not a β-strand. Also, the 70% rule still applies with a minimum of three consecutive 1's needed to define a β-strand. Point 4 also stands the same but applied to Cβ values.[85]

The next figures (3.29 and 3.30) are graphics relative to the chemical shift index of the Cα and Cβ atoms, respectively, of the amino acids from the RRM of LaRP4B with the objective of obtaining a projection of its secondary structure to compare with the RRM from LaRP4A. The peak list used for these CSI calculations is displayed in the Appendix (LaRP4B RRM Peak List) and are data from other lab members.

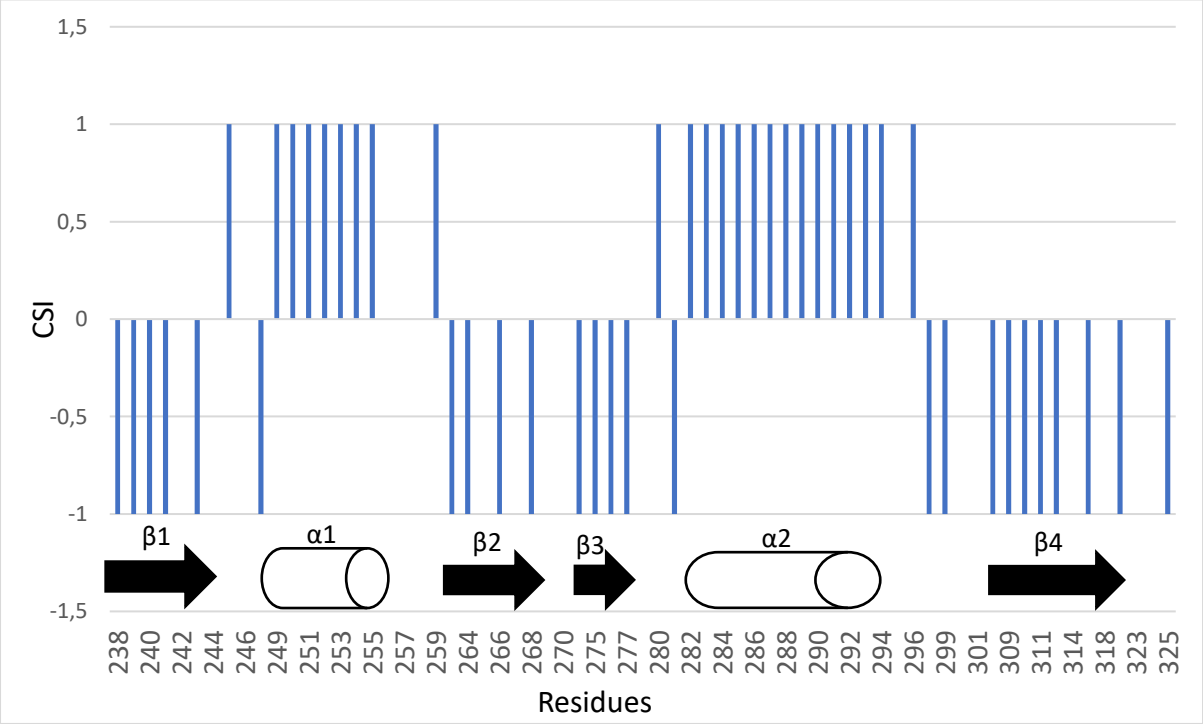

Figure 3.29 Chemical shift index bars graphic for the Cα atoms of the RRM residues

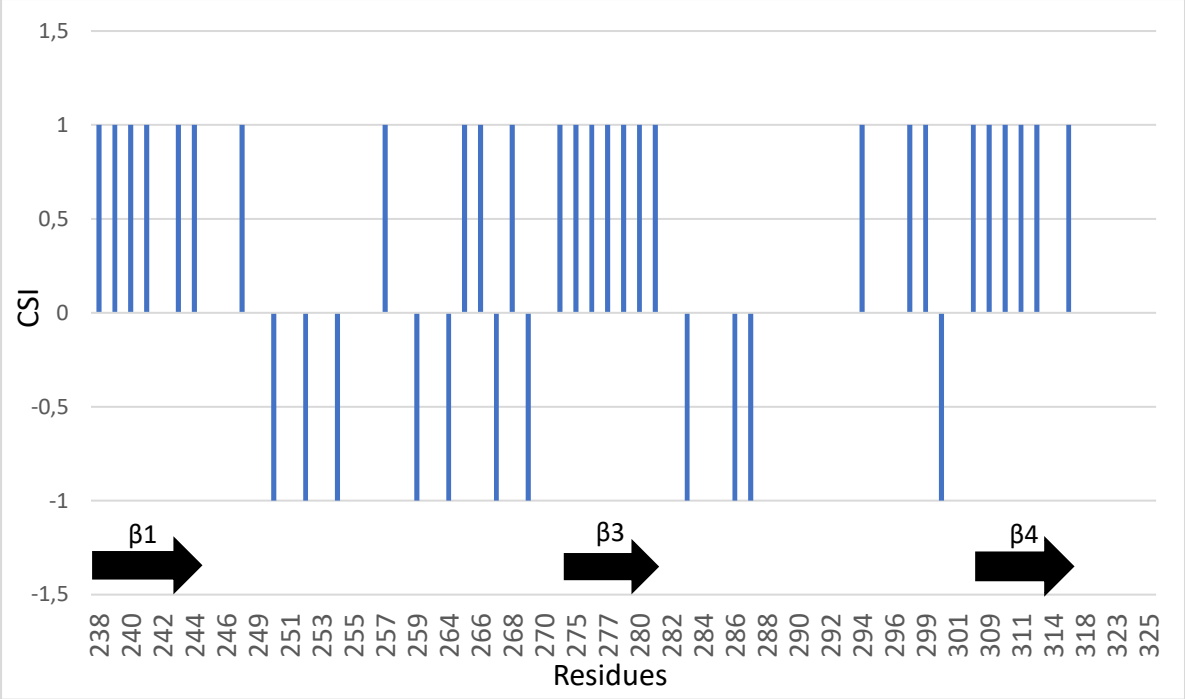

Figure 3.30 Chemical shift index bars graphic for Cβ atoms of the RRM residues

The CSI of the RRM for the Cα showed us that the RRM of LaRP4B possesses four β-strands and two α-helices in a βαββαβ disposition. These results indicate that the secondary structure of the RRM of LaRP4B is like the one found in LaRP4A, since, as stated by Cruz-Gallardo et al, the RRM of this protein also has the same structure, with this being the canonical fold of the RRM.[56,87] The sizes of the helices also go accordingly with the reported LaRP4A structure, with the first helix being smaller and second one larger. With the Cβ graph, strand β3 doesn't appear maybe due to the said overlap between shifts of α-helices and β-strands.[56]

The same was done for the La motif Cα and Cβ atoms of its residues. Below are the figures (3.31 and 3.32) corresponding to the CSI graphs. The peak list is presented in the Appendix (LaRP4B LaM Peak List) and was also obtained by other lab members.

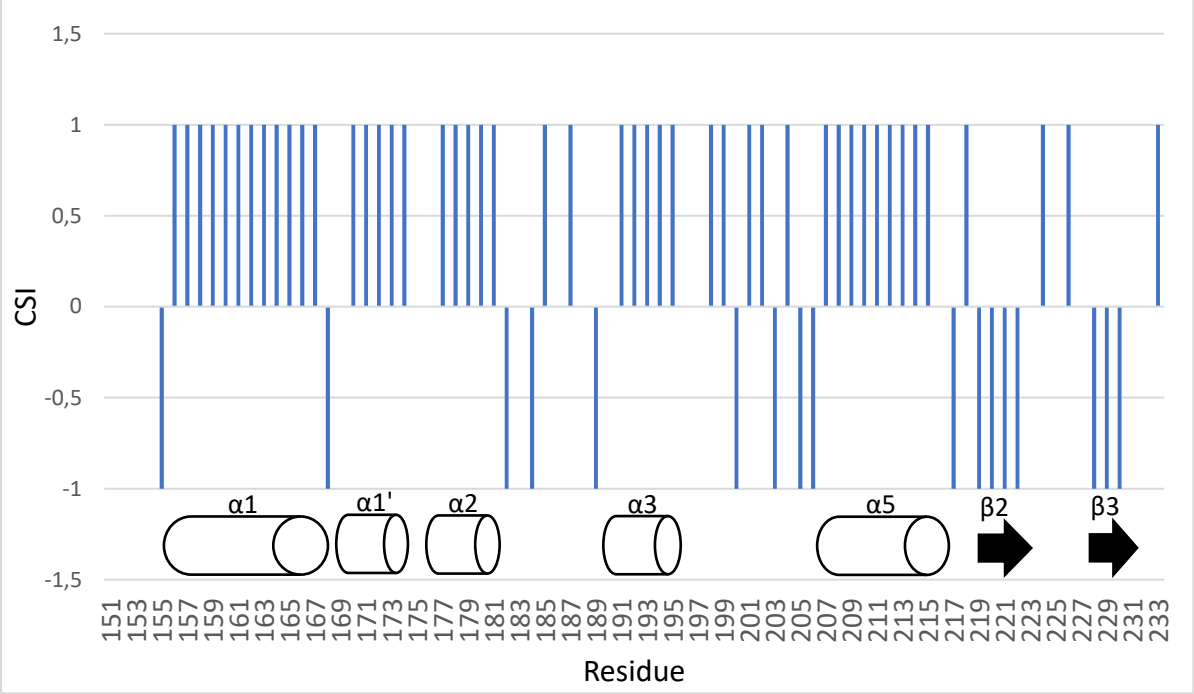

Figure 3.31 Chemical shift index bars graphic for the Cα atoms of the La motif residues

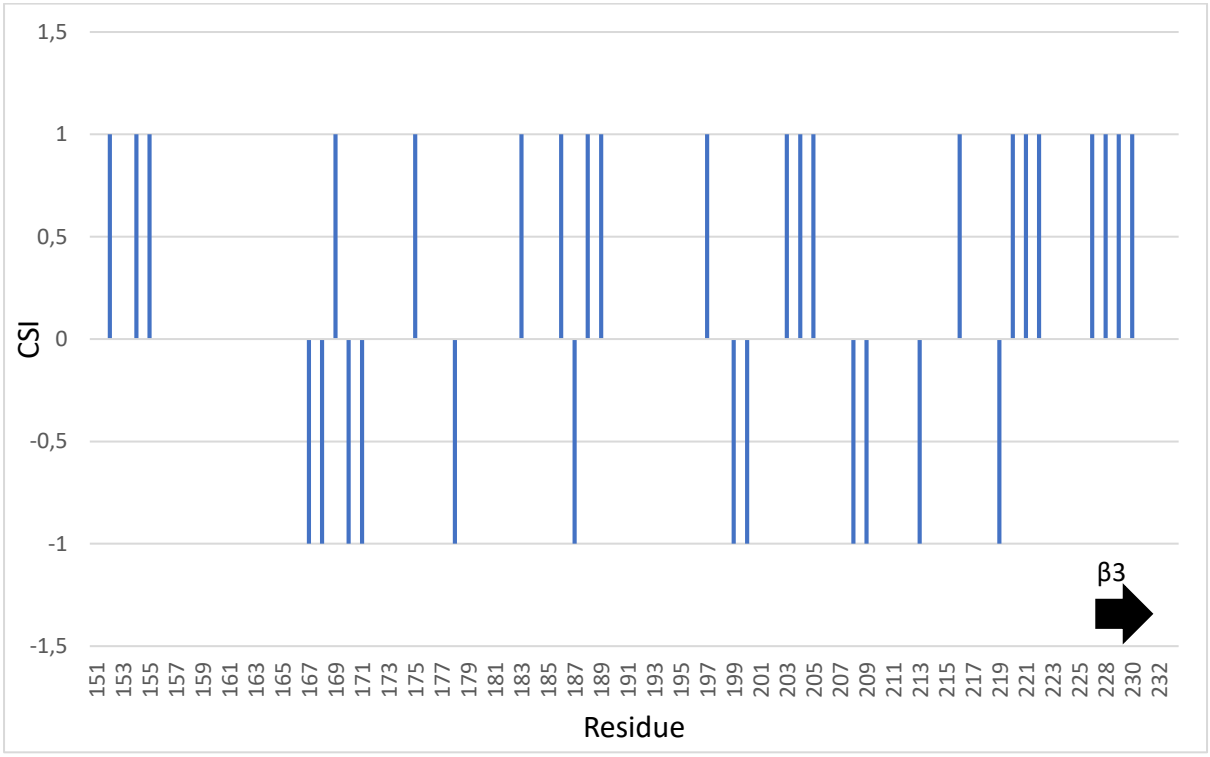

Figure 3.32 Chemical shift index bars graphic for the Cβ atoms of the La motif residues

As shown above, the La motif of LRP4B appears to possess five α-helices and then two β-strands in the end of its structure. This is somewhat different from the reported secondary structure of the LaRP4A La motif.[56] In this report, the La motif of LaRP4A contains six α-helices (α1, α1', α2, α3, α4 and α5) and three β-strands (β1, β2 and β3), in the following arrangement: α1-α1'-α2-β1-α3-α4-α5-β2-β3. The CSI graph, which has the LaRP4A numbering for a better understanding, indicates that

LaRP4B may be missing strand β1 and helix α4, in what would be a major difference in the La motifs of two closely related proteins. However, there is always the possibility of a possible wrong assignment in residues in those areas or bad peak intensity that may change the outcome of this domain secondary structure. As for the Cβ CSI graph, it once again doesn't show one of the strands (β2). In the end, the main conclusion is that although the CSI is providing a different secondary structure for the LaM than the one seen in LaRP4A, this is still an estimate, and should be seen as such, until a future structure of LaRP4B unfolds the secondary structure of this protein. In addition, some of the LaRP4A secondary structure elements, seen in the structure presented before, are quite short and the same could probably happen in LaRP4B which can make it harder to spot in the CSI plot.

The lack of differences between the secondary structure of LaRP4A and LaRP4B RRM and the still to be proved variation in the structure of the La motif, maintain open the possibility of one other hypothesis being impactful to their interactions with different RNA targets: the variation may be connected with specific amino acids that are different to each protein or even the same amino acids being in different conformations which could influence how the protein interacts with different RNAs. Also, the analysis of the size of the linker of LaRP4B shows that it is a three-residue linker (NQN), which is the same size as the linker in LaRP4A. This suggests that the linker does not contribute to the differences in RNA targets between both proteins.

4.

CONCLUSIONS AND FUTURE PERSPECTIVES

LaRPs are important RNA binding proteins responsible for a wide array of RNA related functions and roles. This influence in the life of RNAs is reflected by the fact that several diseases are associated with dysregulations in LaRPs, being overexpression, under expression, mutations and others. With this in mind, the study of LaRP4B is of great importance since its one of the LaRPs that less is known about and with proved impact in cancer, for example.

The objectives proposed in the beginning of this report were, to some extent, accomplished. The main conclusions drawn from the work done were:

- Expression and purification of the La module, RRM and 15N RRM were all successful with little to no changes to the previously optimized protocols.
- Although the assignment of the RRM previous NMR data was halted and picked up by another lab member, it allowed to notice that the uncleaved tag was too structured, which lead to a successful optimization of the RRM purification protocol by adding an on-column cleavage that allowed to purify the RRM without the His-tag.
- Assignment of the RRM of LaRP4B with a completion of 37.89% of the Cα's and Cβ's among all residues.
- A comparison of the secondary structures of LaRP4A and LaRP4B, due to their close relation, was done through CSI and although it showed some possible differences in the La motif, the result is still an estimate and only with the LaRP4B structure it is possible to develop a proper hypothesis.

As for future possible work:

- The optimization of the RRM protocol opens the doors to obtain improved NMR data that may help understand better the RRM structure of LaRP4B and how it interacts with RNA.
- Obtaining the La module, RRM and La motif structures of LaRP4B for a needed detailed comparison with LaRP4A, either by NMR or X-Ray crystallography.

BIBLIOGRAPHY

(1) Dreyfuss, G.; Kim, V. N.; Kataoka, N. Messenger-RNA-Binding Proteins and the Messages They Carry. *Nat. Rev. Mol. Cell Biol.* **2002**, *3* (3), 195–205. https://doi.org/10.1038/nrm760.

(2) Hentze, M. W.; Castello, A.; Schwarzl, T.; Preiss, T. A Brave New World of RNA-Binding Proteins. *Nat. Rev. Mol. Cell Biol.* **2018**, *19* (5), 327–341. https://doi.org/10.1038/nrm.2017.130.

(3) Lunde, B. M.; Moore, C.; Varani, G. RNA-Binding Proteins: Modular Design for Efficient Function. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (6), 479–490. https://doi.org/10.1038/nrm2178.

(4) Steitz, T. A. A Structural Understanding of the Dynamic Ribosome Machine. *Nat. Rev. Mol. Cell Biol.* **2008**, *9* (3), 242–253. https://doi.org/10.1038/nrm2352.

(5) Matera, A. G.; Wang, Z. A Day in the Life of the Spliceosome. *Nat. Rev. Mol. Cell Biol.* **2014**, *15* (2), 108–121. https://doi.org/10.1038/nrm3742.

(6) Singh, G.; Pratt, G.; Yeo, G. W.; Moore, M. J. The Clothes Make the MRNA: Past and Present Trends in MRNP Fashion. *Annu. Rev. Biochem.* **2015**, *84*, 325–354. https://doi.org/10.1146/annurev-biochem-080111-092106.

(7) Gloss, B. S.; Dinger, M. E. The Specificity of Long Noncoding RNA Expression. *Biochim. Biophys. Acta - Gene Regul. Mech.* **2016**, *1859* (1), 16–22. https://doi.org/10.1016/j.bbagrm.2015.08.005.

(8) Anderson, P.; Kedersha, N. RNA Granules: Post-Transcriptional and Epigenetic Modulators of Gene Expression. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (6), 430–436.

https://doi.org/10.1038/nrm2694.

(9) Wright, P. E.; Dyson, H. J. Intrinsically Disordered Proteins in Cellular Signalling and Regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16* (1), 18–29. https://doi.org/10.1038/nrm3920.

(10) Järvelin, A. I.; Noerenberg, M.; Davis, I.; Castello, A. The New (Dis)Order in RNA Regulation. *Cell Commun. Signal.* **2016**, *14* (1). https://doi.org/10.1186/s12964-016-0132-3.

(11) Castello, A.; Fischer, B.; Frese, C. K.; Horos, R.; Alleaume, A. M.; Foehr, S.; Curk, T.; Krijgsveld, J.; Hentze, M. W. Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol. Cell* **2016**, *63* (4), 696–710. https://doi.org/10.1016/j.molcel.2016.06.029.

(12) Casu, F.; Duggan, B. M.; Hennig, M. The Arginine-Rich RNA-Binding Motif of HIV-1 Rev Is Intrinsically Disordered and Folds upon RRE Binding. *Biophys. J.* **2013**, *105* (4), 1004–1017. https://doi.org/10.1016/j.bpj.2013.07.022.

(13) Butter, F.; Scheibe, M.; Mörl, M.; Mann, M. Unbiased RNA-Protein Interaction Screen by Quantitative Proteomics. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (26), 10626–10631. https://doi.org/10.1073/pnas.0812099106.

(14) Castello, A.; Horos, R.; Strein, C.; Fischer, B.; Eichelbaum, K.; Steinmetz, L. M.; Krijgsveld, J.; Hentze, M. W. System-Wide Identification of RNA-Binding Proteins by Interactome Capture. *Nat. Protoc.* **2013**, *8* (3), 491–500. https://doi.org/10.1038/nprot.2013.020.

(15) Baltz, A. G.; Munschauer, M.; Schwanhäusser, B.; Vasile, A.; Murakawa, Y.; Schueler, M.; Youngs, N.; Penfold-Brown, D.; Drew, K.; Milek, M.; et al. The MRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol. Cell* **2012**, *46* (5), 674–690. https://doi.org/10.1016/j.molcel.2012.05.021.

(16) Castello, A.; Fischer, B.; Eichelbaum, K.; Horos, R.; Beckmann, B. M.; Strein, C.; Davey, N. E.; Humphreys, D. T.; Preiss, T.; Steinmetz, L. M.; et al. Insights into RNA Biology from an Atlas of Mammalian MRNA-Binding Proteins. *Cell* **2012**, *149* (6), 1393–1406. https://doi.org/10.1016/j.cell.2012.04.031.

(17) Fernandez-Chamorro, J.; Piñeiro, D.; Gordon, J. M. B.; Ramajo, J.; Francisco-Velilla, R.; Macias, M. J.; Martinez-Salas, E. Identification of Novel Non-Canonical RNA-Binding Sites in Gemin5 Involved in Internal Initiation of Translation. *Nucleic Acids Res.* **2014**, *42* (9), 5742–5754. https://doi.org/10.1093/nar/gku177.

(18) Kwon, S. C.; Yi, H.; Eichelbaum, K.; Föhr, S.; Fischer, B.; You, K. T.; Castello, A.; Krijgsveld, J.; Hentze, M. W.; Kim, V. N. The RNA-Binding Protein Repertoire of Embryonic Stem Cells. *Nat. Struct. Mol. Biol.* **2013**, *20* (9), 1122–1130. https://doi.org/10.1038/nsmb.2638.

(19) Kramer, K.; Sachsenberg, T.; Beckmann, B. M.; Qamar, S.; Boon, K. L.; Hentze, M. W.; Kohlbacher, O.; Urlaub, H. Photo-Cross-Linking and High-Resolution Mass Spectrometry for Assignment of RNA-Binding Sites in RNA-Binding Proteins. *Nat. Methods* **2014**, *11* (10), 1064–1070. https://doi.org/10.1038/nmeth.3092.

(20) He, C.; Sidoli, S.; Warneford-Thomson, R.; Tatomer, D. C.; Wilusz, J. E.; Garcia, B. A.; Bonasio, R. High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Mol. Cell* **2016**, *64* (2), 416–430. https://doi.org/10.1016/j.molcel.2016.09.034.

(21) Eichhorn, C. D.; Yang, Y.; Repeta, L.; Feigon, J. Structural Basis for Recognition of Human 7SK Long Noncoding RNA by the La-Related Protein Larp7. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (28), E6457–E6466. https://doi.org/10.1073/pnas.1806276115.

(22) Phan, A. T.; Kuryavyi, V.; Darnell, J. C.; Serganov, A.; Majumdar, A.; Ilin, S.; Raslin, T.; Polonskaia, A.; Chen, C.; Clain, D.; et al. Structure-Function Studies of FMRP RGG Peptide Recognition of an RNA Duplex-Quadruplex Junction. *Nat. Struct. Mol. Biol.* **2011**, *18* (7), 796–804. https://doi.org/10.1038/nsmb.2064.

(23) Ramos, A.; Grünert, S.; Adams, J.; Micklem, D. R.; Proctor, M. R.; Freund, S.; Bycroft, M.; Johnston, D. S.; Varani, G. RNA Recognition by a Staufen Double-Stranded RNA-Binding Domain. *EMBO J.* **2000**, *19* (5), 997–1009. https://doi.org/10.1093/emboj/19.5.997.

(24) Balachandran, S.; Barber, G. N. PKR in Innate Immunity, Cancer, and Viral Oncolysis. *Methods Mol. Biol.* **2007**, *383*, 277–301. https://doi.org/10.1007/978-1-59745-335-6_18.

(25) Muckenthaler, M. U.; Rivella, S.; Hentze, M. W.; Galy, B. A Red Carpet for Iron Metabolism. *Cell* **2017**, *168* (3), 344–361. https://doi.org/10.1016/j.cell.2016.12.034.

(26) Bousquet-Antonelli, C.; Deragon, J. M. A Comprehensive Analysis of the La-Motif Protein Superfamily. *Rna* **2009**, *15* (5), 750–764. https://doi.org/10.1261/rna.1478709.

(27) Dock-Bregeon, A. C.; Lewis, K. A.; Conte, M. R. The La-Related Proteins: Structures and Interactions of a Versatile Superfamily of RNA-Binding Proteins. *RNA Biol.* **2019**, *00* (00), 1–16. https://doi.org/10.1080/15476286.2019.1695712.

(28) Maraia, R. J.; Mattijssen, S.; Cruz-Gallardo, I.; Conte, M. R. The La and Related RNA-Binding Proteins (LARPs): Structures, Functions, and Evolving Perspectives. *Wiley Interdiscip. Rev. RNA* **2017**, *8* (6). https://doi.org/10.1002/wrna.1430.

(29) Liu, W.; Xie, Y.; Ma, J.; Luo, X.; Nie, P.; Zuo, Z.; Lahrmann, U.; Zhao, Q.; Zheng, Y.; Zhao, Y.; et al. IBS: An Illustrator for the Presentation and Visualization of Biological Sequences. *Bioinformatics* **2015**, *31* (20), 3359–3361. https://doi.org/10.1093/bioinformatics/btv362.

(30) Stefano, J. E. Purified Lupus Antigen La Recognizes an Oligouridylate Stretch Common to the 3′ Termini of RNA Polymerase III Transcripts. *Cell* **1984**, *36* (1), 145–154. https://doi.org/10.1016/0092-8674(84)90083-7.

(31) Yoo, C. J.; Wolin, S. L. The Yeast La Protein Is Required for the 3' Endonucleolytic Cleavage That Matures TRNA Precursors. *Cell* **1997**, *89* (3), 393–402. https://doi.org/10.1016/S0092-8674(00)80220-2.

(32) Bayfield, M. A.; Maraia, R. J. Precursor-Product Discrimination by La Protein during TRNA Metabolism. *Nat. Struct. Mol. Biol.* **2009**, *16* (4), 430–437. https://doi.org/10.1038/nsmb.1573.

(33) Huang, Y.; Bayfield, M. A.; Intine, R. V.; Maraia, R. J. Separate RNA-Binding Surfaces on the Multifunctional La Protein Mediate Distinguishable Activities in TRNA Maturation. *Nat. Struct. Mol. Biol.* **2006**, *13* (7), 611–618. https://doi.org/10.1038/nsmb1110.

(34) Costa-Mattioli, M.; Svitkin, Y.; Sonenberg, N. La Autoantigen Is Necessary for Optimal

Function of the Poliovirus and Hepatitis C Virus Internal Ribosome Entry Site In Vivo and In Vitro. *Mol. Cell. Biol.* **2004**, *24* (15), 6861–6870. https://doi.org/10.1128/mcb.24.15.6861-6870.2004.

(35) Holcik, M.; Korneluk, R. G. Functional Characterization of the X-Linked Inhibitor of Apoptosis (XIAP) Internal Ribosome Entry Site Element: Role of La Autoantigen in XIAP Translation. *Mol. Cell. Biol.* **2000**, *20* (13), 4648–4657. https://doi.org/10.1128/mcb.20.13.4648-4657.2000.

(36) Martino, L.; Pennell, S.; Kelly, G.; Bui, T. T. T.; Kotik-Kogan, O.; Smerdon, S. J.; Drake, A. F.; Curry, S.; Conte, M. R. Analysis of the Interaction with the Hepatitis C Virus MRNA Reveals an Alternative Mode of RNA Recognition by the Human La Protein. *Nucleic Acids Res.* **2012**, *40* (3), 1381–1394. https://doi.org/10.1093/nar/gkr890.

(37) Kuehnert, J.; Sommer, G.; Zierk, A. W.; Fedarovich, A.; Brock, A.; Fedarovich, D.; Heise, T. Novel RNA Chaperone Domain of RNA-Binding Protein La Is Regulated by AKT Phosphorylation. *Nucleic Acids Res.* **2015**, *43* (1), 581–594. https://doi.org/10.1093/nar/gku1309.

(38) Maraia, R. J.; Intine, R. V. La Protein and Its Associated Small Nuclear and Nucleolar Precursor RNAs. *Gene Expr.* **2002**, *10* (1–2), 41–57. https://doi.org/10.0000/096020197390068.

(39) Intine, R. V.; Tenenbaum, S. A.; Sakulich, A. L.; Keene, J. D.; Maraia, R. J. Differential Phosphorylation and Subcellular Localization of La RNPs Associated with Precursor TRNAs and Translation-Related MRNAs. *Mol. Cell* **2003**, *12* (5), 1301–1307. https://doi.org/10.1016/S1097-2765(03)00429-5.

(40) Diribarne, G.; Bensaude, O. 7SK RNA, a Non-Coding RNA Regulating P-TEFb, a General Transcription Factor. *RNA Biol.* **2009**, *6* (2), 122–128. https://doi.org/10.4161/rna.6.2.8115.

(41) Quaresma, A. J. C.; Bugai, A.; Barboric, M. Cracking the Control of RNA Polymerase II Elongation by 7SK SnRNP and P-TEFb. *Nucleic Acids Res.* **2016**, *44* (16), 7527–7539. https://doi.org/10.1093/nar/gkw585.

(42) Muniz, L.; Egloff, S.; Kiss, T. RNA Elements Directing in Vivo Assembly of the 7SK/MePCE/Larp7 Transcriptional Regulatory SnRNP. *Nucleic Acids Res.* **2013**, *41* (8), 4686–

4698. https://doi.org/10.1093/nar/gkt159.

(43) Brown, K. A.; Sharifi, S.; Hussain, R.; Donaldson, L.; Bayfield, M. A.; Wilson, D. J. Distinct Dynamic Modes Enable the Engagement of Dissimilar Ligands in a Promiscuous Atypical RNA Recognition Motif. *Biochemistry* **2016**, *55* (51), 7141–7150. https://doi.org/10.1021/acs.biochem.6b00995.

(44) Lahr, R. M.; Fonseca, B. D.; Ciotti, G. E.; Al-Ashtal, H. A.; Jia, J. J.; Niklaus, M. R.; Blagden, S. P.; Alain, T.; Berman, A. J. La-Related Protein 1 (LARP1) Binds the MRNA Cap, Blocking EIF4F Assembly on TOP MRNAs. *Elife* **2017**, *6*, 1–15. https://doi.org/10.7554/eLife.24146.

(45) Hong, S.; Freeberg, M. A.; Han, T.; Kamath, A.; Yao, Y.; Fukuda, T.; Suzuki, T.; Kim, J. K.; Inoki, K. LARP1 Functions as a Molecular Switch for MTORC1-Mediated Translation of an Essential Class of MRNAs. *Elife* **2017**, *6*, 1–24. https://doi.org/10.7554/eLife.25237.

(46) Cai, L.; Fritz, D.; Stefanovic, L.; Stefanovic, B. Binding of LARP6 to the Conserved 5′ Stem-Loop Regulates Translation of MRNAs Encoding Type I Collagen. *J. Mol. Biol.* **2010**, *395* (2), 309–326. https://doi.org/10.1016/j.jmb.2009.11.020.

(47) Stefanovic, B. RNA Protein Interactions Governing Expression of the Most Abundant Protein in Human Body, Type I Collagen. *Wiley Interdiscip. Rev. RNA* **2013**, *4* (5), 535–545. https://doi.org/10.1002/wrna.1177.

(48) Treiber, T.; Treiber, N.; Plessmann, U.; Harlander, S.; Daiß, J. L.; Eichner, N.; Lehmann, G.; Schall, K.; Urlaub, H.; Meister, G. A Compendium of RNA-Binding Proteins That Regulate MicroRNA Biogenesis. *Mol. Cell* **2017**, *66* (2), 270-284.e13. https://doi.org/10.1016/j.molcel.2017.03.014.

(49) Dermit, M.; Dodel, M.; Lee, F. C. Y.; Azman, M. S.; Schwenzer, H.; Jones, J. L.; Blagden, S. P.; Ule, J.; Mardakheh, F. K. Subcellular MRNA Localization Regulates Ribosome Biogenesis in Migrating Cells. *Dev. Cell* **2020**, *55* (3), 298-313.e10. https://doi.org/10.1016/j.devcel.2020.10.006.

(50) Jolma, A.; Zhang, J.; Mondragón, E.; Morgunova, E.; Kivioja, T.; Laverty, K. U.; Yin, Y.; Zhu, F.; Bourenkov, G.; Morris, Q.; et al. Binding Specificities of Human RNA-Binding Proteins

toward Structured and Linear RNA Sequences. *Genome Res.* **2020**, *30* (7), 962–973. https://doi.org/10.1101/gr.258848.119.

(51) Vukmirovic, M.; Manojlovic, Z.; Stefanovic, B. Serine-Threonine Kinase Receptor-Associated Protein (STRAP) Regulates Translation of Type I Collagen MRNAs. *Mol. Cell. Biol.* **2013**, *33* (19), 3893–3906. https://doi.org/10.1128/mcb.00195-13.

(52) Maraia, R. J.; Mattijssen, S.; Cruz-Gallardo, I.; Conte, M. R. The La and Related RNA-Binding Proteins (LARPs): Structures, Functions, and Evolving Perspectives. *Wiley Interdiscip. Rev. RNA* **2017**, *8* (6). https://doi.org/10.1002/wrna.1430.

(53) Yang, R.; Gaidamakov, S. A.; Xie, J.; Lee, J.; Martino, L.; Kozlov, G.; Crawford, A. K.; Russo, A. N.; Conte, M. R.; Gehring, K.; et al. 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. *Mol. Cell. Biol.* **2011**, *31* (3), 542–556. https://doi.org/10.1128/mcb.01162-10.

(54) Küspert, M.; Murakawa, Y.; Schäffler, K.; Vanselow, J. T.; Wolf, E.; Juranek, S.; Schlosser, A.; Landthaler, M.; Fischer, U. LARP4B Is an AU-Rich Sequence Associated Factor That Promotes MRNA Accumulation and Translation. *Rna* **2015**, *21* (7), 1294–1305. https://doi.org/10.1261/rna.051441.115.

(55) Mattijssen, S.; Arimbasseri, A. G.; Iben, J. R.; Gaidamakov, S.; Lee, J.; Hafner, M.; Maraia, R. J. LARP4 MRNA Codon-TRNA Match Contributes to LARP4 Activity for Ribosomal Protein MRNA Poly(A) Tail Length Protection. *Elife* **2017**, *6*, 1–33. https://doi.org/10.7554/eLife.28889.

(56) Cruz-Gallardo, I.; Martino, L.; Kelly, G.; Andrew Atkinson, R.; Trotta, R.; De Tito, S.; Coleman, P.; Ahdash, Z.; Gu, Y.; Bui, T. T. T.; et al. 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 Res.* **2019**, *47* (8), 4272–4291. https://doi.org/10.1093/nar/gkz144.

(57) Nussbacher, J. K.; Yeo, G. W. Systematic Discovery of RNA Binding Proteins That Regulate MicroRNA Levels. *Mol. Cell* **2018**, *69* (6), 1005-1016.e7. https://doi.org/10.1016/j.molcel.2018.02.012.

(58) Seetharaman, S.; Flemyng, E.; Shen, J.; Conte, M. R.; Ridley, A. J. The RNA-Binding Protein LARP4 Regulates Cancer Cell Migration and Invasion. *Cytoskeleton* **2016**, *73* (11), 680–690. https://doi.org/10.1002/cm.21336.

(59) Koso, H.; Yi, H.; Sheridan, P.; Miyano, S.; Ino, Y.; Todo, T.; Watanabe, S. Identification of RNA-Binding Protein LARP4B as a Tumor Suppressor in Glioma. *Cancer Res.* **2016**, *76* (8), 2254–2264. https://doi.org/10.1158/0008-5472.CAN-15-2308.

(60) Zhang, Y.; Peng, L.; Hu, T.; Wan, Y.; Ren, Y.; Zhang, J.; Wang, X.; Zhou, Y.; Yuan, W.; Wang, Q.; et al. La-Related Protein 4B Maintains Murine MLL-AF9 Leukemia Stem Cell Self-Renewal by Regulating Cell Cycle Progression. *Exp. Hematol.* **2015**, *43* (4), 309-318.e2. https://doi.org/10.1016/j.exphem.2014.12.003.

(61) Li, Y.; Jiao, Y.; Li, Y.; Liu, Y. Expression of La Ribonucleoprotein Domain Family Member 4B (LARP4B) in Liver Cancer and Their Clinical and Prognostic Significance. *Dis. Markers* **2019**, *2019*. https://doi.org/10.1155/2019/1569049.

(62) Teichmann, M.; Dumay-Odelot, H.; Fribourg, S. Structural and Functional Aspects of Winged-Helix Domains at the Core of Transcription Initiation Complexes. *Transcription* **2012**, *3* (1), 2–7. https://doi.org/10.4161/trns.3.1.18917.

(63) Alfano, C.; Sanfelice, D.; Babon, J.; Kelly, G.; Jacks, A.; Curry, S.; Conte, M. R. Structural Analysis of Cooperative RNA Binding by the La Motif and Central RRM Domain of Human La Protein. *Nat. Struct. Mol. Biol.* **2004**, *11* (4), 323–329. https://doi.org/10.1038/nsmb747.

(64) Dong, G.; Chakshusmathi, G.; Wolin, S. L.; Reinisch, K. M. Structure of the La Motif: A Winged Helix Domain Mediates RNA Binding via a Conserved Aromatic Patch. *EMBO J.* **2004**, *23* (5), 1000–1007. https://doi.org/10.1038/sj.emboj.7600115.

(65) Martino, L.; Pennell, S.; Kelly, G.; Busi, B.; Brown, P.; Atkinson, R. A.; Salisbury, N. J. H.; Ooi, Z. H.; See, K. W.; Smerdon, S. J.; et al. 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 Res.* **2015**, *43* (1), 645–660. https://doi.org/10.1093/nar/gku1287.

(66) Merret, R.; Martino, L.; Bousquet-Antonelli, C.; Fneich, S.; Descombin, J.; Billey, É.; Conte, M. R.; Deragon, J. M. 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* **2013**, *19* (1), 36–50. https://doi.org/10.1261/rna.035469.112.

(67) Kotik-Kogan, O.; Valentine, E. R.; Sanfelice, D.; Conte, M. R.; Curry, S. Structural Analysis Reveals Conformational Plasticity in the Recognition of RNA 3′ Ends by the Human La Protein. *Structure* **2008**, *16* (6), 852–862. https://doi.org/10.1016/j.str.2008.02.021.

(68) Cléry, A.; Blatter, M.; Allain, F. H. T. RNA Recognition Motifs: Boring? Not Quite. *Curr. Opin. Struct. Biol.* **2008**, *18* (3), 290–298. https://doi.org/10.1016/j.sbi.2008.04.002.

(69) Teplova, M.; Yuan, Y. R.; Phan, A. T.; Malinina, L.; Ilin, S.; Teplov, A.; Patel, D. J. Structural Basis for Recognition and Sequestration of UUUOH 3′ Temini of Nascent RNA Polymerase III Transcripts by La, a Rheumatic Disease Autoantigen. *Mol. Cell* **2006**, *21* (1), 75–85. https://doi.org/10.1016/j.molcel.2005.10.027.

(70) Gajiwala, K. S.; Burley, S. K. Winged Helix Proteins. *Curr. Opin. Struct. Biol.* **2000**, *10* (1), 110–116. https://doi.org/10.1016/S0959-440X(99)00057-3.

(71) Curry, S.; Kotik-Kogan, O.; Conte, M. R.; Brick, P. Getting to the End of RNA: Structural Analysis of Protein Recognition of 5′ and 3′ Termini. *Biochim. Biophys. Acta - Gene Regul. Mech.* **2009**, *1789* (9–10), 653–666. https://doi.org/10.1016/j.bbagrm.2009.07.003.

(72) Cruz-Gallardo, I.; Martino, L.; Trotta, R.; De Tito, S.; Kelly, G.; Atkinson, R. A.; Randazzo, A.; Conte, M. R. Resonance Assignment of Human LARP4A La Module. *Biomol. NMR Assign.* **2019**, *13* (1), 169–172. https://doi.org/10.1007/s12104-019-09871-4.

(73) Sanfelice, D.; Kelly, G.; Curry, S.; Conte, M. R. NMR Assignment of the N-Terminal Region of Human La Free and in Complex with RNA. *Biomol. NMR Assign.* **2008**, *2* (2), 107–109. https://doi.org/10.1007/s12104-008-9097-5.

(74) Schäffler, K.; Schulz, K.; Hirmer, A.; Wiesner, J.; Grimm, M.; Sickmann, A.; Fischer, U. A Stimulatory Role for the La-Related Protein 4B in Translation. *Rna* **2010**, *16* (8), 1488–1499. https://doi.org/10.1261/rna.2146910.

(75) Xie, J.; Kozlov, G.; Gehring, K. The “Tale” of Poly(A) Binding Protein: The MLLE Domain and PAM2-Containing Proteins. *Biochim. Biophys. Acta - Gene Regul. Mech.* **2014**, *1839* (11), 1062–1068. https://doi.org/10.1016/j.bbagrm.2014.08.001.

(76) Kozlov, G.; Ménade, M.; Rosenauer, A.; Nguyen, L.; Gehring, K. Molecular Determinants of PAM2 Recognition by the MLLE Domain of Poly(A)-Binding Protein. *J. Mol. Biol.* **2010**, *397* (2), 397–407. https://doi.org/10.1016/j.jmb.2010.01.032.

(77) Kota, V.; Sommer, G.; Durette, C.; Thibault, P.; Van Niekerk, E. A.; Twiss, J. L.; Heise, T. SUMO-Modification of the La Protein Facilitates Binding to MRNA in Vitro and in Cells. *PLoS One* **2016**, *11* (5), 1–19. https://doi.org/10.1371/journal.pone.0156365.

(78) Plasmid Mini-Prep Kit. *4037* (949), 7749.

(79) Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: The Proteomics Server for in-Depth Protein Knowledge and Analysis. *Nucleic Acids Res.* **2003**, *31* (13), 3784–3788. https://doi.org/10.1093/nar/gkg563.

(80) Madeira, F.; Park, Y. M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A. R. N.; Potter, S. C.; Finn, R. D.; et al. The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47* (W1), W636–W641. https://doi.org/10.1093/nar/gkz268.

(81) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* **1995**, *6* (3), 277–293. https://doi.org/10.1007/BF00197809.

(82) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins Struct. Funct. Genet.* **2005**, *59* (4), 687–696. https://doi.org/10.1002/prot.20449.

(83) Wilfinger, W. W.; Mackey, K.; Chomczynski, P. Effect of PH and Ionic Strength on the Spectrophotometric Assessment of Nucleic Acid Purity. *Biotechniques* **1997**, *22* (3), 474–481. https://doi.org/10.2144/97223st01.

(84) Wishart, D. S.; Sykes, B. D.; Richards, F. M. The Chemical Shift Index : A Fast and Simple Method for the Assignment of Protein Secondary Structure Through NMR Spectroscopy. *Biochemistry* **1992**, *31* (6), 1647–1651. https://doi.org/10.1021/bi00121a010.

(85) Wishart, D. S.; Sykes, B. D. The 13C Chemical-Shift Index: A Simple Method for the Identification of Protein Secondary Structure Using 13C Chemical-Shift Data. *J. Biomol. NMR* **1994**, *4* (2), 171–180. https://doi.org/10.1007/BF00175245.

(86) Speraf, S.; Bax, A. Empirical Correlation between Protein Backbone Conformation and Cα and Cβ 13C Nuclear Magnetic Resonance Chemical Shifts. *J. Am. Chem. Soc.* **1991**, *113* (14), 5490–5492.

(87) Nagai, K.; Oubridge, C.; Jessen, T. H.; Li, J.; Evans, P. R. Crystal Structure of the RNA-Binding Domain of the U1 Small Nuclear Ribonucleoprotein A. *Nature* **1990**, *348* (6301), 515–520. https://doi.org/10.1038/348515a0.

A.

APPENDIX

Table A.1 RNA ratio values for pure sample

$A_{260/280}$	$A_{260/230}$
1.9-2.2	2.0-2.2

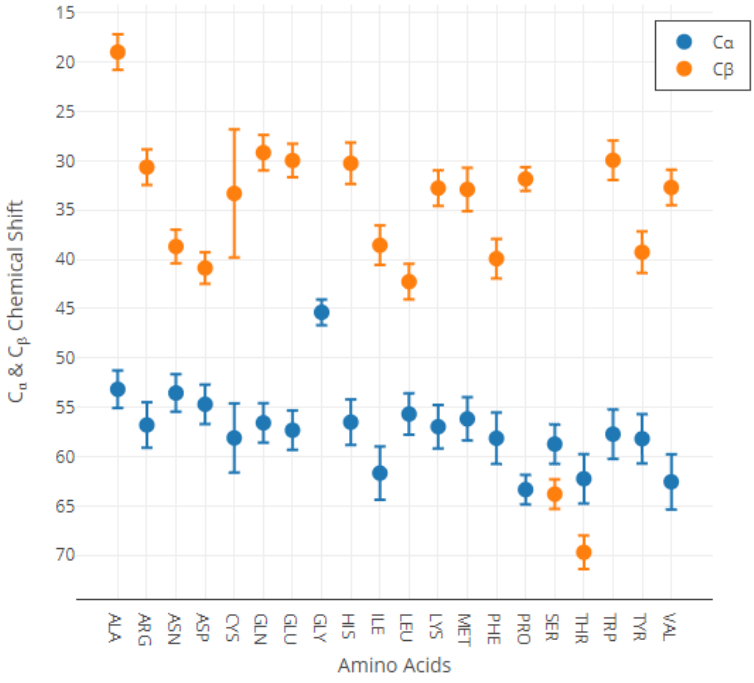

Figure A.1 Cα and Cβ carbon chemical shift distribution for all the 20 amino acids

```
########################################
# Program: needle
# Rundate: Thu 25 Nov 2021 23:58:35
# Commandline: needle
#    -auto
#    -stdout
#    -asequence emboss_needle-I20211125-235517-0219-4458777-p1m.asequence
#    -bsequence emboss_needle-I20211125-235517-0219-4458777-p1m.bsequence
#    -datafile EBLOSUM62
#    -gapopen 10.0
#    -gapextend 0.5
#    -endopen 10.0
#    -endextend 0.5
#    -aformat3 pair
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 83
# Identity:      60/83 (72.3%)
# Similarity:    71/83 (85.5%)
# Gaps:           0/83 ( 0.0%)
# Score: 306.0
#
#
#=======================================

EMBOSS_001         1 VSTEDLKECLKKQLEFCFSRENLSKDLYLISQMDSDQFIPIWTVANMEEI     50
                     .|.||.:|.|||.||||.|||||:.|:||||||||||::||.||||::.|
EMBOSS_001         1 DSQEDPREVLKKTLEFCLSRENLASDMYLISQMDSDQYVPITTVANLDHI     50

EMBOSS_001        51 KKLTTDPDLILEVLRSSPMVQVDEKGEKVRPSH     83
                     |||:||.|||:|||||.|:||||||||||||:.
EMBOSS_001        51 KKLSTDVDLIVEVLRSLPLVQVDEKGEKVRPNQ     83


#---------------------------------------
#---------------------------------------
```

Figure A.2 La motif sequence alignment between LaRP4A and LaRP4B with Clustal Omega

```
########################################
# Program: needle
# Rundate: Fri 26 Nov 2021 00:29:48
# Commandline: needle
#    -auto
#    -stdout
#    -asequence emboss_needle-I20211126-002946-0675-82178650-p2m.asequence
#    -bsequence emboss_needle-I20211126-002946-0675-82178650-p2m.bsequence
#    -datafile EBLOSUM62
#    -gapopen 10.0
#    -gapextend 0.5
#    -endopen 10.0
#    -endextend 0.5
#    -aformat3 pair
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 95
# Identity:      66/95 (69.5%)
# Similarity:    79/95 (83.2%)
# Gaps:           4/95 ( 4.2%)
# Score: 353.5
#
#
#=======================================

EMBOSS_001         1 KRCIVILREIPETTPIEEVKGLFKSENCPKVISCEFAHNSNWYITFQSDT     50
                     .|||||||||.|:||:|||:.|||.:|.||.|:||||:|.||:|||:::.
EMBOSS_001         1 NRCIVILREISESTPVEEVEALFKGDNLPKFINCEFAYNDNWFITFETEA     50

EMBOSS_001        51 DAQQAFKYLREEVKTFQGKPIMARIK----AINTFFAKNGYRLMD     91
                     |||||:|||||||||||||.||||    |||||..|||:|.:|
EMBOSS_001        51 DAQQAYKYLREEVKTFQGKPIKARIKAKAIAINTFLPKNGFRPLD     95


#---------------------------------------
#---------------------------------------
```

Figure A.3 RRM sequence alignment between LaRP4A and LaRP4B with Clustal Omega

LaRP4B LaM Peak List

N	R	A	Shift
151	D	C	176.281
151	D	CA	54.606
151	D	CB	41.357
152	S	HN	8.370
152	S	N	115.921
152	S	C	174.542
152	S	CA	58.489
152	S	CB	63.800
153	Q	HN	8.448
153	Q	N	122.030
153	Q	C	175.750
153	Q	CA	55.977
153	Q	CB	29.536
154	E	HN	8.327
154	E	N	121.969
154	E	HA	4.263
154	E	C	175.725
154	E	CA	56.228
154	E	CB	30.842
155	D	HN	8.615
155	D	N	124.393
155	D	HA	4.886
155	D	CA	51.865
155	D	CB	42.077
156	P	HA	4.387
156	P	C	178.840
156	P	CA	65.114
156	P	CB	32.365
157	R	HN	8.502
157	R	N	118.998
157	R	HA	3.922
157	R	C	178.115
157	R	CA	59.783
157	R	CB	29.827
158	E	HN	7.752
158	E	N	119.404
158	E	HA	4.137
158	E	C	179.057
158	E	CA	58.889
158	E	CB	29.319
159	V	HN	7.934
159	V	N	119.874
159	V	HA	3.729
159	V	C	179.926
159	V	CA	66.256
159	V	CB	31.604
160	L	HN	8.261
160	L	N	124.080

160 L HA 4.016
160 L C 177.560
160 L CA 58.767
160 L CB 42.012
161 K HN 8.366
161 K N 119.839
161 K HA 3.676
161 K C 177.681
161 K CA 61.052
161 K CB 32.619
162 K HN 8.143
162 K N 116.912
162 K HA 4.181
162 K C 179.733
162 K CA 59.021
162 K CB 32.365
163 T HN 8.196
163 T N 117.448
163 T HA 4.122
163 T C 177.005
163 T CA 67.170
163 T CB 68.483
164 L HN 8.730
164 L N 124.176
164 L HA 3.900
164 L C 177.391
164 L CA 58.513
164 L CB 41.758
165 E HN 8.433
165 E N 118.214
165 E HA 3.696
165 E C 178.284
165 E CA 59.783
165 E CB 29.827
166 F HN 7.605
166 F N 116.402
166 F HA 4.356
166 F C 178.816
166 F CA 61.306
166 F CB 39.220
167 C HN 8.462
167 C N 119.036
167 C HA 4.017
167 C C 174.808
167 C CA 62.321
167 C CB 26.780
168 L HN 7.386
168 L N 114.190
168 L HA 4.405
168 L C 175.070

168 L CA 53.182
168 L CB 40.743
169 S HN 7.297
169 S N 115.741
169 S HA 4.339
169 S C 174.977
169 S CA 57.918
169 S CB 64.657
170 R HN 9.032
170 R N 123.307
170 R HA 3.852
170 R C 179.695
170 R CA 59.783
170 R CB 29.573
171 E HN 8.830
171 E N 117.204
171 E HA 4.012
171 E C 178.526
171 E CA 59.783
171 E CB 28.811
172 N HN 7.540
172 N N 117.833
172 N HA 4.481
172 N C 177.753
172 N CA 56.482
172 N CB 39.727
173 L HN 8.473
173 L N 119.465
173 L HA 4.043
173 L C 177.319
173 L CA 57.498
173 L CB 41.504
174 A HN 7.390
174 A N 116.147
174 A HA 4.165
174 A C 178.550
174 A CA 53.690
174 A CB 18.656
175 S HN 7.403
175 S N 109.443
175 S HA 4.599
175 S C 173.601
175 S CA 58.006
175 S CB 64.860
176 D HN 7.825
176 D N 123.701
176 D HA 4.762
176 D C 174.929
176 D CA 53.944
176 D CB 40.743

177 M HN 8.335
177 M N 122.193
177 M HA 4.266
177 M C 178.936
177 M CA 57.752
177 M CB 32.111
178 Y HN 8.436
178 Y N 120.698
178 Y HA 4.370
178 Y C 178.622
178 Y CA 61.306
178 Y CB 37.696
179 L HN 8.457
179 L N 121.651
179 L HA 3.613
179 L C 179.516
179 L CA 58.513
179 L CB 41.250
180 I HN 8.333
180 I N 117.429
180 I HA 3.918
180 I C 179.685
180 I CA 64.606
180 I CB 37.950
181 S HN 7.711
181 S N 115.108
181 S HA 4.334
181 S C 174.736
181 S CA 60.888
181 S CB 63.401
182 Q HN 7.376
182 Q N 119.335
182 Q HA 4.274
182 Q C 174.977
182 Q CA 54.451
182 Q CB 29.573
183 M HN 7.250
183 M N 119.860
183 M HA 4.668
183 M C 177.077
183 M CA 56.228
183 M CB 33.719
184 D HN 8.666
184 D N 122.946
184 D HA 4.823
184 D C 179.081
184 D CA 52.420
184 D CB 41.514
185 S HN 8.474
185 S N 113.640

185 S HA 4.212
185 S C 174.736
185 S CA 61.306
185 S CB 62.944
186 D HN 8.459
186 D N 122.268
186 D HA 4.919
186 D C 174.202
186 D CA 54.451
186 D CB 42.266
187 Q HN 7.988
187 Q N 109.503
187 Q HA 3.806
187 Q C 173.370
187 Q CA 58.513
187 Q CB 26.019
188 Y HN 8.220
188 Y N 114.768
188 Y HA 4.661
188 Y C 177.319
188 Y CA 59.275
188 Y CB 39.727
189 V HN 9.392
189 V N 118.965
189 V HA 4.916
189 V CA 57.752
189 V CB 35.075
190 P HA 4.747
190 P C 178.767
190 P CA 63.083
190 P CB 32.111
191 I HN 9.010
191 I N 127.602
191 I HA 3.493
191 I C 177.777
191 I CA 65.621
191 I CB 36.935
192 T HN 8.568
192 T N 111.308
192 T HA 3.841
192 T C 176.474
192 T CA 65.368
192 T CB 67.906
193 T HN 7.175
193 T N 118.029
193 T HA 4.040
193 T C 176.281
193 T CA 66.383
193 T CB 67.906
194 V HN 7.065

194 V N 121.716
194 V HA 3.485
194 V C 176.546
194 V CA 65.875
194 V CB 32.111
195 A HN 8.705
195 A N 117.614
195 A HA 3.898
195 A C 176.957
195 A CA 53.944
195 A CB 18.403
196 N HN 7.114
196 N N 110.304
196 N HA 4.677
196 N C 175.822
196 N CA 52.928
196 N CB 39.727
197 L HN 7.440
197 L N 122.238
197 L HA 4.274
197 L C 179.516
197 L CA 55.467
197 L CB 41.758
198 D HN 9.054
198 D N 124.755
198 D HA 4.180
198 D C 177.753
198 D CA 58.513
198 D CB 41.014
199 H HN 8.425
199 H N 112.249
199 H HA 4.360
199 H C 175.725
199 H CA 57.752
199 H CB 30.080
200 I HN 6.974
200 I N 118.792
200 I HA 3.938
200 I C 177.536
200 I CA 61.306
200 I CB 35.412
201 K HN 8.416
201 K N 121.558
201 K HA 4.203
201 K C 177.729
201 K CA 58.767
201 K CB 32.111
202 K HN 7.117
202 K N 113.765
202 K HA 4.023

202 K C 176.933
202 K CA 58.006
202 K CB 32.619
203 L HN 7.706
203 L N 118.734
203 L HA 4.347
203 L C 177.922
203 L CA 55.975
203 L CB 43.281
204 S HN 8.087
204 S N 111.496
204 S HA 4.473
204 S C 173.974
204 S CA 59.021
204 S CB 65.621
205 T HN 8.424
205 T N 115.032
205 T HA 4.451
205 T C 173.504
205 T CA 60.544
205 T CB 68.922
206 D HN 8.514
206 D N 124.224
206 D HA 4.672
206 D C 175.339
206 D CA 52.674
206 D CB 40.997
207 V HN 8.333
207 V N 125.911
207 V HA 3.486
207 V C 176.640
207 V CA 66.383
207 V CB 31.604
208 D HN 7.819
208 D N 118.503
208 D HA 4.359
208 D C 179.178
208 D CA 57.752
208 D CB 39.981
209 L HN 7.439
209 L N 123.735
209 L HA 4.161
209 L C 177.995
209 L CA 57.498
209 L CB 40.997
210 I HN 7.830
210 I N 119.509
210 I HA 3.313
210 I C 177.633
210 I CA 66.637

210 I CB 37.588
211 V HN 8.596
211 V N 119.335
211 V HA 3.398
211 V C 177.077
211 V CA 68.160
211 V CB 31.857
212 E HN 7.827
212 E N 120.053
212 E HA 3.908
212 E C 180.081
212 E CA 59.783
212 E CB 29.573
213 V HN 8.772
213 V N 119.694
213 V HA 3.595
213 V C 180.409
213 V CA 66.383
213 V CB 30.842
214 L HN 8.862
214 L N 122.389
214 L HA 3.906
214 L C 179.081
214 L CA 58.767
214 L CB 42.012
215 R HN 8.455
215 R N 116.280
215 R HA 3.950
215 R C 176.643
215 R CA 59.783
215 R CB 30.080
216 S HN 7.745
216 S N 113.944
216 S HA 4.604
216 S C 173.987
216 S CA 58.513
216 S CB 64.860
217 L HN 7.466
217 L N 126.702
217 L HA 4.738
217 L CA 52.420
217 L CB 42.012
218 P HA 4.549
218 P C 177.826
218 P CA 64.352
218 P CB 32.111
219 L HN 7.848
219 L N 115.561
219 L HA 4.689
219 L C 177.174

219 L CA 54.705
219 L CB 40.900
220 V HN 7.495
220 V N 114.520
220 V HA 4.887
220 V C 173.701
220 V CA 59.289
220 V CB 34.904
221 Q N 122.198
221 Q HN 9.207
221 Q HA 4.791
221 Q C 174.204
221 Q CA 54.451
221 Q CB 31.604
222 V HN 9.073
222 V N 128.012
222 V HA 4.957
222 V C 176.570
222 V CA 60.798
222 V CB 33.127
223 D HN 8.669
223 D N 127.795
223 D HA 4.563
223 D C 176.546
223 D CA 54.198
223 D CB 40.997
224 E HN 8.875
224 E N 121.132
224 E HA 4.019
224 E C 178.381
224 E CA 59.783
224 E CB 29.573
225 K HN 7.670
225 K N 114.488
225 K HA 4.198
225 K C 176.860
225 K CA 56.482
225 K CB 32.619
226 G HN 7.932
226 G N 109.465
226 G HA2 3.842
226 G HA3 3.353
226 G C 173.299
226 G CA 46.582
227 E HN 9.534
227 E N 117.348
227 E HA 4.407
227 E C 177.077
227 E CA 56.482
227 E CB 33.381

228 K HN 8.603
228 K N 121.956
228 K HA 5.198
228 K C 172.925
228 K CA 55.467
228 K CB 36.681
229 V HN 9.304
229 V N 118.120
229 V HA 5.778
229 V C 171.525
229 V CA 58.604
229 V CB 35.412
230 R HN 8.997
230 R N 124.417
230 R HA 5.332
230 R CA 52.420
230 R CB 32.873
231 P HA 4.673
231 P C 175.880
231 P CA 62.575
231 P CB 31.857
232 N HN 8.062
232 N N 120.456
232 N HA 4.537
232 N C 174.349
232 N CA 53.182
232 N CB 38.458
233 Q HN 7.858
233 Q N 125.042
233 Q HA 4.106
233 Q CA 57.244
233 Q CB 30.588

LaRP4B RRM Peak List

N	R	H	N	CA	CB
238	V	9,3416	126,1993	61,9183	33,5715
239	I	9,2340	127,9409	60,0104	40,1184
240	L	9,1615	127,5174	53,3204	44,3858
241	R	8,3733	120,2699	54,7012	32,9225
242	E	8,0626	113,7136	57,3081	0,0000
243	I	8,4474	117,1116	53,5419	38,6707
244	S	7,9238	115,5201	58,6723	63,8154
245	E	8,7849	127,2511	58,6513	29,3234
246	S	7,7707	110,6859	58,4892	63,0503
247	T	7,8247	124,0839	62,3056	69,9486
248	P	0,0000	0,0000	0,0000	0,0000
249	V	8,7627	126,4623	65,9278	31,8230
250	E	9,3728	117,7220	59,5356	28,4713
251	E	7,5896	118,9979	58,8724	29,6054
252	V	6,7984	119,6744	65,9651	30,9173
253	E	8,4397	116,7640	59,7546	30,3624
254	A	7,3611	117,1821	54,5020	18,2278
255	L	7,2652	117,3192	57,0036	41,7454
256	F	7,4028	111,7590	57,6775	39,8123
257	K	7,1634	119,9892	56,0049	33,3209
258	G	7,9228	110,2619	45,0372	0,0000
259	D	7,9894	119,9324	57,7555	39,5782
260	N	0,0000	0,0000	0,0000	0,0000
261	L	0,0000	0,0000	0,0000	0,0000
262	P	0,0000	0,0000	0,0000	0,0000
263	K	8,5944	121,6861	55,8546	32,2152
264	F	6,9067	120,5546	55,1285	38,4281
265	I	8,8600	117,7824	62,1833	38,4645
266	N	7,1933	113,9480	52,6416	41,2786
267	C	8,4220	123,8754	57,9803	27,3677
268	E	8,7338	124,1810	54,3477	33,9665
269	F	8,3334	125,8458	58,5815	38,1952
270	A	7,3923	131,7913	52,0161	18,3903
271	Y	0,0000	0,0000	0,0000	0,0000
272	N	0,0000	0,0000	0,0000	0,0000
273	D	0,0000	0,0000	0,0000	0,0000
274	N	7,6230	113,6375	52,1404	41,1968
275	W	9,4151	122,0789	56,5174	30,3624
276	F	9,1830	120,1562	53,1992	40,0185
277	I	9,5350	129,2664	61,3949	38,2745
278	T	0,0000	0,0000	0,0000	0,0000
279	F	9,0949	124,5737	57,6921	42,6949
280	E	10,1138	119,8689	59,5208	31,1445
281	T	6,9948	131,2776	58,3654	73,5503
282	E	8,9101	122,7201	59,1171	29,2058
283	A	8,2993	121,7201	59,1171	29,2058
284	D	7,8185	119,7585	57,0845	40,7449
285	A	7,3737	121,4478	54,7139	18,4503
286	Q	8,5998	115,9320	58,9434	27,3778
287	Q	7,8714	120,5903	59,0013	28,1738
288	A	7,9132	122,6629	55,3939	19,4237
289	Y	8,6891	118,5599	61,2940	39,2955
290	K	8,0596	119,1730	59,7528	32,5835
291	Y	7,7775	119,7987	61,2816	38,5179
292	L	8,4323	119,9150	57,7360	42,2727
293	R	7,7003	114,8484	58,5490	30,4290
294	E	8,4701	113,9779	57,5922	30,8107
295	E	7,8751	116,1723	56,9606	30,2485

296	V	7,1598	120,7474	66,3074	31,6739
297	K	0,0000	0,0000	0,0000	0,0000
298	T	9,0217	115,3912	59,8216	72,4303
299	F	8,8694	119,4724	57,1301	41,4026
300	Q	8,9944	125,4524	56,4138	24,9091
301	G	8,4920	102,9398	45,2379	0,0000
302	K	7,5301	120,2540	53,2191	33,5610
303	P	0,0000	0,0000	0,0000	0,0000
304	I	0,0000	0,0000	0,0000	0,0000
305	K	0,0000	0,0000	0,0000	0,0000
306	A	0,0000	0,0000	0,0000	0,0000
307	R	0,0000	0,0000	0,0000	0,0000
308	I	0,0000	0,0000	0,0000	0,0000
309	K	9,6860	132,4122	55,2430	33,4579
310	A	8,3083	124,2090	49,8696	23,0777
311	K	8,8765	118,5442	54,8015	34,1864
312	A	0,0000	0,0000	0,0000	0,0000
313	I	7,9272	119,7825	60,7131	38,8516
314	A	8,2489	128,0931	52,1741	19,2491
315	I	8,0487	120,0087	61,3369	38,6696
316	N	0,0000	0,0000	0,0000	0,0000
317	T	0,0000	0,0000	0,0000	0,0000
318	F	8,0476	121,7997	57,5259	39,4629

LaRP4A La Module Peak List

N	R	A	Shift
112	SER	CA	58,655
112	SER	CB	64,003
113	ALA	N	125,998
113	ALA	H	8,293
113	ALA	CA	52,959
113	ALA	HA	4,313
113	ALA	CB	19,261
113	ALA	HB1	1,332
113	ALA	HB2	1,332
113	ALA	HB3	1,332
114	VAL	N	119,343
114	VAL	H	8,009
114	VAL	CA	62,564
114	VAL	HA	4,118
114	VAL	CB	33,363
114	VAL	HB	2,037
114	VAL	CG1	20,543
114	VAL	HG1	0,9
114	VAL	HG1	0,9
114	VAL	HG1	0,9
115	SER	N	120,647
115	SER	H	8,604
115	SER	CA	58,051
115	SER	CB	64,491
116	THR	N	117,537
116	THR	H	8,444
116	THR	CA	64,807
116	THR	HA	4,087
116	THR	CB	68,652
116	THR	HB	4,054
116	THR	CG2	22,145
116	THR	HG2	1,194
116	THR	HG2	1,194
116	THR	HG2	1,194
117	GLU	N	121,469
117	GLU	H	8,461
117	GLU	CA	59,038
117	GLU	HA	3,861
117	GLU	CB	29,517
117	GLU	HB2	2,059
117	GLU	HB3	2,118
117	GLU	CG	35,927
117	GLU	HG2	2,012
117	GLU	HG3	2,232
118	ASP	N	120,885
118	ASP	H	8,041
118	ASP	CA	56,474
118	ASP	HA	4,453
118	ASP	CB	40,734
118	ASP	HB2	2,646
118	ASP	HB3	2,702
119	LEU	N	123,972
119	LEU	H	8,176
119	LEU	CA	58,718
119	LEU	HA	4,028
119	LEU	CB	41,696
119	LEU	HB2	1,686
119	LEU	HB3	1,783

119	LEU	CG	27,273
119	LEU	HG	1,502
119	LEU	CD2	24,068
119	LEU	HD2	0,915
119	LEU	HD2	0,915
119	LEU	HD2	0,915
120	LYS	N	117,417
120	LYS	H	8,594
120	LYS	CA	60,641
120	LYS	HA	3,78
120	LYS	CB	28,555
120	LYS	HB2	1,919
120	LYS	HB3	2,108
120	LYS	CE	37,85
120	LYS	HE2	2,015
120	LYS	HE3	2,657
121	GLU	N	119,086
121	GLU	H	7,803
121	GLU	CA	59,359
121	GLU	HA	4,004
121	GLU	CB	29,196
121	GLU	HB2	1,914
121	GLU	HB3	1,982
122	CYS	N	119,205
122	CYS	H	8,155
122	CYS	CA	63,525
122	CYS	HA	4,065
122	CYS	CB	26,953
122	CYS	HB2	2,893
122	CYS	HB3	3,089
123	LEU	N	121,589
123	LEU	H	8,793
123	LEU	CA	58,077
123	LEU	HA	3,895
123	LEU	CB	42,657
123	LEU	HB2	1,653
123	LEU	HB3	1,709
123	LEU	CD1	22,466
123	LEU	HD1	0,781
123	LEU	HD1	0,781
123	LEU	HD1	0,781
123	LEU	CD2	26,312
123	LEU	HD2	0,775
123	LEU	HD2	0,775
123	LEU	HD2	0,775
124	LYS	N	120,874
124	LYS	H	8,553
124	LYS	CA	61,282
124	LYS	HA	3,52
124	LYS	CB	32,722
124	LYS	HB2	1,81
124	LYS	HB3	1,932
124	LYS	CG	24,709
124	LYS	HG2	1,186
124	LYS	HG3	1,227
124	LYS	CD	28,235
124	LYS	HD2	1,971
124	LYS	HD3	2,133
125	LYS	N	117,179
125	LYS	H	7,901

125	LYS	CA	59,679
125	LYS	HA	4,066
125	LYS	CB	32,401
125	LYS	HB3	1,877
125	LYS	CG	25,35
125	LYS	HG2	1,413
125	LYS	HG3	1,584
125	LYS	CD	35,606
125	LYS	HD2	2,291
126	GLN	N	118,728
126	GLN	H	7,887
126	GLN	CA	58,397
126	GLN	HA	4,138
126	GLN	CB	28,555
126	GLN	HB2	1,924
126	GLN	HB3	2,114
126	GLN	CG	32,722
126	GLN	HG2	1,805
126	GLN	HG3	1,924
127	LEU	N	120,635
127	LEU	H	8,815
127	LEU	CA	58,718
127	LEU	HA	4,033
127	LEU	CB	41,696
127	LEU	HB2	1,188
127	LEU	HB3	1,933
127	LEU	CG	27,273
127	LEU	HG	1,75
127	LEU	CD1	21,825
127	LEU	HD1	0,777
127	LEU	HD1	0,777
127	LEU	HD1	0,777
127	LEU	CD2	23,748
127	LEU	HD2	0,788
127	LEU	HD2	0,788
127	LEU	HD2	0,788
128	GLU	N	117,775
128	GLU	H	8,265
128	GLU	CA	60,961
128	GLU	HA	3,712
128	GLU	CB	32,401
128	GLU	HB2	1,776
128	GLU	HB3	1,887
128	GLU	CG	35,927
128	GLU	HG2	2,225
128	GLU	HG3	2,289
129	PHE	N	119,086
129	PHE	H	7,791
129	PHE	CA	62,243
129	PHE	HA	4,247
129	PHE	CB	38,811
129	PHE	HB2	3,225
129	PHE	HD1	7,234
129	PHE	HE1	7,163
130	CYS	N	118,848
130	CYS	H	8,34
130	CYS	CA	63,205
130	CYS	HA	3,788
130	CYS	CB	26,953
130	CYS	HB2	2,53

130	CYS	HB3	2,908
131	PHE	N	111,101
131	PHE	H	7,396
131	PHE	CA	59,038
131	PHE	HA	4,209
131	PHE	CB	38,811
131	PHE	HB2	2,453
131	PHE	HB3	3,193
131	PHE	CD1	131,909
131	PHE	HD1	7,128
131	PHE	CE1	132,346
131	PHE	HE1	7,115
132	SER	CA	58,463
133	ARG	N	124,33
133	ARG	H	8,913
133	ARG	CA	59,861
133	ARG	CB	29,771
134	GLU	N	118,013
134	GLU	H	8,963
134	GLU	CA	60
134	GLU	HA	3,855
134	GLU	CB	28,876
134	GLU	HB2	1,803
134	GLU	HB3	1,924
134	GLU	CG	36,568
134	GLU	HG2	2,142
134	GLU	HG3	2,226
135	ASN	N	116,822
135	ASN	H	7,348
135	ASN	CA	56,795
135	ASN	HA	4,394
135	ASN	CB	39,452
135	ASN	HB2	1,881
135	ASN	HB3	2,376
135	ASN	ND2	116,596
135	ASN	HD2	6,885
135	ASN	HD2	7,765
136	LEU	N	120,17
136	LEU	H	8,515
136	LEU	CA	58,077
136	LEU	HA	3,931
136	LEU	CB	41,375
136	LEU	HB2	1,275
136	LEU	HB3	1,605
136	LEU	CG	26,632
136	LEU	HG	1,6
136	LEU	CD1	24,389
136	LEU	HD1	0,577
136	LEU	HD1	0,577
136	LEU	HD1	0,577
136	LEU	CD2	25,671
136	LEU	HD2	0,616
136	LEU	HD2	0,616
136	LEU	HD2	0,616
137	SER	N	108,36
137	SER	H	7,473
137	SER	CA	60,961
137	SER	HA	4,13
137	SER	CB	64,166
137	SER	HB2	3,851

138	LYS	N	118,252
138	LYS	H	7,188
138	LYS	CA	56,474
138	LYS	HA	4,413
138	LYS	CB	34,324
138	LYS	HB2	1,775
138	LYS	HB3	1,837
138	LYS	CG	24,389
138	LYS	HG2	1,388
138	LYS	HG3	1,408
138	LYS	CD	32,722
138	LYS	HD2	1,669
138	LYS	HD3	1,713
139	ASP	N	121,946
139	ASP	H	7,79
139	ASP	CA	53,91
139	ASP	HA	4,588
139	ASP	CB	40,414
139	ASP	HB2	2,461
139	ASP	HB3	2,963
140	LEU	N	123,376
140	LEU	H	7,955
140	LEU	CA	57,756
140	LEU	HA	3,956
140	LEU	CB	41,055
140	LEU	HB2	1,537
140	LEU	HB3	1,72
140	LEU	CG	27,273
140	LEU	HG	1,685
140	LEU	CD1	22,786
140	LEU	HD1	0,793
140	LEU	HD1	0,793
140	LEU	HD1	0,793
140	LEU	CD2	25,03
140	LEU	HD2	0,888
140	LEU	HD2	0,888
140	LEU	HD2	0,888
141	TYR	N	120,516
141	TYR	H	8,254
141	TYR	CA	61,602
141	TYR	HA	4,177
141	TYR	CB	37,85
141	TYR	HB2	3,052
141	TYR	HB3	3,127
141	TYR	CD1	133,878
141	TYR	HD1	7,014
141	TYR	CE1	118,562
141	TYR	HE1	6,641
142	LEU	N	121,361
142	LEU	H	8,161
142	LEU	CA	58,718
142	LEU	HA	3,514
142	LEU	CB	41,375
142	LEU	HB2	1,062
142	LEU	HB3	1,583
142	LEU	CG	26,953
142	LEU	HG	1,612
142	LEU	CD1	23,107
142	LEU	HD1	0,563
142	LEU	HD1	0,563

142	LEU	HD1	0,563
142	LEU	CD2	25,671
142	LEU	HD2	0,439
142	LEU	HD2	0,439
142	LEU	HD2	0,439
143	ILE	N	116,406
143	ILE	H	8,107
143	ILE	CA	64,807
143	ILE	HA	3,781
143	ILE	CB	37,85
143	ILE	HB	1,743
143	ILE	CG2	17,658
143	ILE	HG2	0,828
143	ILE	HG2	0,828
143	ILE	HG2	0,828
143	ILE	CG1	28,555
143	ILE	HG1	1,227
143	ILE	HG1	1,636
143	ILE	CD1	13,171
143	ILE	HD1	0,687
143	ILE	HD1	0,687
143	ILE	HD1	0,687
144	SER	N	115,034
144	SER	H	7,768
144	SER	CA	61,282
144	SER	HA	4,203
144	SER	CB	63,846
144	SER	HB2	3,938
144	SER	HB3	4,016
145	GLN	N	119,205
145	GLN	H	7,283
145	GLN	CA	54,872
145	GLN	HA	4,195
145	GLN	CB	29,517
145	GLN	HB2	1,642
145	GLN	HB3	2,293
145	GLN	CG	33,042
145	GLN	HG2	1,683
145	GLN	HG3	1,974
146	MET	N	119,086
146	MET	H	7,169
146	MET	CA	56,474
146	MET	HA	4,418
146	MET	CB	34,965
146	MET	HB2	1,983
146	MET	HB3	2,023
146	MET	CG	33,683
146	MET	HG2	2,287
146	MET	HG3	2,786
146	MET	CE	18,299
146	MET	HE1	1,92
146	MET	HE2	1,92
146	MET	HE3	1,92
147	ASP	N	123,138
147	ASP	H	8,641
147	ASP	CA	52,628
147	ASP	HA	4,745
147	ASP	CB	41,375
147	ASP	HB2	2,828
147	ASP	HB3	3,378

148	SER	N	113,604
148	SER	H	8,398
148	SER	CA	61,923
148	SER	HA	4,125
148	SER	CB	63,205
148	SER	HB2	3,864
148	SER	HB3	3,938
149	ASP	N	122,185
149	ASP	H	8,445
149	ASP	CA	54,872
149	ASP	HA	4,851
149	ASP	CB	42,657
149	ASP	HB2	2,738
149	ASP	HB3	2,882
150	GLN	N	109,194
150	GLN	H	7,946
150	GLN	CA	59,359
150	GLN	HA	3,73
150	GLN	CB	26,312
150	GLN	HB2	2,329
150	GLN	HB3	2,438
150	GLN	CG	34,965
150	GLN	HG3	2,213
150	GLN	HG2	2,007
151	PHE	N	115,153
151	PHE	H	8,284
151	PHE	CA	59,679
151	PHE	HA	4,474
151	PHE	CB	40,414
151	PHE	HB3	2,576
151	PHE	HB2	3,513
151	PHE	HD1	7,101
152	ILE	N	121,838
152	ILE	H	9,41
152	ILE	CA	58,397
152	ILE	HA	3,898
152	ILE	CB	42,657
152	ILE	HB	1,685
152	ILE	CG2	17,338
152	ILE	HG2	0,84
152	ILE	HG2	0,84
152	ILE	HG2	0,84
152	ILE	CG1	27,914
152	ILE	HG1	0,741
152	ILE	HG1	1,685
152	ILE	CD1	15,094
152	ILE	HD1	0,745
152	ILE	HD1	0,745
152	ILE	HD1	0,745
153	PRO	CD	51,987
153	PRO	CA	63,525
153	PRO	HA	4,592
153	PRO	CB	32,081
153	PRO	HB2	1,497
153	PRO	HB3	2,383
153	PRO	CG	28,555
153	PRO	HG2	1,893
153	PRO	HG3	2,084
153	PRO	HD2	3,432
153	PRO	HD3	4,158

154	ILE	N	126,841
154	ILE	H	8,804
154	ILE	CA	66,089
154	ILE	HA	3,439
154	ILE	CB	37,209
154	ILE	HB	1,57
154	ILE	CG2	17,017
154	ILE	HG2	0,709
154	ILE	HG2	0,709
154	ILE	HG2	0,709
154	ILE	CG1	30,799
154	ILE	HG1	0,702
154	ILE	HG1	1,295
154	ILE	CD1	13,492
154	ILE	HD1	0,558
154	ILE	HD1	0,558
154	ILE	HD1	0,558
155	TRP	N	118,371
155	TRP	H	8,271
155	TRP	CA	60,32
155	TRP	HA	4,295
155	TRP	CB	28,235
155	TRP	HB2	3,162
155	TRP	HB3	3,384
155	TRP	CD1	128,627
155	TRP	CE3	121,625
155	TRP	NE1	130,198
155	TRP	HD1	7,373
155	TRP	HE3	7,425
155	TRP	CZ3	122,5
155	TRP	CZ2	115,499
155	TRP	HE1	10,29
155	TRP	HZ3	7,083
155	TRP	CH2	125,126
155	TRP	HZ2	7,439
155	TRP	HH2	7,172
156	THR	N	115,749
156	THR	H	6,897
156	THR	CA	66,41
156	THR	HA	3,717
156	THR	CB	68,653
156	THR	HB	4,013
156	THR	CG2	22,145
156	THR	HG2	0,769
156	THR	HG2	0,769
156	THR	HG2	0,769
157	VAL	N	122,542
157	VAL	H	7,061
157	VAL	CA	66,089
157	VAL	HA	3,432
157	VAL	CB	31,76
157	VAL	HB	2,218
157	VAL	CG1	21,825
157	VAL	HG1	0,885
157	VAL	HG1	0,885
157	VAL	HG1	0,885
157	VAL	CG2	22,786
157	VAL	HG2	0,956
157	VAL	HG2	0,956
157	VAL	HG2	0,956

158	ALA	N	118,49
158	ALA	H	8,388
158	ALA	CA	54,231
158	ALA	HA	3,839
158	ALA	CB	18,62
158	ALA	HB1	1,289
158	ALA	HB2	1,289
158	ALA	HB3	1,289
159	ASN	N	110,386
159	ASN	H	6,999
159	ASN	CA	53,269
159	ASN	HA	4,505
159	ASN	CB	39,773
159	ASN	HB2	2,424
159	ASN	HB3	2,735
160	MET	N	121,231
160	MET	H	7,479
160	MET	CA	57,436
160	MET	HA	4,109
160	MET	CB	33,363
160	MET	HB2	2,132
160	MET	HB3	2,29
160	MET	CG	32,401
160	MET	HG2	2,493
160	MET	HG3	3,106
160	MET	CE	17,338
160	MET	HE1	2,105
160	MET	HE2	2,105
160	MET	HE3	2,105
161	GLU	N	126,117
161	GLU	H	9,007
161	GLU	CA	60,961
161	GLU	HA	3,717
161	GLU	CB	29,837
161	GLU	HB2	1,882
161	GLU	HB3	1,984
161	GLU	CG	35,927
161	GLU	HG2	2,212
162	GLU	N	114,438
162	GLU	H	9,498
162	GLU	CA	59,359
162	GLU	HA	3,857
162	GLU	CB	29,196
162	GLU	HB2	1,734
162	GLU	HB3	1,877
162	GLU	CG	36,568
162	GLU	HG2	2,213
163	ILE	N	116,702
163	ILE	H	7,127
163	ILE	CA	61,282
163	ILE	HA	3,906
163	ILE	CB	35,286
163	ILE	HB	2,287
163	ILE	CG2	18,299
163	ILE	HG2	0,812
163	ILE	HG2	0,812
163	ILE	HG2	0,812
163	ILE	CG1	27,914
163	ILE	HG1	1,422
163	ILE	HG1	1,477

163	ILE	CD1	9,966
163	ILE	HD1	0,543
163	ILE	HD1	0,543
163	ILE	HD1	0,543
164	LYS	N	120,635
164	LYS	H	8,276
164	LYS	CA	59,038
164	LYS	HA	4,21
164	LYS	CB	32,081
164	LYS	HB2	1,644
164	LYS	HB3	1,704
164	LYS	CG	26,953
164	LYS	HG2	1,553
164	LYS	HG3	1,603
164	LYS	CD	26,953
164	LYS	HD2	1,393
165	LYS	N	114,319
165	LYS	H	7,208
165	LYS	CA	58,397
165	LYS	HA	3,925
165	LYS	CB	32,401
165	LYS	HB2	1,653
165	LYS	HB3	1,71
165	LYS	CG	26,953
165	LYS	HG2	1,681
165	LYS	CD	29,196
165	LYS	HD2	1,599
166	LEU	N	118,848
166	LEU	H	7,617
166	LEU	CA	56,474
166	LEU	HA	4,283
166	LEU	CB	43,298
166	LEU	HB2	1,181
166	LEU	HB3	2,089
166	LEU	CG	26,632
166	LEU	HG	1,677
166	LEU	CD1	22,786
166	LEU	HD1	0,786
166	LEU	HD1	0,786
166	LEU	HD1	0,786
166	LEU	CD2	26,312
166	LEU	HD2	0,825
166	LEU	HD2	0,825
166	LEU	HD2	0,825
167	THR	N	111,816
167	THR	H	8,132
167	THR	CA	61,923
167	THR	HA	4,525
167	THR	CB	68,306
167	THR	HB	4,201
167	THR	CG2	19,902
167	THR	HG2	0,916
167	THR	HG2	0,916
167	THR	HG2	0,916
168	THR	N	113,842
168	THR	H	7,994
168	THR	CA	60,641
168	THR	HA	4,429
168	THR	CB	69,294
168	THR	HB	4,565

168	THR	CG2	21,504
168	THR	HG2	1,151
168	THR	HG2	1,151
168	THR	HG2	1,151
169	ASP	N	126,833
169	ASP	H	8,657
169	ASP	CA	51,667
169	ASP	HA	4,871
169	ASP	CB	41,696
169	ASP	HB2	2,458
169	ASP	HB3	3,027
170	PRO	CD	51,346
170	PRO	CA	65,128
170	PRO	HA	4,084
170	PRO	CB	32,081
170	PRO	HB2	1,99
170	PRO	HB3	2,193
170	PRO	CG	26,953
170	PRO	HG2	1,829
170	PRO	HG3	2,015
170	PRO	HD2	3,935
170	PRO	HD3	3,977
171	ASP	N	119,217
171	ASP	H	7,94
171	ASP	CA	57,759
171	ASP	HA	4,395
171	ASP	CB	40,093
171	ASP	HB2	2,562
171	ASP	HB3	2,794
172	LEU	N	123,972
172	LEU	H	7,291
172	LEU	CA	57,756
172	LEU	HA	4,135
172	LEU	CB	41,696
172	LEU	HB2	1,507
172	LEU	HB3	1,747
172	LEU	CG	27,273
172	LEU	HG	1,375
172	LEU	CD1	21,825
172	LEU	HD1	0,773
172	LEU	HD1	0,773
172	LEU	HD1	0,773
172	LEU	CD2	24,068
172	LEU	HD2	0,844
172	LEU	HD2	0,844
172	LEU	HD2	0,844
173	ILE	N	117,417
173	ILE	H	7,289
173	ILE	CA	66,089
173	ILE	HA	3,285
173	ILE	CB	37,85
173	ILE	HB	1,723
173	ILE	CG2	16,697
173	ILE	HG2	0,684
173	ILE	HG2	0,684
173	ILE	HG2	0,684
173	ILE	CG1	30,158
173	ILE	HG1	0,701
173	ILE	HG1	1,632
173	ILE	CD1	13,812

173	ILE	HD1	0,643
173	ILE	HD1	0,643
173	ILE	HD1	0,643
174	LEU	N	119,086
174	LEU	H	7,805
174	LEU	CA	58,718
174	LEU	HA	3,743
174	LEU	CB	42,016
174	LEU	HB2	1,023
174	LEU	HB3	1,938
174	LEU	CD1	24,389
174	LEU	HD1	0,847
174	LEU	HD1	0,847
174	LEU	HD1	0,847
174	LEU	CD2	27,273
174	LEU	HD2	0,657
174	LEU	HD2	0,657
174	LEU	HD2	0,657
175	GLU	N	118,848
175	GLU	H	7,859
175	GLU	CA	59,679
175	GLU	HA	3,838
175	GLU	CB	29,644
176	VAL	N	118,555
176	VAL	H	8,142
176	VAL	CA	66,73
176	VAL	HA	3,581
176	VAL	CB	31,119
176	VAL	HB	1,926
176	VAL	CG1	21,825
176	VAL	HG1	0,774
176	VAL	HG1	0,774
176	VAL	HG1	0,774
176	VAL	CG2	23,427
176	VAL	HG2	1,009
176	VAL	HG2	1,009
176	VAL	HG2	1,009
177	LEU	N	121,6
177	LEU	H	8,591
177	LEU	CA	58,718
177	LEU	HA	3,742
177	LEU	CB	41,696
177	LEU	HB2	1,025
177	LEU	HB3	1,938
177	LEU	CG	25,991
177	LEU	HG	1,879
177	LEU	CD1	24,068
177	LEU	HD1	0,691
177	LEU	HD1	0,691
177	LEU	HD1	0,691
178	ARG	N	115,63
178	ARG	H	8,398
178	ARG	CA	60
178	ARG	HA	3,77
178	ARG	CB	29,837
178	ARG	HB2	1,484
178	ARG	HB3	1,805
179	SER	N	112,65
179	SER	H	7,457
179	SER	CA	58,397

179	SER	HA	4,516
179	SER	CB	64,487
179	SER	HB2	3,891
179	SER	HB3	4,03
180	SER	N	117,795
180	SER	H	7,301
180	SER	CA	55,192
180	SER	HA	4,886
180	SER	CB	64,487
180	SER	HB2	3,706
180	SER	HB3	3,869
181	PRO	CD	51,667
181	PRO	CA	64,807
181	PRO	HA	4,087
181	PRO	CB	32,081
181	PRO	HB2	1,993
181	PRO	HB3	2,193
181	PRO	CG	27,273
181	PRO	HG2	2
181	PRO	HG3	2,039
181	PRO	HD2	3,986
181	PRO	HD3	4,332
182	MET	N	116,345
182	MET	H	8,005
182	MET	CA	56,154
182	MET	HA	4,342
182	MET	CB	32,55
182	MET	HB3	2,132
182	MET	HB2	1,834
182	MET	CG	32,401
182	MET	HG2	2,385
182	MET	HG3	2,477
182	MET	CE	27,273
182	MET	HE1	2,017
182	MET	HE2	2,017
182	MET	HE3	2,017
183	VAL	N	110,386
183	VAL	H	7,342
183	VAL	CA	58,718
183	VAL	HA	4,825
183	VAL	CB	35,446
183	VAL	HB	1,609
183	VAL	CG1	20,863
183	VAL	HG1	-0,034
183	VAL	HG1	-0,034
183	VAL	HG1	-0,034
183	VAL	CG2	18,94
183	VAL	HG2	0,543
183	VAL	HG2	0,543
183	VAL	HG2	0,543
184	GLN	N	120,159
184	GLN	H	8,901
184	GLN	CA	54,677
184	GLN	CB	31,769
184	GLN	HB2	1,587
184	GLN	HG2	1,745
185	VAL	N	127,794
185	VAL	H	9,08
185	VAL	CA	61,075
185	VAL	HA	4,766

185	VAL	CB	33,042
185	VAL	HB	1,928
185	VAL	CG1	22,786
185	VAL	HG1	0,737
185	VAL	HG1	0,737
185	VAL	HG1	0,737
185	VAL	CG2	21,825
185	VAL	HG2	1,164
185	VAL	HG2	1,164
185	VAL	HG2	1,164
186	ASP	N	127,309
186	ASP	H	8,599
186	ASP	CA	54,231
186	ASP	HA	4,411
186	ASP	CB	41,055
186	ASP	HB2	1,925
186	ASP	HB3	2,702
187	GLU	N	120,754
187	GLU	H	8,767
187	GLU	CA	60
187	GLU	HA	3,838
187	GLU	CB	29,517
187	GLU	HB2	1,937
187	GLU	HB3	1,979
187	GLU	CG	35,927
187	GLU	HG2	2,005
187	GLU	HG3	2,245
188	LYS	N	114,676
188	LYS	H	7,569
188	LYS	CA	57,115
188	LYS	HA	4,107
188	LYS	CB	32,401
188	LYS	HB2	1,66
188	LYS	HB3	1,711
188	LYS	CG	25,671
188	LYS	HG2	1,286
188	LYS	HG3	1,4
189	GLY	N	109,449
189	GLY	H	7,858
189	GLY	CA	46,823
189	GLY	HA3	3,329
189	GLY	HA2	3,752
190	GLU	N	117,549
190	GLU	H	9,352
190	GLU	CA	57,115
190	GLU	HA	4,365
190	GLU	CB	33,363
190	GLU	HB2	1,836
190	GLU	HB3	2,016
190	GLU	CG	36,888
190	GLU	HG2	2,076
190	GLU	HG3	2,123
191	LYS	N	120,754
191	LYS	H	8,398
191	LYS	CA	55,513
191	LYS	HA	5,144
191	LYS	CB	36,963
191	LYS	HB2	1,264
191	LYS	HB3	1,628
191	LYS	CG	24,709

191	LYS	HG2	1,045
191	LYS	CD	29,196
191	LYS	HD2	1,211
191	LYS	HD3	1,29
191	LYS	CE	42,016
191	LYS	HE2	2,367
191	LYS	HE3	2,477
192	VAL	N	117,788
192	VAL	H	9,384
192	VAL	CA	59,038
192	VAL	HA	5,484
192	VAL	CB	35,286
192	VAL	HB	1,706
192	VAL	CG1	20,222
192	VAL	HG1	0,782
192	VAL	HG1	0,782
192	VAL	HG1	0,782
192	VAL	CG2	22,466
192	VAL	HG2	0,537
192	VAL	HG2	0,537
192	VAL	HG2	0,537
193	ARG	N	124,459
193	ARG	H	8,806
193	ARG	CA	52,308
193	ARG	HA	5,319
193	ARG	CB	33,042
193	ARG	HB2	0,572
193	ARG	HB3	1,194
194	PRO	CD	50,705
194	PRO	CA	62,564
194	PRO	HA	4,446
194	PRO	CB	32,081
194	PRO	HB2	2,257
194	PRO	CG	27,914
194	PRO	HG2	1,726
194	PRO	HG3	1,978
194	PRO	HD2	3,48
194	PRO	HD3	3,872
195	SER	CA	58,077
195	SER	HA	4,132
195	SER	CB	64,166
195	SER	HB2	3,376
195	SER	HB3	3,467
199	CYS	CA	59,359
199	CYS	HA	4,615
199	CYS	CB	28,876
199	CYS	HB3	2,465
199	CYS	HB2	2,725
200	ILE	N	123,496
200	ILE	H	8,263
200	ILE	CA	60,329
200	ILE	HA	4,893
200	ILE	CB	41,375
200	ILE	HB	1,603
200	ILE	CG2	17,979
200	ILE	HG2	0,517
200	ILE	HG2	0,517
200	ILE	HG2	0,517
200	ILE	CG1	27,594
200	ILE	HG1	0,966

200	ILE	HG1	1,213
200	ILE	CD1	14,133
200	ILE	HD1	0,348
200	ILE	HD1	0,348
200	ILE	HD1	0,348
201	VAL	N	125,045
201	VAL	H	9,248
201	VAL	CA	62,243
201	VAL	HA	4,527
201	VAL	CB	34,324
201	VAL	HB	2,134
201	VAL	CG1	21,504
201	VAL	HG1	0,962
201	VAL	HG1	0,962
201	VAL	HG1	0,962
201	VAL	CG2	23,748
201	VAL	HG2	1,035
201	VAL	HG2	1,035
201	VAL	HG2	1,035
202	ILE	N	127,794
202	ILE	H	9,274
202	ILE	CA	60,32
202	ILE	HA	5,095
202	ILE	CB	40,414
202	ILE	HB	1,6
202	ILE	CG2	17,338
202	ILE	HG2	0,831
202	ILE	HG2	0,831
202	ILE	HG2	0,831
202	ILE	CG1	27,594
202	ILE	HG1	0,72
202	ILE	HG1	1,375
202	ILE	CD1	14,453
202	ILE	HD1	0,625
202	ILE	HD1	0,625
202	ILE	HD1	0,625
203	LEU	N	126,786
203	LEU	H	9,233
203	LEU	CA	53,59
203	LEU	HA	5,134
203	LEU	CB	44,259
203	LEU	HB2	1,51
203	LEU	HB3	1,702
203	LEU	CG	26,953
203	LEU	HG	1,699
203	LEU	CD2	24,389
203	LEU	HD2	0,785
203	LEU	HD2	0,785
203	LEU	HD2	0,785
203	LEU	CD1	25,991
203	LEU	HD1	0,726
203	LEU	HD1	0,726
203	LEU	HD1	0,726
204	ARG	N	120,275
204	ARG	H	8,405
204	ARG	CA	55,192
204	ARG	HA	5,245
204	ARG	CB	33,363
204	ARG	HB2	1,702
204	ARG	HB3	1,937

204	ARG	CG	27,594
204	ARG	HG2	1,508
204	ARG	HG3	1,707
204	ARG	CD	43,939
204	ARG	HD2	3,03
204	ARG	HD3	3,151
205	GLU	N	112,412
205	GLU	H	8,323
205	GLU	CA	57,756
205	GLU	HA	3,782
205	GLU	CB	26,632
205	GLU	HB2	2,234
205	GLU	CG	36,888
205	GLU	HG2	2,145
205	GLU	HG3	2,256
206	ILE	N	120,923
206	ILE	H	8,402
206	ILE	CA	58,397
206	ILE	HA	4,658
206	ILE	CB	39,452
206	ILE	HB	1,691
206	ILE	CG2	17,338
206	ILE	HG2	0,452
206	ILE	HG2	0,452
206	ILE	HG2	0,452
206	ILE	CG1	26,632
206	ILE	HG1	0,934
206	ILE	HG1	1,389
206	ILE	CD1	13,171
206	ILE	HD1	0,345
206	ILE	HD1	0,345
206	ILE	HD1	0,345
207	PRO	CD	51,667
207	PRO	CA	63,525
207	PRO	HA	4,409
207	PRO	CB	32,722
207	PRO	HB2	1,806
207	PRO	HB3	2,494
207	PRO	HD2	3,621
207	PRO	HD3	4,093
208	GLU	N	124,661
208	GLU	H	9,292
208	GLU	CA	59,039
208	GLU	CB	29,397
209	THR	N	104,069
209	THR	H	6,952
209	THR	CA	61,923
209	THR	HA	4,04
209	THR	CB	68,974
209	THR	HB	4,441
209	THR	CG2	22,145
209	THR	HG2	1,129
209	THR	HG2	1,129
209	THR	HG2	1,129
210	THR	N	123,301
210	THR	H	7,819
210	THR	CA	61,602
210	THR	HA	4,064
210	THR	CB	70,576
210	THR	HB	3,872

210	THR	CG2	21,184
210	THR	HG2	1,07
210	THR	HG2	1,07
210	THR	HG2	1,07
211	PRO	CA	63,341
211	PRO	CB	32,645
212	ILE	N	126,603
212	ILE	H	8,796
212	ILE	CA	65,448
212	ILE	HA	3,731
212	ILE	CB	38,491
212	ILE	HB	2,035
212	ILE	CG2	17,338
212	ILE	HG2	1,063
212	ILE	HG2	1,063
212	ILE	HG2	1,063
212	ILE	CG1	30,158
212	ILE	HG1	1,326
212	ILE	HG1	1,805
212	ILE	CD1	14,453
212	ILE	HD1	1,154
212	ILE	HD1	1,154
212	ILE	HD1	1,154
213	GLU	N	118,502
213	GLU	H	9,078
213	GLU	CA	60
213	GLU	HA	3,914
213	GLU	CB	28,876
213	GLU	HB2	1,931
213	GLU	HB3	1,991
213	GLU	CG	36,247
213	GLU	HG2	2,285
213	GLU	HG3	2,32
214	GLU	N	119,205
214	GLU	H	7,603
214	GLU	CA	59,038
214	GLU	HA	4,063
214	GLU	CB	29,837
214	GLU	HB2	2,015
214	GLU	HB3	2,093
214	GLU	CG	36,888
214	GLU	HG2	2,165
214	GLU	HG3	2,274
215	VAL	N	119,694
215	VAL	H	6,827
215	VAL	CA	66,73
215	VAL	HA	3,026
215	VAL	CB	31,119
215	VAL	HB	1,426
215	VAL	CG1	20,863
215	VAL	HG1	-0,488
215	VAL	HG1	-0,488
215	VAL	HG1	-0,488
215	VAL	CG2	21,184
215	VAL	HG2	0,252
215	VAL	HG2	0,252
215	VAL	HG2	0,252
216	LYS	N	116,702
216	LYS	H	8,424
216	LYS	CA	60,641

216	LYS	HA	3,865
216	LYS	CB	32,722
216	LYS	HB2	1,684
216	LYS	HB3	1,784
216	LYS	CG	26,953
216	LYS	HG2	1,264
216	LYS	HG3	1,576
216	LYS	CD	29,837
216	LYS	HD2	1,573
216	LYS	HD3	1,623
216	LYS	CE	42,016
216	LYS	HE2	2,873
217	GLY	N	106,114
217	GLY	H	7,734
217	GLY	CA	46,4
217	GLY	HA3	3,753
217	GLY	HA2	3,874
218	LEU	N	121,123
218	LEU	H	7,271
218	LEU	CA	57,436
218	LEU	HA	3,872
218	LEU	CB	42,016
218	LEU	HB2	0,602
218	LEU	HB3	1,263
218	LEU	CG	25,991
218	LEU	HG	1,369
218	LEU	CD1	24,709
218	LEU	HD1	0,181
218	LEU	HD1	0,181
218	LEU	HD1	0,181
218	LEU	CD2	23,107
218	LEU	HD2	0,474
218	LEU	HD2	0,474
218	LEU	HD2	0,474
219	PHE	N	113,499
219	PHE	H	7,347
219	PHE	CA	57,756
219	PHE	HA	4,566
219	PHE	CB	38,811
219	PHE	HB2	2,721
219	PHE	HB3	3,765
219	PHE	CD1	133,878
219	PHE	HD1	7,632
219	PHE	CE1	131,471
219	PHE	HE1	7,141
219	PHE	CZ	129,502
219	PHE	HZ	7,095
220	LYS	N	121,6
220	LYS	H	7,097
220	LYS	CA	56,795
220	LYS	HA	4,279
220	LYS	CB	31,44
220	LYS	HB2	1,705
220	LYS	HB3	1,932
220	LYS	CG	25,03
220	LYS	HG2	1,334
220	LYS	HG3	1,413
220	LYS	CD	29,517
220	LYS	HD2	1,671
220	LYS	CE	42,016

220	LYS	HE2	2,96
221	SER	N	118,728
221	SER	H	7,639
221	SER	CA	58,077
221	SER	HA	4,396
221	SER	CB	64,487
221	SER	HB2	2,973
221	SER	HB3	4,118
222	GLU	N	126,719
222	GLU	H	9,067
222	GLU	CA	58,631
222	GLU	CB	29,279
223	ASN	N	115,643
223	ASN	H	8,399
223	ASN	CA	53,269
223	ASN	HA	4,618
223	ASN	CB	38,811
223	ASN	HB2	2,56
223	ASN	HB3	2,956
224	CYS	N	120,647
224	CYS	H	7,229
224	CYS	CA	55,833
224	CYS	HA	4,552
224	CYS	CB	28,876
224	CYS	HB2	0,837
224	CYS	HB3	2,66
225	PRO	CA	62,922
225	PRO	CB	33,028
226	LYS	N	119,694
226	LYS	H	8,275
226	LYS	CA	57,115
226	LYS	HA	4,025
226	LYS	CB	33,042
226	LYS	HB2	1,711
226	LYS	HB3	1,805
226	LYS	CG	24,709
226	LYS	HG2	1,305
226	LYS	HG3	1,403
226	LYS	CD	29,196
226	LYS	HD2	1,607
227	VAL	N	123,615
227	VAL	H	8,016
227	VAL	CA	62,564
227	VAL	HA	3,107
227	VAL	CB	33,042
227	VAL	HB	1,601
227	VAL	CG1	22,145
227	VAL	HG1	0,302
227	VAL	HG1	0,302
227	VAL	HG1	0,302
227	VAL	CG2	22,466
227	VAL	HG2	0,431
227	VAL	HG2	0,431
227	VAL	HG2	0,431
228	ILE	N	126,117
228	ILE	H	8,358
228	ILE	CA	61,602
228	ILE	HA	3,972
228	ILE	CB	36,568
228	ILE	HB	1,513

228	ILE	CG2	17,017
228	ILE	HG2	0,729
228	ILE	HG2	0,729
228	ILE	HG2	0,729
228	ILE	CG1	27,273
228	ILE	HG1	1,083
228	ILE	HG1	1,155
228	ILE	CD1	10,607
228	ILE	HD1	0,584
228	ILE	HD1	0,584
228	ILE	HD1	0,584
229	SER	N	111,355
229	SER	H	7,05
229	SER	CA	57,756
229	SER	HA	4,485
229	SER	CB	65,769
229	SER	HB2	3,651
229	SER	HB3	3,739
230	CYS	N	123,019
230	CYS	H	8,508
230	CYS	CA	58,077
230	CYS	HA	5,106
230	CYS	CB	28,555
230	CYS	HB2	2,437
230	CYS	HB3	2,681
231	GLU	N	124,211
231	GLU	H	8,719
231	GLU	CA	55,192
231	GLU	HA	4,789
231	GLU	CB	34,324
231	GLU	HB2	1,829
231	GLU	HB3	1,928
231	GLU	CG	35,927
231	GLU	HG2	2,008
231	GLU	HG3	2,162
232	PHE	N	126,237
232	PHE	H	8,607
232	PHE	CA	58,397
232	PHE	HA	3,377
232	PHE	CB	38,491
232	PHE	HB2	1,332
232	PHE	HB3	2,038
232	PHE	CD1	132,565
232	PHE	HD1	5,894
232	PHE	CE1	131,033
232	PHE	HE1	6,874
232	PHE	CZ	129,502
232	PHE	HZ	7,154
233	ALA	N	131,361
233	ALA	H	7,658
233	ALA	CA	51,987
233	ALA	HA	3,926
233	ALA	CB	19,261
233	ALA	HB1	0,138
233	ALA	HB2	0,138
233	ALA	HB3	0,138
235	ASN	CA	55,192
235	ASN	HA	4,353
235	ASN	CB	37,529
235	ASN	HB2	2,682

235	ASN	HB3	2,856
236	SER	N	108,496
236	SER	H	7,942
236	SER	CA	60
236	SER	HA	4,308
236	SER	CB	62,564
236	SER	HB2	3,865
236	SER	HB3	4,062
237	ASN	N	115,868
237	ASN	H	7,502
237	ASN	CA	51,938
237	ASN	HA	5,822
237	ASN	CB	41,055
237	ASN	HB2	2,147
237	ASN	HB3	2,546
238	TRP	N	120,874
238	TRP	H	9,226
238	TRP	CA	56,8
238	TRP	HA	4,692
238	TRP	CB	31,119
238	TRP	HB2	2,934
238	TRP	HB3	3,04
238	TRP	CD1	127,314
238	TRP	CE3	119,875
238	TRP	NE1	130,577
238	TRP	HD1	7,275
238	TRP	HE3	7,173
238	TRP	CZ3	121,82
238	TRP	CZ2	115,718
238	TRP	HE1	9,42
238	TRP	HZ3	6,898
238	TRP	CH2	125,564
238	TRP	HZ2	7,44
238	TRP	HH2	7,151
239	TYR	N	121,123
239	TYR	H	9,332
239	TYR	CA	54,551
239	TYR	HA	5,179
239	TYR	CB	38,811
239	TYR	HB2	2,635
239	TYR	HB3	3,045
239	TYR	CD1	131,909
239	TYR	HD1	6,792
239	TYR	CE1	118,562
239	TYR	HE1	6,529
240	ILE	N	129,756
240	ILE	H	9,672
240	ILE	CA	59,038
240	ILE	HA	4,7
240	ILE	CB	37,529
240	ILE	HB	2,118
240	ILE	CG2	20,863
240	ILE	HG2	0,065
240	ILE	HG2	0,065
240	ILE	HG2	0,065
240	ILE	CG1	27,594
240	ILE	HG1	1,362
240	ILE	HG1	1,655
240	ILE	CD1	12,851
240	ILE	HD1	0,813

240	ILE	HD1	0,813
240	ILE	HD1	0,813
241	THR	N	121,361
241	THR	H	8,415
241	THR	CA	63,505
241	THR	HA	4,531
241	THR	CB	69,935
241	THR	HB	3,855
241	THR	CG2	21,825
241	THR	HG2	1,039
241	THR	HG2	1,039
241	THR	HG2	1,039
242	PHE	N	123,029
242	PHE	H	8,842
242	PHE	CA	58,077
242	PHE	HA	4,593
242	PHE	CB	42,977
242	PHE	HB2	2,653
242	PHE	HB3	3,489
242	PHE	CD1	132,565
242	PHE	HD1	7,182
243	GLN	N	118,967
243	GLN	H	10,118
243	GLN	CA	58,397
243	GLN	HA	4,09
243	GLN	CB	29,517
243	GLN	HB2	2,067
243	GLN	HB3	2,23
243	GLN	CG	34,004
243	GLN	HG2	2,27
243	GLN	HG3	2,364
244	SER	N	107,543
244	SER	H	7,406
244	SER	CA	56,474
244	SER	HA	4,891
244	SER	CB	67,371
244	SER	HB2	3,858
244	SER	HB3	4,16
245	ASP	N	122,822
245	ASP	H	9,12
245	ASP	CA	57,756
245	ASP	HA	4,389
245	ASP	CB	40,093
245	ASP	HB2	2,667
246	THR	N	115,391
246	THR	H	8,073
246	THR	CA	66,41
246	THR	HA	3,873
246	THR	CB	68,653
246	THR	HB	4,05
246	THR	CG2	22,145
246	THR	HG2	1,2
246	THR	HG2	1,2
246	THR	HG2	1,2
247	ASP	N	123,982
247	ASP	H	7,839
247	ASP	CA	57,756
247	ASP	HA	4,253
247	ASP	CB	40,734
247	ASP	HB2	2,546

247	ASP	HB3	2,854
248	ALA	N	121,78
248	ALA	H	7,737
248	ALA	CA	55,513
248	ALA	HA	3,113
248	ALA	CB	17,979
248	ALA	HB1	1,326
248	ALA	HB2	1,326
248	ALA	HB3	1,326
249	GLN	N	115,987
249	GLN	H	8,036
249	GLN	CA	59,679
249	GLN	HA	3,841
249	GLN	CB	27,914
249	GLN	HB2	2,083
249	GLN	HB3	2,139
249	GLN	CG	33,683
249	GLN	HG2	2,356
249	GLN	HG3	2,449
250	GLN	N	120,754
250	GLN	H	8,028
250	GLN	CA	59,038
250	GLN	HA	4,066
250	GLN	CB	27,914
250	GLN	HB2	1,989
250	GLN	HB3	2,151
250	GLN	CG	33,683
250	GLN	HG2	2,358
250	GLN	HG3	2,462
251	ALA	N	123,496
251	ALA	H	8,406
251	ALA	CA	55,512
251	ALA	HA	4,217
251	ALA	CB	20,222
251	ALA	HB1	0,844
251	ALA	HB2	0,844
251	ALA	HB3	0,844
252	PHE	N	118,967
252	PHE	H	8,402
252	PHE	CA	61,602
252	PHE	HA	4,069
252	PHE	CB	39,773
252	PHE	HB2	3,033
252	PHE	HB3	3,241
252	PHE	CD1	132,127
252	PHE	HD1	7,064
252	PHE	CE1	132,346
252	PHE	HE1	7,178
253	LYS	N	119,92
253	LYS	H	7,827
253	LYS	CA	60,32
253	LYS	HA	3,77
253	LYS	CB	33,042
253	LYS	HB2	1,861
253	LYS	HB3	1,914
253	LYS	CG	24,709
253	LYS	HG2	1,278
253	LYS	HG3	1,438
253	LYS	CD	29,837
253	LYS	HD2	1,631

254	TYR	N	120,993
254	TYR	H	8,177
254	TYR	CA	61,28
254	TYR	HA	4,276
254	TYR	CB	37,529
254	TYR	HB2	3,215
254	TYR	HB3	3,318
255	LEU	N	121,361
255	LEU	H	8,516
255	LEU	CA	58,077
255	LEU	HA	3,424
255	LEU	CB	42,797
255	LEU	HB2	1,071
255	LEU	HB3	1,927
255	LEU	CG	26,312
255	LEU	HG	1,969
255	LEU	CD1	22,466
255	LEU	HD1	0,446
255	LEU	HD1	0,446
255	LEU	HD1	0,446
255	LEU	CD2	26,312
255	LEU	HD2	0,832
255	LEU	HD2	0,832
255	LEU	HD2	0,832
256	ARG	N	115,749
256	ARG	H	7,608
256	ARG	CA	59,038
256	ARG	HA	3,778
256	ARG	CB	30,478
256	ARG	HB2	1,562
256	ARG	HB3	1,604
256	ARG	CG	26,953
256	ARG	HG2	1,237
256	ARG	HG3	1,252
256	ARG	CD	43,298
256	ARG	HD2	2,71
256	ARG	HD3	2,859
257	GLU	N	114,319
257	GLU	H	8,528
257	GLU	CA	58,077
257	GLU	HA	4,141
257	GLU	CB	30,84
257	GLU	HB2	1,885
257	GLU	HB3	1,941
257	GLU	CG	36,888
257	GLU	HG2	2,103
257	GLU	HG3	2,296
258	GLU	N	116,856
258	GLU	H	7,986
258	GLU	CA	57,756
258	GLU	HA	4,398
258	GLU	CB	30,478
258	GLU	HB2	1,693
258	GLU	HB3	1,874
258	GLU	CG	36,568
258	GLU	HG2	1,993
258	GLU	HG3	2,167
259	VAL	N	120,397
259	VAL	H	7,036
259	VAL	CA	66,41

259	VAL	HA	3,617
259	VAL	CB	31,76
259	VAL	HB	1,868
259	VAL	CG1	22,145
259	VAL	HG1	0,886
259	VAL	HG1	0,886
259	VAL	HG1	0,886
259	VAL	CG2	22,466
259	VAL	HG2	0,424
259	VAL	HG2	0,424
259	VAL	HG2	0,424
260	LYS	N	112,174
260	LYS	H	8,333
260	LYS	CA	63,205
260	LYS	HA	3,615
260	LYS	CB	32,081
260	LYS	HB2	1,626
260	LYS	HB3	2,304
260	LYS	CG	25,35
260	LYS	HG2	0,785
260	LYS	HG3	1,164
260	LYS	CD	29,196
260	LYS	HD2	1,48
260	LYS	HD3	1,548
260	LYS	CE	42,016
260	LYS	HE2	2,865
261	THR	N	114,676
261	THR	H	9,038
261	THR	CA	60,32
261	THR	HA	4,908
261	THR	CB	73,14
261	THR	HB	3,625
261	THR	CG2	21,184
261	THR	HG2	0,839
261	THR	HG2	0,839
261	THR	HG2	0,839
262	PHE	N	118,979
262	PHE	H	8,874
262	PHE	CA	57,436
262	PHE	HA	4,644
262	PHE	CB	41,375
262	PHE	HB2	2,744
262	PHE	HB3	2,885
262	PHE	CD1	132,565
262	PHE	HD1	7,184
262	PHE	CE1	131,69
262	PHE	HE1	7,084
262	PHE	CZ	129,939
262	PHE	HZ	6,805
263	GLN	N	125,76
263	GLN	H	9,009
263	GLN	CA	56,795
263	GLN	HA	3,464
263	GLN	CB	25,03
263	GLN	HB2	1,655
263	GLN	HB3	1,804
263	GLN	CG	32,722
263	GLN	HG2	0,705
263	GLN	HG3	1,375
264	GLY	N	102,878

264	GLY	H	8,471
264	GLY	CA	45,231
264	GLY	HA3	3,481
264	GLY	HA2	4,031
265	LYS	N	120,874
265	LYS	H	7,59
265	LYS	CA	53,59
265	LYS	HA	4,877
265	LYS	CB	33,683
265	LYS	HB2	1,754
265	LYS	HB3	1,875
265	LYS	CG	24,709
265	LYS	HG2	1,417
265	LYS	CD	29,196
265	LYS	HD2	1,719
265	LYS	CE	43,618
265	LYS	HE2	3,153
266	PRO	CD	51,026
266	PRO	CA	63,525
266	PRO	HA	4,413
266	PRO	CB	32,081
266	PRO	HB2	1,766
266	PRO	HB3	2,283
266	PRO	CG	27,594
266	PRO	HG2	1,896
266	PRO	HG3	2,092
266	PRO	HD2	3,606
266	PRO	HD3	3,812
267	ILE	N	123,744
267	ILE	H	7,804
267	ILE	CA	61,923
267	ILE	HA	3,874
267	ILE	CB	38,811
267	ILE	HB	1,323
267	ILE	CG2	17,658
267	ILE	HG2	0,661
267	ILE	HG2	0,661
267	ILE	HG2	0,661
267	ILE	CG1	29,196
267	ILE	HG1	0,849
267	ILE	HG1	1,329
267	ILE	CD1	14,133
267	ILE	HD1	0,607
267	ILE	HD1	0,607
267	ILE	HD1	0,607
268	MET	N	131,628
268	MET	H	9,715
268	MET	CA	54,29
268	MET	HA	4,688
268	MET	CB	33,042
268	MET	HB2	1,807
268	MET	HB3	2,073
268	MET	CG	31,76
268	MET	HG2	2,447
268	MET	HG3	2,554
268	MET	CE	19,581
268	MET	HE1	1,245
268	MET	HE2	1,245
268	MET	HE3	1,245
269	ALA	N	124,935

269	ALA	H	8,266
269	ALA	CA	50,385
269	ALA	HA	5,814
269	ALA	CB	23,427
269	ALA	HB1	1,215
269	ALA	HB2	1,215
269	ALA	HB3	1,215
270	ARG	N	118,252
270	ARG	H	8,805
270	ARG	CA	55,192
270	ARG	HA	4,597
270	ARG	CB	34,324
270	ARG	HB2	1,661
270	ARG	HB3	1,79
270	ARG	CG	26,953
270	ARG	HG2	1,394
270	ARG	HG3	1,551
270	ARG	CD	43,618
270	ARG	HD2	3,141
271	ILE	N	122,076
271	ILE	H	8,286
271	ILE	CA	60,64
271	ILE	HA	4,658
271	ILE	CB	39,452
271	ILE	HB	1,237
271	ILE	CG2	17,017
271	ILE	HG2	0,526
271	ILE	HG2	0,526
271	ILE	HG2	0,526
271	ILE	CG1	27,914
271	ILE	HG1	0,571
271	ILE	HG1	1,068
271	ILE	CD1	13,492
271	ILE	HD1	0,477
271	ILE	HD1	0,477
271	ILE	HD1	0,477
272	LYS	N	128,747
272	LYS	H	8,682
272	LYS	CA	56,058
272	LYS	HA	4,291
273	ALA	N	128,143
273	ALA	H	8,381
273	ALA	CA	51,987
273	ALA	HA	4,66
273	ALA	CB	19,581
273	ALA	HB1	1,233
273	ALA	HB2	1,233
273	ALA	HB3	1,233
274	ILE	N	120,278
274	ILE	H	8,313
274	ILE	CA	60,961
274	ILE	HA	4,207
274	ILE	CB	40,093
274	ILE	HB	1,69
274	ILE	CG2	17,658
274	ILE	HG2	0,739
274	ILE	HG2	0,739
274	ILE	HG2	0,739
274	ILE	CG1	27,273
274	ILE	HG1	0,993

274	ILE	HG1	1,678
274	ILE	CD1	13,171
274	ILE	HD1	0,635
274	ILE	HD1	0,635
274	ILE	HD1	0,635
277	PHE	CA	58,251
277	PHE	CB	39,659
278	PHE	N	120,919
278	PHE	H	7,965
278	PHE	CA	57,659
278	PHE	CB	39,778
279	ALA	N	124,806
279	ALA	H	7,935
279	ALA	CA	52,43
279	ALA	HA	4,449
279	ALA	CB	19,773
279	ALA	HB1	1,246
279	ALA	HB2	1,246
279	ALA	HB3	1,246
280	LYS	N	119,966
280	LYS	H	8,126
280	LYS	CA	57,103
280	LYS	CB	32,818
281	ASN	CA	53,41
281	ASN	CB	38,653
282	GLY	N	108,479
282	GLY	H	8,065
282	GLY	CA	45,541
282	GLY	HA3	3,853
282	GLY	HA2	3,687
283	TYR	N	119,563
283	TYR	H	7,89
283	TYR	CA	57,891
283	TYR	CB	38,857
283	TYR	HB2	2,802
283	TYR	HD1	6,925
284	ARG	N	122,304
284	ARG	H	8,025
284	ARG	CA	55,787
284	ARG	CB	31,199
284	ARG	HB2	1,632
285	LEU	N	123,138
285	LEU	H	8,137
285	LEU	CA	55,292
285	LEU	HA	4,199
285	LEU	CB	42,147
285	LEU	HB2	1,473
285	LEU	HD1	0,769
285	LEU	HD1	0,769
285	LEU	HD1	0,769
286	MET	N	121,038
286	MET	H	8,172
286	MET	CA	55,231
286	MET	CB	33,353
287	ASP	N	126,995
287	ASP	H	7,824
287	ASP	CA	55,906
287	ASP	CB	42,207

2021

CLÁUDIO ALEXANDRE LOPES COLAÇO

UNDERSTANDING THE DETAILED MECHANISM OF INTERACTION BETWEEN LA-RELATED PROTEINS AND THEIR RNA TARGETS

NOVA SCHOOL OF SCIENCE & TECHNOLOGY