RNA chaperone activity of the RRM2 domain of human La

Samel Sharifi

A THESIS SUBMITTED TO FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

August 2015

ABSTRACT

Genuine La proteins are abundant RNA binding factors that primarily facilitate maturation of nascent precursor RNAs in eukaryotes. Chaperone activity is a conserved feature in the La superfamily, yet the structural basis for this activity remains obscure. RNAs bound by human La (hLa) primarily make contacts with the conserved La module in the N-terminal domain. However, a role for the non-canonical RRM2 fold in the C-terminal domain (CTD) has been proposed recently. Using gel shift assays, nuclear magnetic resonance (NMR) spectroscopy, and circular dichroism (CD) spectroscopy, we aim to understand hLa's uncharacterized RRM2 dependent binding. Our results demonstrate that the RRM2 relies on non-canonical features including a subsequent disordered C terminal to bind RNA. Consequently, we predict that this region becomes ordered to facilitate hLa's chaperone function. Taken together, these data support a model for disordered regions in chaperone activity and expand our understanding of La's function.

ACKNOWLEDGMENTS

First and foremost I offer my sincerest gratitude to my supervisor, Dr. Mark Bayfield, who has supported me throughout my thesis. His guidance, constructive criticism, and constant encouragement throughout the course of my thesis shall carry me a long way through life.

I would like to acknowledge my advisors, Dr. Peter Cheung, and Dr. Derek Wilson for reviewing this thesis and providing me with helpful feedback. I am indebted to Dr. Logan Donaldson and his graduate student Sarah Ramsook, for their assistance with NMR spectroscopy and feedback through this project.

Thanks to the Bayfield lab group; Rawaa, Ana, Jyotsna, and Stefano, as well as past members, for their kindness and friendship.

Lastly, I thank my parents, sister, and friends for providing me with encouragement and support throughout this tough but wonderful experience. It would not have been possible to write this thesis without the help and support of the kind people mentioned above.

TABLE OF CONTENTS

Abstract ii
Acknowledgements iii
Table of Contents iv
List of Tables vi
List of Figures vii
Abbreviations ix

1. Introduction 1

1.2 La's function 2

1.3 La's structure 6
1.3.1 The La Module in the N terminal domain 7
1.3.2 La's C terminal domain 9
1.3.3 RNA Recognition Motifs 11

1.4 Conservation of RNA binding domains in genuine La and LARPS 15
1.4.1 LARP7 p65 RRM2 crystal structure reveals a novel-binding mode 17

1.5 Role of structural disorder in RNA chaperones 18

1.6 Thesis Proposal 22

2. Materials and Methods 25

2.1 Study design 25
2.1.1 Template and vector isolation 25
2.1.2 Primer design 27
2.1.3 Amplifying plasmid DNA by Polymerase Chain Reaction (PCR)
2.1.4 PCR product purification 28
2.1.5 DNA digest and purification 29
2.1.6 DNA purification using gel electrophoresis 30
2.1.7 Ligation of insert and plasmid DNA 30
2.1.8 Transformation of purified plasmids into RbCl component *E. coli* cells 30

2.2 Ni^{2+} affinity purification 31

2.3 Assessing secondary structure by Circular Dichroism Spectroscopy 33
2.3.1 Secondary structure analysis 34

2.4 Assessing CTD fragments affinity by Electrophoretic Mobility Shift Assay 35
2.4.1 Probe labelling and purification 36

2.5 1H-15N NMR Spectroscopy of hLa CTD fragments 37

3. Results 39

3.1 Protein purification 39

3.2 The RRM2s C terminal extension is required for RNA binding 40

3.3 Increased α-helical content may mediate Protein-RNA interaction 47

3.4 The RRM2 of hLa binds 12nt trailer RNA non-canonically 52

4. Discussion

4.1 A role for the CTD of hLa in RNA affinity 59

4.2 The RRM2 alone is not sufficient for binding RNA 60

4.3 RRM2 uses a surface other than the canonical one to bind RNA 62

4.4 Increased α-helical content may mediate hLa's chaperone activity 65

4.5 Conclusions 68

4.6 Future directions 68

Supplementary Figures 70

References 71

LIST OF TABLES

Table 1. Primers used for cloning truncated CTD mutants. 26

Table 2. Quantitative representation of affinity of hLa mutants. 47

Table. 3. Secondary structure analysis from Circular Dichroism spectra of hLa CTD fragments interacting with 12nt trailer. 52

LIST OF FIGURES

Figure 1. Alignment of the La ortholog sequences in eukaryotes. 2

Figure 2. Structure of the three RNA-binding surfaces within human La. 7

Figure 3. Comparing canonical RRMs with RRM1 and RRM2 of hLa. 14

Figure 4. Alignment of the conserved structural features of genuine La proteins and Human LARPs. 16

Figure 5. Ribbon diagram of Larp7 family member p65 indicating a novel-binding mode. 18

Figure 6. The entropy transfer model of the role of structural disorder in chaperone function. 22

Figure 7. Schematic depiction of constructed hLa CTD fragments. 25

Figure 8. Circular Dichroism Spectroscopy methodology using a Quartz two compartment cuvette. 34

Figure 9. 2μg of purified hLa CTD protein fragments resolved on a 10% SDS-PAGE gel. 39

Figure 10. 225-332 EMSA results of hLa CTD fragments binding radiolabelled 12nt trailer performed under different KCl concentrations. 42

Figure 11. 225-339 EMSA results of hLa CTD fragments binding radiolabelled 12nt trailer performed under different KCl concentrations. 43

Figure 12. 225-344 EMSA results of hLa CTD fragments binding radiolabelled 12nt trailer performed under different KCl concentrations. 44

Figure 13. 225-375 EMSA results of hLa CTD fragments binding radiolabelled 12nt trailer performed under different KCl concentrations. 45

Figure 14. 225-408 EMSA results of hLa CTD fragments binding radiolabelled 12nt trailer performed under different KCl concentrations. 46

Figure 15. 2uM to 5uM is the optimal protein concentration of hLa 225-332 CTD fragment for Circular Dichroism. 50

Figure 16. Circular Dichroism spectra of hLa CTD fragments interacting with 12nt trailer. 51

Figure 17. Purified ^{15}N labeled hLa CTD protein fragments resolved on a 10% SDS-PAGE gel. 55

Figure 18. 1H-15N HSQC spectra for hLa 225–332 and 225–408 protein fragments. 56

Figure 19. Chemical shift perturbation difference of 1H-15N HSQC spectra for hLa 225–332 between bound and unbound forms. 57

Figure 20. 3D schematic representations of protein structure depicting chemical shift perturbations of hLa 225-332. 58

Figure 21. Model of RNA binding by the RRM2 and SBM of hLa. 67

Figure 22. Secondary structure prediction of the short basic motif (SBM) of hLa. 70

LIST OF ABBREVIATIONS

3'OH	3' hydroxyl group
3'P	3' phosphate
5' leader	5'end of pre-tRNAs
5'UTR	5' untranslated
5'TOP	5' terminal oligopyrimidines
12nt trailer	3'end of pre-tRNAs containing a poly-uridylate rich sequence
^{13}C	Carbon isotope
^{15}N	Nitrogen isotope
α3x	extension of the α3 helix in the p65/RNA complex
Arg	Arginine
APE	A Plasmid Editor
bp	Base pairs
BIP	Binding immunoglobulin protein
BSA	Bovine serum albumin
BL21	*E.coli* competent cell strain
CD	Circular Dichroism Spectroscopy
CTD	C-terminal domain
CCD1	Cyclin D1
CK2	Casein Kinase 2
ddH_2O	DNase/RNase free Double-distilled water
DEPC	Diethylpyrocarbonate
DH5α	*E.coli* competent cell strain
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside Triphosphate
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediamine tetra-acetic acid
EMSA	Electrophoretic mobility shift assay
FRET	Fluorescence resonance energy transfer
FP	Fluorescence Polarization
[γ-^{32}P]	Radiolabel gamma adenosine 5'-triphosphate [γ-^{32}P]
g	Force of Gravity (centrifugal speed unit)
HBV	Hepatitis B virus
HCV	Hepatitis C virus
Hexahis	6X Histidine tag
hLa	Human La
IDT	Integrated DNA Technologies
IPTG	Isopropyl thiol galactopyranoside
IRES	Internal Ribosomal Entry Site
Kan	Kanamycin
kb	Kilobase pairs

KCl	Potassium Chloride
kDa	Kilodalton
K_d	Equilibrium constant
k_{on}	Association constant
k_{off}	Dissociation constant
L	Liters
LAM	La-motif
LARP	La-related protein
LB	Luria-Bertani growth medium (Luria Broth)
loop3	predicted loop between the β2 and β3 sheets of RRM1
µg	Microgram
µL	Microliter
µM	Micromolar
mM	Millimolar
mRNA	messenger RNA
NaCl	Sodium Chloride
NLS	Nuclear localization signal
NMR	Nuclear magnetic resonance spectroscopy
NRE	Nuclear retention element
NTD	N-terminal Domain
ng	Nanogram
OD_{600nm}	Optical density
p65	Telomerase associated protein
PKB/AKT	Protein kinase B
PMSF	Phenylmethylsulfonyl fluoride
P-TEFb	Positive transcription elongation factor β
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RNP1	Conserved RRM Ribonucleoprotein sequence 1
RNP2	Conserved RRM Ribonucleoprotein sequence 2
RRM	RNA- recognition motif
S366	Serine 366 phosphorylation site
SBM	Short basic motif
snRNA	Small nuclear RNA
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
tRNA	transfer RNA
uORF	Upstream open reading frame
U	Units of enzyme activity
U1A	U1 small nuclear ribonucleoprotein-associated protein
UUU-3'OH	3' end of poly-uridylate rich sequence

1. Introduction

Genuine La proteins are abundant and conserved RNA binding factors that target a variety of both coding and non-coding RNAs (Wolin & Cedervall, 2002). Although first identified as an auto-antigen in systemic lupus erythematous and in Sjogren's syndrome more than 30 years ago, the mechanism of La's involvement is still unknown (Lerner et al., 1981; Bayfield et al., 2010). Since then, phylogenetic analysis has revealed that La is found in nearly all eukaryotes, with the exception of the *Plasmodium* genus (Bousquet-Antonelli & Deragon, 2009). Studies indicate that La is essential in mice but is dispensable in yeast, which is where most studies of La's function have been conducted (Yoo & Wolin, 1994; Van Horn et al., 1997). Homologs of La range between 32 and 50 Kilodalton (Kda) across species and Human La (hLa) is a 50 Kda protein that is estimated to have a copy number of 2 $x10^7$ per cell (Fan, et al., 1998; Yoo & Wolin, 1994; Gottlieb & Steitz, 1989).

La associates with several classes of RNAs that function in various cellular processes (Wolin & Cedervall, 2002). All RNAs bound by La make contacts with the N-terminal domain (NTD), which is conserved in La orthologs (Fig.1). This activity is mapped to the La module, which is the La motif (LAM) and RNA Recognition Motif 1 (RRM1) acting synergistically in the NTD. The NTD has been well characterized, however recent data suggest a role for the C-terminal domain (CTD) in RNA affinity, bearing in mind that some La's don't have this feature (Martino et al., 2012).

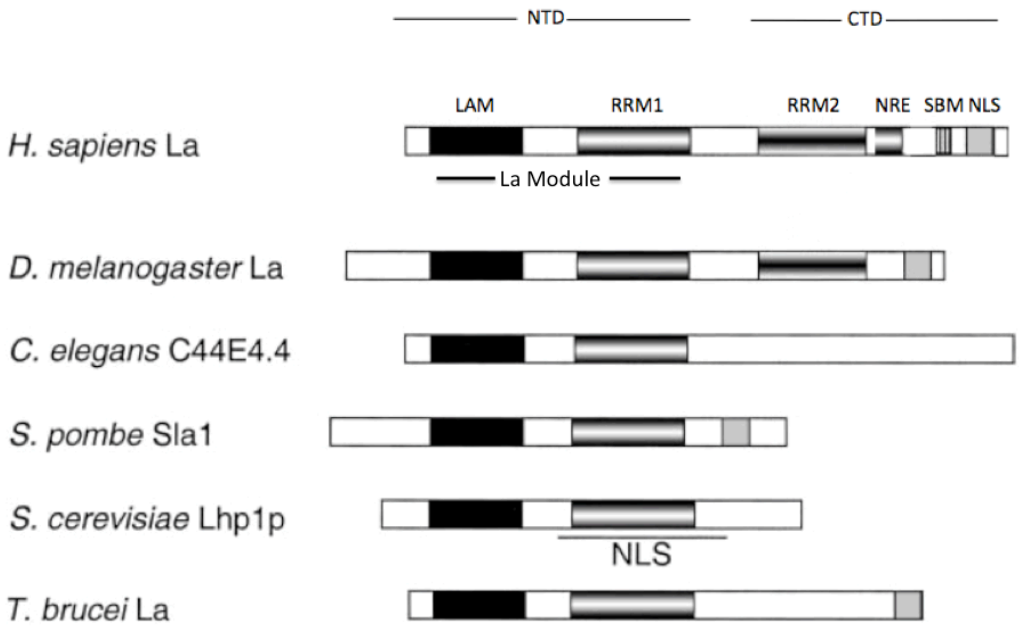


Figure 1. Alignment of the La ortholog sequences in eukaryotes. Conservation of La in human, Drosophila, *C. elegans, S. pombe, S. cerevisiae, and T. brucei*. La is bipartite structured and has a highly conserved NTD and less conserved CTD. La's NTD contains the La module, which is comprised of the La motif (LAM) and RNA recognition motif (RRM1) and are separated by a short linker sequence. Together, the LAM and RRM1 act cooperatively to form a binding pocket to bind RNA polymerase III transcripts. The CTD is less well conserved and comprised of a putative RNA recognition motif 2 (RRM2), a nuclear retention element (NRE), an intrinsically disordered short basic motif (SBM), and a nuclear localization signal (NLS). Adapted from Wolin & Cedervall, 2002.

1.2 La's function

La is primarily localized in the nucleus and is hypothesized to be responsible for nuclear retention of small RNAs (Hendrick et al., 1981; Simons et al., 1996; Wolin & Cedervall, 2002). The best-characterized function of La is to bind polymerase III precursor transcripts including pre-transfer RNAs (tRNAs), pre-U6

small nuclear RNA (snRNA), and pre-5S ribosomal RNA (rRNA), among others (Bayfield & Maraia, 2009; Wolin & Cedervall, 2002). These RNAs contain a characteristic poly-uridylate rich sequence (UUU-3'OH) at their 3' ends (Bayfield & Maraia, 2009). By recognizing the 3' termini, La protects RNAs from exonuclease digestion and also assists with their maturation (Bayfield & Maraia, 2009). Moreover, La also is hypothesized to be involved in pre-tRNA remodeling through its chaperone function (Naeeni et al., 2012).

Of the RNAs containing a terminal UUU-3'OH, La's best-characterized targets are nascent pre-tRNAs, which require extensive processing during maturation (Bertrand et al., 1998; Phizicky, 2005). Certain pre-tRNAs require assistance for efficient folding and La becomes essential to protect these from degradation and assist in processing into their mature forms (Yoo & Wolin, 1997; Chakshusmathi et al., 2003). Both Polymerase II and III transcripts acquire a terminal UUU-3'OH during transcription and La preferentially binds nascent precursor transcripts as the UUU-3'OH tract is cleaved during maturation (Hashimoto & Steitz, 1983; Xue et al., 2000; Bayfield & Maraia, 2009). La uses two distinctive binding modes to assist in pre-tRNA maturation and assembly; A UUU-3'OH dependent and a UUU-3'OH independent binding mode, which are both mapped to the NTD (Kotik-Kogan et al., 2008). The primary binding mode is the UUU-3'OH dependent mode to the pre-tRNA trailer by the LAM, which protects the 3' trailer of nascent pre-tRNAs from exonuclease digestion. The UUU-3'OH independent binding uses the RRM1 to bind to the main body of the tRNA and assist with proper folding (Teplova et al., 2006;

Maraia & Lamichhane, 2011). La stays on nascent pre-tRNA until the trailer is cleaved to ensure proper folding and is required to when there are mutations in essential tRNA that destabilize secondary structure (Bayfield et al., 2009). Binding efficiency of hLa increases with an increased content of uridylates. Conversely, La's binding efficiency is significantly reduced with RNAs containing fewer than three uridylates, indicating the importance of terminal UUU-3'OH length (Bayfield et al., 2010). The 3' hydroxyl group (3'OH) on the terminal end of RNAs are required for La's recognition as conversion of the yeast pre-tRNA and U6 snRNA 3'OH to a 3' phosphate (3'P) decreased La's affinity for these substrates significantly (Stefano, 1984; Yoo & Wolin, 1994). Moreover, La binds to pre-tRNAs at their 5' ends (5' leader), containing 5'triphosphate more efficiently than dephosphorylated RNA (Fan et al., 1998). These data suggest that La binds its substrates via electrostatic interactions and requires a negatively charged group on RNAs for efficient binding. While the pre-tRNA protection and stabilization activity of La has been well characterized, how La targets cytoplasmic mRNA remains unclear. Although most RNAs La binds have UUU-3'OH termini sequence, La also associates with host mRNAs and viral mRNAs in the cytoplasm, which extends the intricacy of La's binding repertoire (Martino et al., 2012).

Although La is a predominately nuclear protein it also shuttles to the cytoplasm and is implicated in the translational initiation of mRNA that lack a UUU-3'OH termini sequence (Martino et al., 2012; Hussain et al., 2013). These include mRNAs with structured 5' untranslated regions (5'UTR), 5' terminal

oligopyrimidines (5'TOP), and upstream open reading frames (uORF), as well as viral mRNAs that contain Internal Ribosome Entry Sites (IRES) (Ali et al., 2000; Pudi et al., 2004; Sommer, et al., 2011). La has been proposed to destabilize mRNA secondary structure impeding translation through its chaperone function to facilitate mRNA translation, however this mechanism is still under investigation (Kuehnert et al., 2015). Specific examples of these substrates are binding immunoglobulin protein (BIP) and cyclin D1 (CCD1) mRNA (Crosio et al., 2000; Kuehnert et al., 2015).

Interestingly, The first evidence of La associating with viral RNA came when mammalian cells were infected with poliovirus (Meerovitch et al., 1993). Subsequently, La was reaffirmed to be recruited during viral infection and stimulate the translation of viral mRNA (Ali et al., 2000; Martino et al., 2012). Since the early discover of La's role in viral mRNA translation, it has been shown to promote translation of Hepatitis B virus (HBV), Hepatitis C virus (HCV), and Encephalomyocarditis virus (Horke et al., 2002; Costa-Mattioli et al., 2004; Martino et al., 2012). Presumably, the synergistic action of the three RNA-binding surfaces within human La is responsible, but this mechanism is poorly understood (Martino et al., 2012).

1.3 La's structure

La proteins have a highly conserved N terminal domain (NTD) and a less conserved C terminal domain (CTD) in size and sequence homology between species (Wolin & Cedervall, 2002). The variation of La's size across species is due to the variation in size of the CTD (Fan et al., 1998). The primary binding mode for the maturation of nascent pre-tRNAs is located in the NTD as discussed earlier, whereas the CTD has an alternative-binding mode and may be responsible for enhancing translation of both host and viral mRNA (Bayfield & Maraia, 2009; Ali et al., 2000; Sommer et al., 2011). The NTD structure and binding mode has been well characterized, however the CTD has not been well characterized (Teplova et al., 2006; Kotik-Kogan et al., 2008). To date, three RNA binding surfaces have been described for hLa (Fig.2). Studies on truncated and mutant La proteins as well as structural comparisons with RNA recognition motif (RRM) containing proteins have provided insight about conserved binding domains within La as the full-length protein structure has yet to be to be resolved.

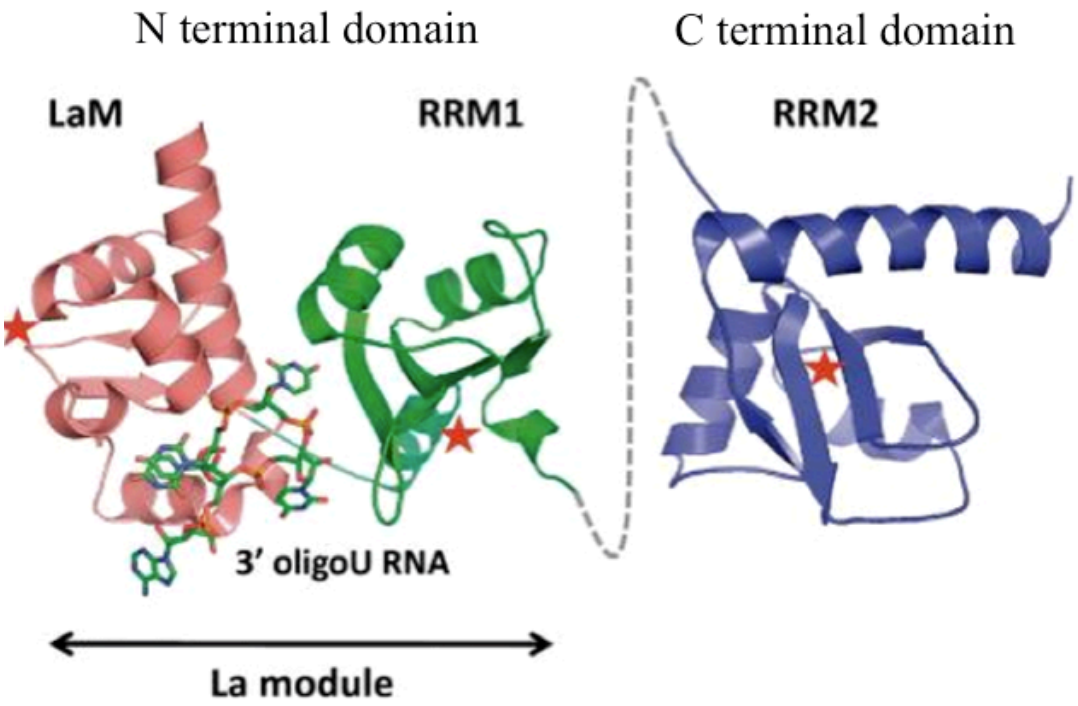


Figure 2. Structure of the three RNA-binding surfaces within human La. The La module (LAM and RRM1) was co-crystalized in complex with UGCUGUUUU ssRNA, with a UUU-3'OH end by Teplova et al., 2006. The NMR solution structure of RRM2 was solved by Jacks et al., 2003. The full-length protein structure has yet to be resolved and so the grey line indicates a linker between the NTD and CTD. Red stars indicate canonical nucleic acid binding surfaces. Adapted from Martino et al., 2012.

1.3.1 The La Module in the N terminal domain

HLa's La module in the NTD contains a tandem arrangement of the La motif (LAM) and RNA recognition motif 1 (RRM1) that is separated by short linker (Bayfield et al., 2009). Together, the LAM and RRM1, form a binding pocket and act cooperatively to bind RNA (Ohndorf et al., 2001; Alfano, et al., 2004). The La module of hLa spans residues 10-184 and has conserved structural elements common

in RNA binding proteins found in eukaryotes (Sobel et al., 1999). The LAM is a unifying feature in the La superfamily, which consists of genuine La and La related proteins (LARPs) (Bousquet-Antonelli & Deragon, 2009). The LAM is responsible for specificity and protection from exonuclease digestion of La's substrates as the majority of La's substrates end in UUU-3'OH (Bayfield & Maraia, 2009). Consistent with many RNA-binding proteins, the LAM forms a winged helix-turn-helix fold with an α1α1'α2β1α3α4α5β2β3 topology (Dong et al., 2004; Gajiwala & Burley, 2000). A co-crystal structure of hLa in complex with UGCUGUUUU ssRNA revealed that all contacts to the UUU-3'OH termini were made on the backside of the LAM, away from the expected nucleic acid binding, winged helix-turn-helix motif (Teplova et al., 2006). The LAMs activity is dependent on residues that include Glutamic Acid 20 (Q20), Tyrosine 23 (Y23), and Aspartic acid 33 (D33), which are involved in base stacking interactions with the UUU-3'OH end of RNA. Point mutations of conserved aromatic LAM residues and deletion of the LAM results in a significant decrease in RNA affinity (Huang et al., 2006; Teplova et al., 2006; Dong et al., 2004).

C terminal to the LAM is the RRM1, which adopts a canonical RRM fold with the addition of a α3 helix followed by a disordered region. This region is predicted to bind specifically to the La's substrates and provides a basis for La's chaperone activity by assisting pre-tRNA folding (Teplova et al., 2006; Bayfield & Maraia, 2009). More specifically, the highly conserved and charged basic residues in the loop region between β2 and β3 sheets (loop3) have been shown to be responsible

for interactions with the pre-tRNA body, independently of the LAMs UUU-3'OH end recognition (Bayfield & Maraia, 2009; Naeeni et al., 2012). Additionally, the canonical RNA binding surface of hLa's RRM1 has conserved aromatic residues that are predicted to stack with pre-tRNA within the canonical RRM binding surface (Huang et al., 2006; Naeeni et al., 2012). Moreover, RNA chaperone activity by La was extensively investigated in Naeeni et al., 2012, and was shown to rely on the structurally disordered extension C terminal to the α3 helix of RRM1. Similarly, the RRM2 located in the CTDs binding mode has an α3 helix followed by a disordered region, however the CTD has not been well characterized.

1.3.2 La's C Terminal domain

The variation of La's size across species is due to an extended and structurally more complex CTD (Fan, et al., 1998; Wolin & Cedervall, 2002). In vertebrates, nearly half of the CTD is comprised of charged basic residues, with a high conservation of arginine, lysine and phenylalanine (Yoo & Wolin, 1994; Kuehnert et al., 2015). In humans, the CTD is comprised of a putative RRM2, a nuclear retention element (NRE), an intrinsically disordered short basic motif (SBM), a conserved phosphorylation site at serine 366 (S366), and a nuclear localization signal (NLS) (Fig.1).

Shuttling of La between the nucleus and cytoplasm is dependent on various elements including the NLS and NRE. The NLS is required for La's ability to be imported into the nucleus, while the NRE is required for nuclear retention of La.

(Intine, et al., 2002; Bayfield et al., 2007). Therefore, hLa's interaction with various substrates is a result of their different subcellular locations (Broekhuis et al., 2000; Schwartz et al., 2004). La's phosphorylation at S366 is also influenced by localization. HLa is phosphorylated in the nucleus by casein kinase II (CKII) and unphosphorylated in the cytoplasm (Maraia & Intine, 2001). Additionally, La is phosphorylated at multiple other sites that are not conserved amongst different species, but its best-characterized phosphorylation site remains S366 (Maraia & Intine, 2001; Long et al., 2001; Brenet et al., 2009; Kuehnert et al., 2015).

Central to this thesis is the role of RRM2 in hLa function, bearing in mind that this feature is not conserved in most La's (Wolin & Cedervall, 2002). The RRM2 of hLa conforms to a non-canonical RRM fold with a long α3 helix, which masks the RRM2's canonical RNA binding surface, and an addition of a β4 sheet (β4') (Jacks et al., 2003). Predicted to span residues 231-327, the RRM2 also contains a C terminal disordered region (Horke et al., 2004). Interestingly, unlike other RRMs, the RNA binding surface of RRM2 has atypical consensus sequences as well as the presence of charged acidic residues that bind unfavorably to RNA (explored in more detail in the discussion) (Jacks et al., 2003; Maris et al., 2005). Because of the atypical RRM fold and the negative electrostatic potential of the RNA binding surface, it is likely that the RRM2 of hLa may use a surface other than the canonical one.

Regions outside the canonical RRM fold to engage RNA targets, including C terminal disordered regions (Maris et al., 2005). Consequently, the basic and intrinsically disordered SBM, which is C-terminal to the RRM2 is predicted to be

important to hLa's RRM2 dependent binding (Fan et al., 1998; Intine & Maraia, 2000). The SBM of hLa spans residues 328-344 and is hypothesized to bind the 5' end of the pre-tRNAs and attenuate 5' leader processing, which is inhibited by phosphorylation of La at S366 (Fan et al., 1998). In the absence of this activity pre-tRNAs have a tendency to misfold, which are then degraded through the nuclear surveillance machinery (Callahan et al., 2010; Schmidt et al., 2012). Recently, biochemical assays have implicated the disordered region in the CTD of hLa in RNA chaperone activity of CCD1 mRNA, which can also be impaired by a novel phosphorylation site at threonine 389 (T389) by Protein kinase B (PKB/AKT) (Kuehnert et al., 2015). Both *in vitro* and i*n vivo* studies have proposed a general role for the SBM's involvement in RNA processing and chaperone activity, however the structural basis for this remains obscure.

1.3.3 RNA Recognition Motifs

RRMs are ubiquitous features of proteins that are required for RNA metabolism and are typically 90 amino acids long (Fig.3A) (Afroz, et al., 2015). There are several structural variations of the RRM family, however they are canonically arranged in a β1α1β2β3α2β4 topology, with 4 β-sheets and 2 α-helices (Loerch & Kielkopf, 2015). The RRM domain exists in all kingdoms of life and has diverse biological roles, however less than 1% of predicted RRMs have been structurally characterized to date (Afroz, et al., 2015). Conserved consensus sequences within RRMs include the ribonucleoprotein 1 (RNP1) and

ribonucleoprotein 2 (RNP2) motifs that map to the β1-sheet and β3-sheets within a core RNA binding surface (Maris et al., 2005; Cléry, et al. 2008). The RNP contacts are non-specific and contain aromatic and basic residues that interact with RNA by forming stacking interactions, hydrogen bonds, and electrostatic interactions with up to eight nucleotides of cognate RNAs (Tan et al., 1993; Oubridge et al., 1994; Loerch & Kielkopf, 2015).

As discussed earlier, regions outside the RRM can also participate in RNA binding. Both N and C terminal extremities of RRMs can take on a variety of structures, including α-helices and structurally disordered regions. These features often transition between each other and are flexible in the free form to assist with binding RNA molecules lying on the β-sheet surface. Alignment sequences of 25 RRMs containing disordered regions suggest a broader role for these atypical extremities in RNA binding (Maris et al., 2005; Loerch & Kielkopf, 2015). For instance, α-helical extremities found in an unrelated human U1 small nuclear ribonucleoprotein-associated protein (U1A) and La related protein 7 (LARP7) family member p65 transition between ordered and unordered states to assist with ligand binding (Allain et al., 1997; Singh et al., 2012).

Both RRM1 and RRM2 of hLa are atypical in that they contain a third α3 helix, followed by a disordered C terminal region (Fig.3B; Fig.3C) (Birney et al., 1993; Jacks et al., 2003; Kotik-Kogan et al., 2008). The RRM1 of hLa coupled to a short C-terminal α3 helix, followed by a disordered region may extend upon binding RNA and is hypothesized to promote chaperone activity (Naeeni, et al. 2012). In

contrast the RRM2 contains a longer α3 helix, which masks the RRM2's RNA binding surface. To date, it is unclear whether either structure extends, unfolds, or is displaced upon RNA binding (Fig.3D).

Multiple RRMs are a common feature of RNA binding proteins and hLa has two RRMs. Structural comparisons with proteins containing tandem RRMs bound to RNA provide insight to the function of hLa's RRMs. Most commonly, tandem RRMs, which are required to bind longer RNA sequences, tumble independently when unbound and form a rigid structure in their bound form (Afroz et al., 2015). The linkers between tandem RRMs are disordered in their free form, however they are hypothesized to bridge the RRMs in their bound form and interact with each other. The two RRMs form a binding pocket where the 5'end of the RNA is bound by one RRM and the 3'end of RNA by the other RRM (Maris et al., 2005). This is consistent with a recent finding that implicates both the RRM1 and RRM2 of hLa acting cooperatively to mediate translation of Hepatitis C viral (HCV) mRNA (Martino et al., 2012). In spite of all of the resolved RRM structures, there are several distinct variations, which point to a need of sub-classification of these proteins.

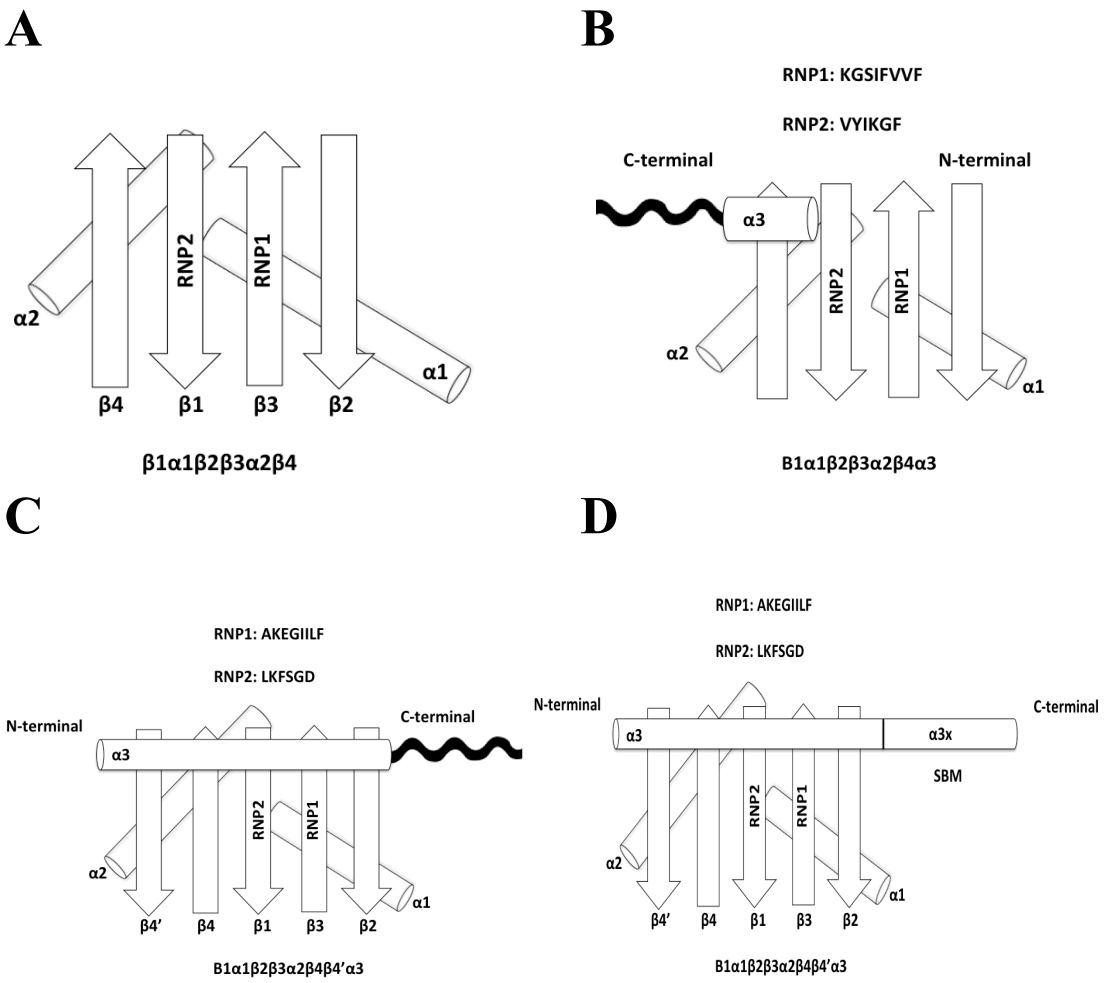


Figure 3. Comparing canonical RRMs with RRM1 and RRM2 of hLa. A) RRMs are canonically arranged in a β1α1β2β3α2β4 topology, with the RNP1 and RNP2 mapped to the β1-sheet and β3-sheets within a core RNA binding surface. B) The RRM1 of hLa coupled to a short C-terminal α3 helix is hypothesized to promote chaperone activity. C) RRM2 of hLa contains a longer α3 helix and additional β4 sheet (β4'), which masks the RRM2's RNA binding surface and is followed by a disordered region. D) Upon RNA binding the disordered region C terminal to RRM2 may extend. However, it is unclear whether either structure extends or unfolds. Consensus RNP sequences are indicated above each RRM. Adapted from Singh et al., 2013, Teplova et al., 2006, and Jacks et al., 2003.

1.4 Conservation of RNA binding domains in genuine La and LARPS

The La Module RNA binding surface in the NTD is also found in the evolutionary divergent La related-proteins (LARPs) and may provide insight on the broad involvement of La in both RNA metabolism (Bousquet-Antonelli & Deragon, 2009; Bayfield et al., 2010). The La superfamily consists of five family members including genuine La and four LARPs, LARP1, LARP4, LARP6, and LARP7 (Bousquet-Antonelli & Deragon, 2009; Bayfield et al., 2010). LARPs and La have amino acid sequence overlap and similar structural elements in the LAM, however they have diverged more significantly outside this region (Fig. 4). The RRM2 contained in genuine La can be absent in LARP1, and variable in LARP4 and LARP6, however it is structurally similar in LARP7. Recently, LARPs 4, 6, and 7 were shown to harbor RNA chaperone activity, which provide a unified function for the La superfamily (Hussain et al., 2013).

LARPs engage with a variety of RNA targets and play a critical role in RNA metabolism and mRNA translation, however their mode of RNA binding is not well characterized (Hussain et al., 2013). The most divergent LARPs from genuine La are LARP1 and LARP4, which are both cytoplasmic proteins and are involved in mRNA stability (Yang et al., 2011; Nykampet al., 2008). In contrast, LARP6 associates with collagen mRNAs and is hypothesized to promote nuclear export (Cai et al., 2010). Human LARP7 is the most closely related to genuine La proteins as its La domain consists of a LAM, a canonical RRM1, and a non-canonical RRM2 (Bousquet-Antonelli & Deragon, 2009; Singh et al., 2012). Also similar to La, the primary

targets of the LARP7 family are RNA pol II transcripts: in humans, LARP7 binds to 7SK snRNA, a critical component of the 7SK RNP, which inhibits the activity of positive transcription elongation factor β (P-TEFb) (Krueger et al., 2008; Singh et al., 2012). In ciliates, LARP7 family members p43 and p65 bind telomerase RNAs (Bayfield et al., 2010). Both human and ciliate LARP7 proteins are postulated to bind their targets via the same UUU-3'OH dependent recognition as genuine La (Bayfield et al., 2010).

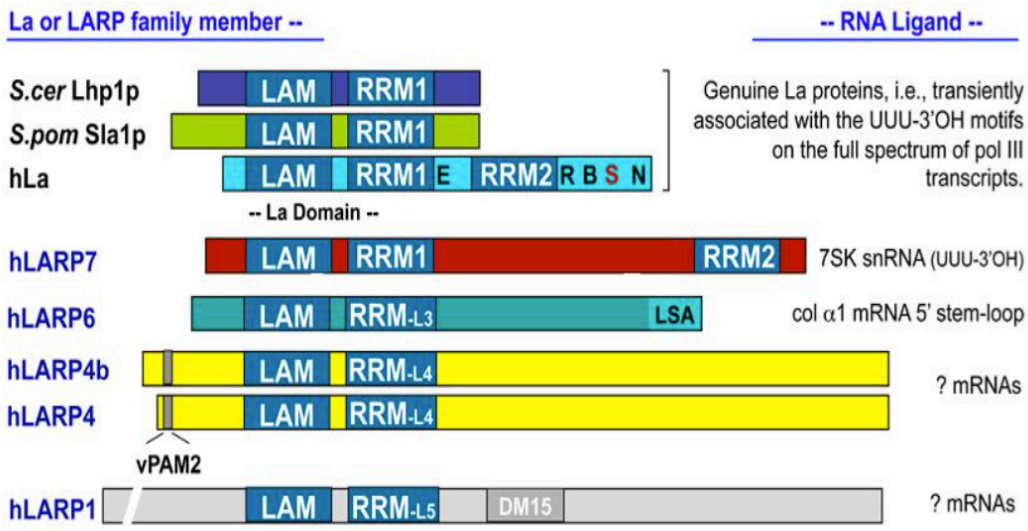


Figure 4. Alignment of the conserved structural features of genuine La proteins and Human LARPs. Genuine La homologs are from *S. cerevisiae S. pombe*, and human La. The La superfamily consists of five family members including genuine La and four LARPs, LARP1, LARP4, LARP6, and LARP7. The LAM in LARPs is conserved, however they have diverged more significantly outside this region. LARP1 and LARP4, which are both cytoplasmic proteins and are involved in mRNA stability. LARP6 associates with collagen mRNAs and is hypothesized to promote nuclear export. Human LARP7 is the most closely related to genuine La proteins as its La domain consists of a LAM, a canonical RRM1, and non-canonical RRM2. LARP7 is a critical component of the 7SK RNP. Adapted from Bayfield et al., 2010.

1.4.1 LARP7 p65 RRM2 crystal structure reveals a novel-binding mode

Recently, Singh et al (2012) revealed a novel-binding mode for the RRM2 in the CTD of the LARP7 family member and telomerase ribonucleprotein (RNP) subunit p65 that is hypothesized to be homologous to the RRM2 of hLa. The atypical RRM2 of p65 is essential for telomerase RNP assembly (Singh et al., 2013). They reported the first co-crystal structure of p65 RRM2 in complex with telomerase stem IV RNA, which is also the first structure of an RRM2 of the La superfamily in complex with RNA. Similar to hLa's RRM2, the RRM2 of p65 is atypical and consists of an additional β4' sheet as well as an extended α3 helix, which masks the RNA binding surface and is displaced when bound (Fig.5). Interestingly, the RNP1 and RNP2 sequences are absent in p65, and instead it uses aromatic residues on the β2-sheet and β3-sheet to stack onto the RNA, while forming hydrogen bonds with the RNA with residues on the β2-sheet and β3-sheet. This is highly divergent from the canonical RRMs where main contacts are made by the β1 and β3-sheets to bind RNA. C-terminal to the α3 helix is a 22 residue long intrinsically disordered region, which forms an extension of the α3 helix (α3x) in the p65/RNA complex. Additionally, three aromatic residues on α3 helix of p65 are inserted between these base pairs of the RNA to stabilize the complex. Together, p65 uses the atypical RNA binding surface, the α3 helix, and the α3x helix to bind RNA. All three structures are homologous in hLa and it has been hypothesized that hLa uses this novel-binding mode to engage RNA (Jacks et al., 2003; Singh et al., 2012, Singh et al., 2013). This finding suggests a broader role for disordered regions in La's chaperone function.

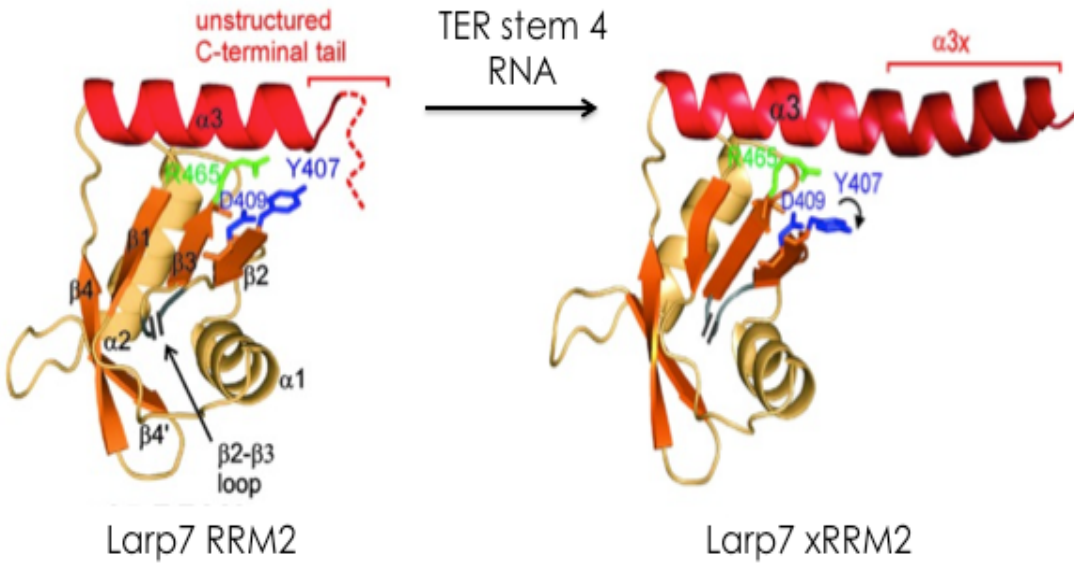


Figure 5. Ribbon diagram of LARP7 family member p65 indicating a novel-binding mode. A co-crystal structure of p65 RRM2 in complex with telomerase stem IV (TER stem 4) RNA that is hypothesized to be homologous to the RRM2 of hLa. The α3 helix, masking the RNA binding surface is displaced when bound, which allows for aromatic residues on the β3-sheet to stack onto RNA, while forming hydrogen bonds with the RNA with residues on the β2-sheet. Aromatic residues also stack onto RNA when an extended α3 helix (α3x) in a structurally disordered region C terminal to the RRM2 region is formed. Adapted from Singh et al.. 2013.

1.5 Role of structural disorder in RNA chaperones

Generally, chaperone proteins assist with catalyzing folding reactions and destabilizing RNA-RNA interactions (Doetsch et al., 2011). Folding is a kinetically slow process with a high probability for RNA to misfold, however *in vivo,* RNAs are stabilized by proteins (Zuker, 1989; Weeks, 1997). RNA molecules form a multitude of complex three-dimensional structures. There are multiple steps during the folding

process with various stable intermediate species (Mayer et al., 2007; Doetsch et al., 2011). The primary structure of RNA is single stranded and can form standard and non-standard Watson-Crick base pairs. Protein-RNA interactions are characterized by transient interactions, which destabilize misfolded RNA structures and allow for subsequent attempts at folding RNA into its native structure (Tompa & Csermely, 2004).

Chaperones are found in all living organisms and viruses and are ATP-independent (Bardwell & Jakob, 2012). The weak nonspecific interactions point to their promiscuity in binding a wide variety of substrates (Cristofari & Darlix, 2002) They serve important functions including RNA maturation, post-transcriptional modifications, translation, and cell surveillance (Woodson, 2010). RNA chaperones can have strand annealing or dissociation activity, or both. Strand annealing activity promotes the interaction with RNA and catalyzes folding reactions, while strand dissociation facilitates the destabilization RNA-RNA interactions (Rajkowitsch & Schroeder, 2007). Because of the strong structural stability of these various forms, if an RNA molecule forms an unfavorable structure, it is kinetically difficult to unwind and therefore it may be trapped in the misfolded inactive state.

La assists with RNA maturation by unwinding misfolded RNA through its chaperone function of the NTD and CTD (Naeeni et al., 2012; Kuehnert et al., 2015). This ensures proper folding of RNA and prevents newly synthesized RNA from forming non-functional structures, such as kinetic traps (Chakshusmathi et al., 2003). La not only protects the pre-tRNAs UUU-3'OH trailer's from degradation but also

rescues mutated pre-tRNAs that are misfolded (Yoo & Wolin, 1997). When tRNA modification enzymes that are important for pre-tRNA folding are deleted, La becomes essential (Chakshusmathi et al., 2003). Through fluorescent energy transfer (FRET), cis-splicing, and folding trap assays, La has been shown to harbor both strand annealing and strand displacement activities, which are mapped to RRM1 coupled to a C-terminal α3 helix (Belisova et al., 2005; Naeeni et al., 2012). As previously discussed, La's C terminal was shown to have chaperone activity and facilitate translation by binding close to the start codon of Cyclin D1 mRNA (Kuehnert et al., 2015). This activity is hypothesized to rely on strand dissociation to remove structural obstructions of mRNA during translation and was mapped to the structurally disordered region of hLa.

To date, a unified binding domain between RNA chaperones has not been defined and their mechanism of promoting strand annealing and strand displacement activities remain unclear (Rajkowitsch et al., 2005). However, intrinsic disorder is hypothesized to be required for RNA chaperones (Rajkowitsch & Schroeder, 2007). In eukaryotes, RNA chaperones contain over 50% unstructured amino acids (Tompa & Csermely, 2004). The entropy transfer model of RNA chaperone activity suggests that a structurally disordered protein becomes ordered upon binding and transfers this gain in entropy to RNA, in order to break RNA self base pairing and facilitate proper folding (Fig.6). Briefly, protein-RNA contacts make hydrogen bonds, the enthalpy of this process is hypothesized to offset the decrease in entropy in the RNA chaperone as it becomes a structurally ordered protein. This entropy is transferred onto the RNA,

which would distort RNA secondary structure, providing the RNA with a subsequent chance to access the native fold (Tompa & Csermely, 2004). In this model, RNA chaperones do not promote folding directly, but by help folding indirectly by unfolding misfolded RNA, which is impeding function. This allows the RNA a subsequent change to fold correctly.

Through entropy transfer, structurally disordered coils that are comprised of basic residues are thermodynamically favourable to form α-helices upon binding as hydrogen-bonding and electrostatic interactions between polar side chains, increase enthalpy, and influences conformational entropy change during the helix-coil transition (Chakrabartty et al., 1994; Lomize & Mosberg 1997). Consequently, this may be the basis for La's chaperone activity as residues that have a high propensity to form α-helices could not shift between structured and unstructured conformations easily upon RNA binding (Biswas et al., 2008). This may provide a unified function of disordered regions in promoting chaperone activity and explain how hLa's chaperone activity relies on the RRM2 and the SBM.

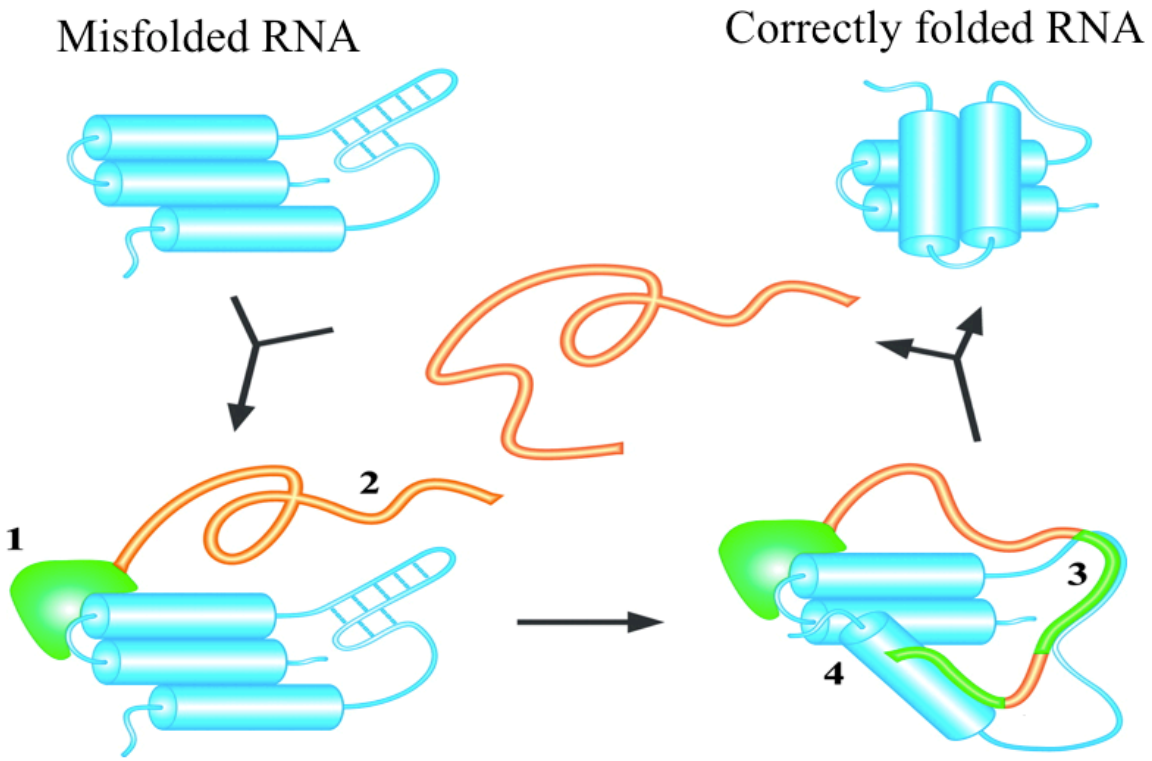


Figure 6. The entropy transfer model of the role of structural disorder in chaperone function. RNA chaperones are composed of over 50% unstructured amino acids and intrinsic disorder is hypothesized to be required for RNA chaperones. The entropy transfer model suggests that a structurally disordered protein becomes ordered upon binding and transfers this gain in entropy to RNA, in order to break RNA self base pairing and facilitate proper folding 1) A chaperone (orange) binds a misfolded RNA (blue), using the RRM (green). 2) An intrinsically disordered segment of the chaperone then binds RNA. 3) Protein-RNA contacts make hydrogen bonds, which increases enthalpy, the chaperone effectively gaining this as entropy and becoming a structurally ordered protein 4) this entropy is transferred onto the RNA, which would distort RNA secondary structure.

1.6 Thesis Proposal

La binds a variety of both coding and non-coding RNAs and has multiple functions including 3'end protection and rescuing misfolded RNAs. The mechanism of how La can fold RNAs and which regions of the protein contribute to this function in the NTD has been well characterized, however the CTDs RRM2 and SBM has not been well characterized. Because the RRM2 of hLa is atypical in structure and is highly divergent from the canonical RRMs, we use structural comparisons with RRM containing proteins, like LARP7 p65 to provide insight about conserved functions of hLa and suggest a role for disordered regions in RNA chaperone activity.

Previous work in our laboratory has demonstrated that the disordered region C-terminal to RRM2, the SBM, has chaperone activity, however the structural basis of recognition has not been determined. The SBM is predicted to primarily be disordered in hLa's unbound form, however it contains stretches of basic residues, which are thermodynamically favourable to provide RNA destabilization through the formation of an α helix (Fig. 22). Therefore, the disordered SBMs basic residues provide a platform for RNA to interact with. We have shown the SBM can help bind RNA by gel shift but we haven't shown its importance in promoting RNA native fold. We hypothesize that the SBM of hLa may become ordered and form a α-helix extension upon RNA binding. This has been postulated in the entropy transfer model and our objective is to provide experimental evidence for this model by using hLa.

The aim of this project is to structurally characterize the chaperone activity of hLa's CTD. We took a mechanistic approach by systematically generating hLa CTD

mutants to identify important regions in chaperone activity. By making targeted deletions of hLa's CTD we attempt to map the binding surface and contact residues of hLa to RNA and to better understand hLa's RRM2 dependent binding. To assess affinity of hLa's interaction with 12nt pre-tRNA trailer, Electrophoretic mobility shift assays (EMSA) were carried out. The amount and type of secondary structure was assessed using Circular Dichroism (CD) during complex formation. Finally, Nuclear Magnetic Resonance (NMR) was employed to determine the structure and dynamics of this interaction. By incorporating $^{13}C^{15}N$ isotopes into truncated mutant hLa proteins, we can determine the identity and spatial organization of amino acids and 3D structural information like α-helix formation upon binding. By using spectroscopic methods like NMR and CD as well as activity assays (EMSA), we hope to uncover the RRM2 dependent binding mode of hLa and as a result have a better understanding of La's role in RNA metabolism.

2. Materials and Methods

2.1 Study design

Several truncated La CTD fragments were constructed to test their ability to form complexes with RNA. 225-332, 225-339, 225-344, 225-363, and 225-375 hLa CTD protein fragments were subcloned into the pET28a expression vector (Novagen) containing an N-terminal Hexahis (6XHis) tag and Kanamycin marker. 225-332 lacks the SBM and NLS, 225-339 lacks the full SBM and NLS, 225-344 contains the full SBM but not the NLS, 225-363 includes the full SBM as well as residues C-terminal to the SBM, and 225-375 contains the SBM and S366 phosphorylation site (Fig.7). The full-length CTD (225-408) of hLa was constructed by a previous lab member in the same vector, but with a C-terminal 6XHis tag. The vector contains a T7lac promoter, which can be induced by isopropyl-D-thiogalactopyranoside (IPTG). These clones were generated by conventional cloning methods using full-length Human La protein as a template and clones were transformed into XL1 *E. coli* cells and plasmid DNA was purified and then transformed into BL21 *E. coli* cells. Sequences were verified by the Centre for Applied Genomics DNA sequencing (TCAG) sequencing facility (Toronto).

2.1.1 Template and vector isolation

Pet28a expression vector containing full-length hLa was constructed by a previous lab member and isolated from *E. coli* strain XL1 (DH5α) by streaking onto a

LB (Luria Broth) Kan+ plate and incubating at 37°C overnight. Pet28a empty plasmid was also streaked onto an LB Kan+ plate and incubated at 37°C overnight. A single colony from each plate was picked and inoculated 5ml of LB media supplemented with 10µg/µl Kanamycin and incubated overnight at 37°C. Plasmids were isolated using Plasmid Miniprep Kit Spin Protocol II (OMEGA Bio-Tek), according to the instructions outlined by the manufacturer.

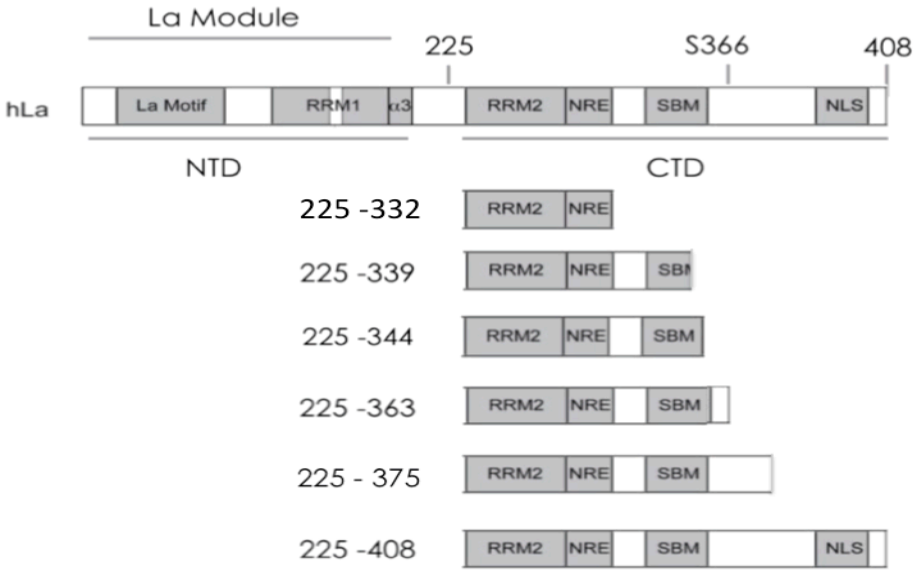


Figure 7. Schematic depiction of constructed hLa CTD fragments. The CTD is comprised of a RNA recognition motif (RRM2), a nuclear retention element (NRE), an intrinsically disordered short basic motif (SBM), and a nuclear localization signal (NLS). Additionally, a phosphorylation site at serine 366 (S366) that may play a role in La binding modes. HLa CTD protein fragments were constructed to identify important structures in chaperone activity. 225-332, 225-339, 225-344, 225-363, 225-375, and 225-408 truncated CTD fragments were subcloned into Pet28a vectors. All recombinant proteins contain the RRM2. 225-332 lacks the SBM and NLS, 225-339 lacks the full SBM and NLS, 225-344 contains the full SBM, but not the NLS, 225-363 includes the full SBM as well as residues C terminal to the SBM, 225-375 contains the SBM and S366 phosphorylation site, 225-408 contains the full length CTD. Adapted from Naeeni et al. (2012).

2.1.2 Primer design

Forward and reverse primers for the truncated hLa CTD sequences were designed using A Plasmid Editor (APE) software. To amplify this type of template DNA, oligos were purchased from Integrated DNA Technologies (IDT) and diluted to 20 μM in DNase/RNase free double distilled H20 (ddH20) (Invitrogen). A NcoI cut site, 6XHis-tag sequence, and Methionine start codon were added to the 5' end (forward primer), while the BamHI cut site and stop codon were added to the 3' end (reverse primer).

Table 1. Primers used for cloning truncated CTD mutants. HLa CTD sequences were designed using A Plasmid Editor (APE) software. To amplify this template DNA, oligos were purchased from Integrated DNA Technologies (IDT). A NcoI cut site, 6XHis-tag sequence, and Methionine start codon were added to the 5' end (forward primer), while the BamHI cut site and stop codon were added to the 3' end (reverse primer). All sequences used the same 225 forward primer and a subsequent reverse primer.

	Primer sequence (5' to 3' direction)
225 Forward	GCAGCCCATGGGCCACCATCACCATCACCATCTAGAAGAAAAGA TTGGATGCTTGCTGAA
332 Reverse	CGCGCGGATCCCTATTTTGACTTCCATTTGTTTAGGGATTCTTGTT GGTCTTCTATTATT
339 Reverse	CGCGCGGATCCCTATTTTCCTTTAAATCTACGACCTTTTGACTTCC ATTTGTTTAGGGA
344 Reverse	CGCGCGGATCCCTATTTATTACCCTTTCCTTTTCCTTTAAATCTAC GACCTTTTGACTTC
363 Reverse	CGCGCGGATCCCTATTTGCGCGGATCCTCATTTCGTTTTCTTGCCC TGAAACTGTACTTTTCC
375 Reverse	CGCGCGGATCCCTATTTGCGCGGATCCTCATTCATCATGTTCATC ATGTTCATCATCACTAGC

2.1.3 Amplifying plasmid DNA by Polymerase Chain Reaction (PCR)

To clone truncated sequences into pET28a expression vectors, inserts were prepared by setting up a Polymerase Chain Reaction (PCR) reaction. A 50ul reaction was set up in triplicate that contained 0.4 uM of Forward in Reverse primer, 200 uM Deoxynucleotide (dNTPs) (NEB), 5X Phusion buffer (NEB), 0.4 Enzyme Units (U) of High Fidelity Q5 Enzyme (NEB), 50ng/µL of full length hLa DNA in pET28a, and 35 µL of DNase/RNase free ddH20 (Invitrogen). The thermal cycler (Eppendorf) was programmed for 35 cycles, with each cycle at a denaturing temperature of 95°C for 30 seconds, annealing temperature of 55°C for 2 minutes and extension temperature of 72°C for 30 seconds. To verify amplification of inserts, 2µL of this PCR product was then electrophoresed with 6X Loading dye (2.5% Ficoll®-400,11 mM EDTA, 3.3 mM Tris-HCl, 0.017% SDS, 0.015% bromophenol blue, pH 8.0) (NEB) on a 1% agarose gel and separated at 100V for 1 hour.

2.1.4 PCR product purification

Following amplification, the PCR product was purified by phenol-chloroform/ethanol precipitation. An equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) (Fisher Scientific), was added to samples and was then centrifuged for 3 minutes at 13 000 xg. The aqueous layer was then transferred to a new Eppendorf tube. 200 µL of Chloroform was added and vortexed. The mixture was then centrifuged for 3 minutes at 13 000 xg, and the aqueous layer was transferred to

a new Eppendorf tube. 20 µL of 3M sodium acetate (pH 5.2) and 500 µL of 100% cold ethanol was then added to each sample, which were stored in the -20°C freezer overnight. On the following day, samples were centrifuged for 10 minutes at 13000 xg. The ethanol was decanted and was replaced with 500 µL of fresh 70% ethanol, which was centrifuged for 10 minutes at 13000 xg. The ethanol was decanted again, and the sample was allowed to air dry 5 minutes. The DNA pellet was then resuspended in 20 µL of DNAse/RNAse free ddH2O (Invitrogen). The insert was electrophoresed on a 1% w/v agarose gel with 6X Loading dye at 100V with three µl of EtBr for 1 hour to verify purity.

2.1.5 DNA digest and purification

The ends of the DNA inserts and Pet28a expression vector were double digested using NcoI and BamHI (NEB) endonuclease enzymes at 37°C overnight. The reaction setup for digestion of the DNA inserts included 0.2mM BSA (NEB), 10U of Nco1 (NEB), 10U BamHI (NEB), 50ng/µL of Plasmid or PCR amplified DNA, 1X NEB CutSmart Buffer, and ddH2O (Invitrogen). The total reaction volume was 20 µl for the insert DNA and 100 µl for the plasmid Pet28a vector. The Pet28a vector included 15U of both Nco1 and BamHI (NEB). Before purification and ligation, the plasmid was treated with 10U of Alkaline phosphatase/Calf intestinal phosphatase (CIP) to prevent self-ligation and was incubated at 37°C for 2 hours

2.1.6 DNA purification using gel electrophoresis

Both Pet28a plasmid DNA and insert DNA were resolved on a 1% agarose gel for 15min at 100V and gel extracted using MicroElute Gel Extraction Spin protocol (Omega Bio-Tek). One µL of each of the gel-purified samples were resolved on a 1% agarose gel to check for sizes and quality of the DNA. The DNA was quantified by comparing the intensity of insert and vector to the markers intensity (Lambda DNA; Bioshop).

2.1.7 Ligation of insert and plasmid DNA

The digested plasmid DNA was then ligated with inserts using a 3:1 molar ratio of insert to plasmid. The reaction setup for ligation included 100ng pET28a vector, 300ng of insert DNA, 1X T4 Ligase buffer (NEB), 40U of T4 DNA ligase (NEB), and 4 µL of ddH2O in a total reaction volume of 10µl. This reaction was incubated at 16°C overnight. A negative control was also incubated overnight that did not contain an insert in the reaction. The positive and negative control ligation was visualized on a 1% w/v agarose with 6X Loading dye electrophoresed at 100V with 3 µl of EtBr for 1 hour.

2.1.8 Transformation of purified plasmids into RbCl component *E. coli* cells

To transform plasmids into RbCl component XL1 *E. coli* cells, initially cells were thawed on ice for 20 minutes. Aliquots of 50 µL of cells were used per

transformation and placed into pre-chilled Eppendorf tubes. 5 µL of DNA was then added to each pre-chilled Eppendorf tubes and incubated on ice for 1 hour. The samples were then heat shocked at 37°C for 90 seconds, chilled on ice for 2 minutes, and then 400 µL of LB media was added to each sample. The cells were incubated at 37°C for 90 minutes and plated on LB + 10µg/µl Kanamycin plates and incubated overnight at 37°C. On the following day 5ml of each XL1 *E. coli* cells containing plasmids were cultured in LB + 10µg/µl Kan+ media and incubated overnight at 37°C. Plasmids were isolated using Plasmid Miniprep Kit II (OMEGA bio-tek), according to the instructions outlined by the manufacturer.

Isolated plasmids were also double digested using NcoI and BamHI to verify they were carrying an insert and sequenced at the TCAG sequencing facility (Toronto). The reaction setup for digestion included 0.2mM BSA (NEB), 10U Nco1 (NEB), 10U BamHI, 1µl Plasmid DNA, 1X NEB CutSmart Buffer, and ddH2O (Invitrogen) in a total reaction volume of 20 µL, incubated at 37°C. Insert analysis was visualized on a 1% w/v agarose with 6X Loading dye at 100V with 3 µl of EtBr for 1 hour. After confirming that plasmids contained correct sequences, these plasmids were transformed into BL21 (DE3) cells (see section 2.1.8) and plated on LB supplemented with 10µg/µl Kanamycin plates and incubated at 37°C overnight.

2.2 Ni^{2+} affinity purification

BL21 (DE3) *E. coli* cells expressing hLa protein fragments were grown at 37°C in 1.5L of LB media supplemented with 10µg/µl Kanamycin. Cells were grown

to exponential phase reaching an optical density (OD600nm) of 1.0, followed by 1mM IPTG induction for 3 hours at 37°C for soluble protein production. Cell pellets were collected by centrifugation (Beckman) at 4°C, 5000rpm for 10 minutes, and resuspended in 50mL of lysis buffer (20mM Tris-HCl pH 7.6, 200mM NaCl, 0.05% NP40, 20mM imidazole, 1mM BME, 1X Halt Protease Inhibitor Cocktail (PIC; Thermo Scientific) and 1mM PMSF (Phenylmethylsulfonyl fluoride)). Cells were pulse sonicated (Branson Digital Sonifier) using 5mV for 15 seconds each, with a 15-second break between each pulse for a total of 15minutes (7.5 minutes on). The lysate was cleared by centrifugation at 30000rpm (Eppendorf) for 30 minutes. The extract was purified over a 1mL of pre-packed HisTrap FF with precharged Ni Sepharose (GE Healthcare) by sequential washes with wash buffer (Lysis Buffer, no NP40, 40mM imidazole) and eluted in 10ml of elution buffer (Lysis buffer, 300mM imidazole). The eluant was concentrated to 500uL by ultrafiltration (10K Amicon Ultra 15mL centrifugal filter; Millipore) and was analyzed on a 10% SDS-PAGE gel alongside BSA standards to determine molar concentrations. Samples were visualized using Coomassie staining (45% methanol, 10% acetic acid, 0.25% Coomassie Brilliant Blue, R-250) for 60 minutes and destain solution (40% methanol, 10% acetic acid) overnight. 2µg of purified of recombinant hLa proteins were electrophoresed a 10% SDS-PAGE to verify concentrations. Approximate sizes for hLa CTD protein in Kilo daltons (Kda) were calculated using ExPASy online ProtParam tool (http://web.expasy.org/protparam).

2.3 Assessing secondary structure by Circular Dichroism Spectroscopy

Circular Dichroism (CD) Spectroscopy was used to assess the type and amount of secondary structure of hLa CTD fragments interacting with 12 nucleotide Arginine trailer RNA, was based on the human tRNAArgICG26 and purchased from Integrated DNA technologies (IDT) with a sequence of 5'-GUGUAAGCUUUU-3'. CD experiments were carried out as described in Fourmann et al, (2013) and ellipticity was monitored at 208, 222 nm, to determine helical content. To determine B sheets content ellipticity was monitored at 190, 212, and 128nm. The optimal protein concentration of hLa 225-332 CTD fragment for CD experiments was determined to be 2uM to 5uM. 4 total spectra were acquired for each protein-RNA interaction on a Jasco J-810 instrument at 20°C with a Quartz two compartment cuvette (Precision Cells Inc.) and ellipticity was measured in milldegrees (mdeg). A 1 to 1 stoichiometry of protein to RNA was used and the final protein and RNA concentration in the last spectra was 1uM while the first three spectra had a 2uM protein or RNA concentration, respectively. Assays were carried out in CD Buffer (10mM NaH2pO4 pH 7.5, 100mM mM NaF). Spectra were recorded from 170 – 320 nm with a scan rate of 100 nm/min and a 1.0 nm bandwidth. First, the protein solution was placed in one compartment and only buffer in the second one, and then RNA solution was placed in one compartment and only buffer in the second one to obtain baseline values. Protein and RNA were then put in separate compartments to obtain a third spectrum. Finally, protein and RNA was mixed at 1uM concentration, and spectra were recorded after 5 minutes. If there were no interaction between the

protein and RNA, spectra four would be similar to Spectra 3, however if there were interaction scan 3 and four would not correspond (Fig.8). The cuvette was rinsed with ddH2O followed by rinsing 70% or ethanol before each scan and stored in 70% nitric acid (Sigma).

2.3.1 Secondary structure analysis

Secondary structure analysis was done using Dichro-Web Online CD Analysis software using CONTIN set 4 (190nm-240nm) optimized analysis program (Whitmore & Wallace, 2004; Whitmore & Wallace, 2008). CONTIN is a database that encompasses X-ray crystallography of 16 proteins whose secondary structures are known. Spectra of hLa CTD fragments were compared to these 16 proteins and the theoretical best fit curve was determined (Greenfield, 2006). Spectra 2 (RNA alone) was subtracted from spectra 4 (Protein-RNA interaction) to determine the interactions baseline value. The input units, Machine units θobs (millidegrees; mdeg) was converted to delta epsilon Δε by calculating the mean residue weight (MRW; = molecular mass/(N – 1); N = number of amino acids).
(http://dichroweb.cryst.bbk.ac.uk/html/home.shtml).

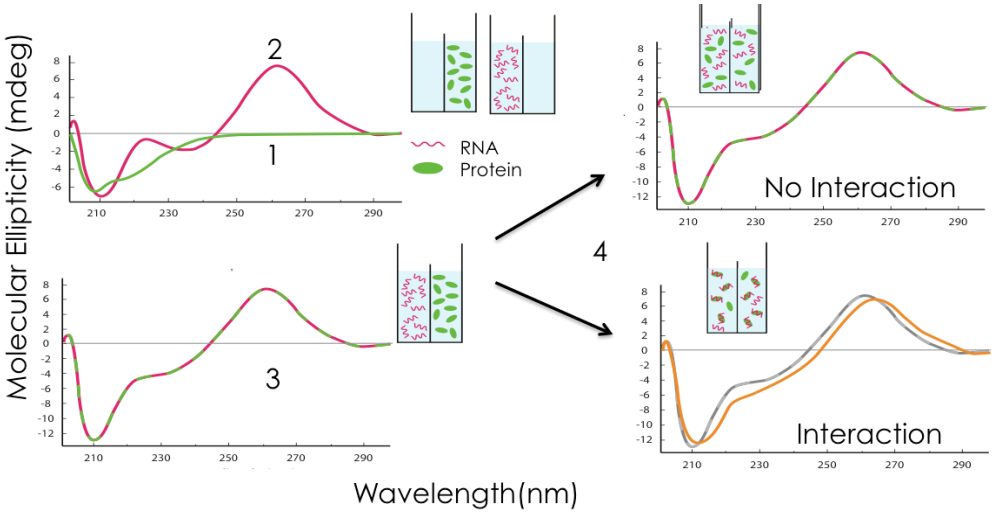


Figure 8. Circular Dichroism Spectroscopy methodology using a Quartz two compartment cuvette. 1) First, the protein solution was placed in one compartment and only buffer in the second one. 2) RNA solution was placed in one compartment and only buffer in the second one to obtain base line values. 3) Protein and RNA were then put in separate compartments to obtain a third spectrum. 4) Finally, protein and RNA was mixed at 1uM concentration, and spectra were recorded after 5 minutes. If there were no interaction with between protein and RNA, spectra four would be similar to spectra 3, however if there were interaction scan 3 and four would not correspond. Adapted from from Anne-Sophie Tillault's Ph.D. thesis in the lab of Bruno Charpentier, Université de Lorraine, 2014.

2.4 Assessing CTD fragments affinity by Electrophoretic Mobility Shift Assay

Master mix was prepared for 15 lanes per gel with EMSA Buffer (200mM Tris pH 7.6, 1mM BME, 0.1M EDTA 50 to 300mM KCl) and 3000 Ci/mmol of γ-32P labeled 12nt RNA and was incubated at 95°C for 5 min, and slow cooled to 37°C. 50mM KCl was used for low salt conditions and 300mM KCl was used for high salt conditions. Then bromophenol blue dye and one µl RNase Inhibitor was then added. Stock protein was prepared by serial dilution using La Dilution Buffer (LDB) (1X EMSA and 80% glycerol). The master mix was incubated with various

amounts of hLa proteins at 37°C for 20 minutes, chilled on ice for 5 minutes, and separated on a 10% Native Separating Gel (1X TBE, 40% polyacrylamide, ddH2O up to a final volume of 10 mL, 0.08% APS, 0.037% TEMED) at 100V and 4°C for 1 hour. The gel was then dried for 20 minutes using a gel drier (Biorad) and was exposed to the phosphorimager plate (film) overnight. The next day the gel was scanned using Phosphor imager Typhoon Trio+ (GE Healthcare). Gel images were quantified using the Imagequant TL (Amersham) software, and the Dissociation Constant (Kd) was calculated by measuring the disappearance of free RNA in subsequent lanes and was calculated by plotting values into GraphPad Prism 5 (Graphpad software).

2.4 Probe labelling and purification

Electrophoretic mobility shift assay (EMSA) was carried out to evaluate the ability of hLa CTD fragments to bind target RNA. The 12nt trailer RNA was radiolabeled with γ-32P ATP (PerkinElmer) by a T4 Polynucleotide Kinase reaction. The reaction consisted of 1X T4 Polynucleotide Kinase Buffer (NEB), two µL γ-32P ATP (PerkinElmer), 10U of T4 Polynucleotide Kinase (NEB), 25µM RNA Sequence (IDT), and ddH2O. The total reaction volume was 20 µL and was incubated for 2 hours at 37°C. Then, 10µL of 3X RNA Loading Buffer (80% deionized formamide, 10 mM EDTA, 0.06% bromophenol blue, 0.06% xylene cyanol) was added and loaded on a12% Urea gel (12%Polyacrylamide, 1X TBE, 7M Urea, ddH2O, 0.1% APS, and 0.05% of TEMED) and electrophoresed at 100V for 1 hour.

The gel was then was exposed to a HyBlot CL Autoradiography Film (Denville Scientific), next to a GLOGOS II Autocad Marker (Stratagene). The film was then developed, and the labeled marker was excised from the gel and placed in a solution of 0.5M NaCl overnight at room temperature. Activity level was measured by using a 300 XL scintillation counter (HiDex).

2.5 1H-15N NMR Spectroscopy of hLa CTD fragments

HLa CTD mutants 225-332 and 225-408 were generated for 1H-15N heteronuclear single quantum coherence (HSQC) spectroscopy. BL21 *E. coli* cells expressing hLa protein fragments were grown at 37°C in 1.5L of M9 minimal media (33.7 mM Na2HPO4, 22mM KH2PO4, 8.55 mM NaCl, and 20% Glucose, 1X trace elements, 1 mM MgSO4, 100 µM CaCl2) supplemented with 10µg/µl Kanamycin and 1.5g/L (15N) Ammonium Chloride as the sole nitrogen source. Cells were grown to exponential phase reaching an OD600nm of 1.0, followed by 1mM IPTG induction for 18 hours at 16°C for soluble hLa protein fragment production. Cell pellets were collected by centrifugation at 5000rpm for 10 minutes at 4°C (Beckman) and resuspended in 30mL of 1X T300 (10mM Tris-HCl pH 8.0, 300mM NaCl) supplemented with 1X Halt Protease Inhibitor Cocktail (PIC; Thermo Scientific) and 1mM PMSF (Phenylmethylsulfonyl fluoride) Protease Inhibitor. The cell suspension was subject to three passes of homogenization at 15000psi using a French cell press. The lysate was cleared by centrifugation at 30000rpm (Beckman) for 30 minutes. The extract was purified over 10mL of Ni-NTA resin (Qiagen) followed by sequential

washes with 1X T300 and 1X T300 supplemented with 2M urea, 1M Urea, and finally 40mM imidazole. The proteins were eluted in 80mL of 1X T300 supplemented 300mM imidazole. The elutant was concentrated to 5mL by ultrafiltration (10K Amicon Ultra 15mL centrifugal filter; Millipore) and injected into a gel filtration system (Fast protein liquid chromatography; FPLC) using a Sephacryl S100 16/60 column (GE Biosciences) equilibrated with 20mM sodium phosphate buffer pH 7.0, 250mM NaCl. Fractions of interest were analyzed on a 10% SDS-PAGE for verifying purity. The purest fractions were pooled by ultrafiltration (10K Amicon Ultra 15mL centrifugal filter; Millipore) and concentrated to 0.15 mM. Samples were analyzed by 10% SDS-PAGE alongside BSA standards to determine molar concentrations. Samples were visualized using Coomassie staining in stain solution for 60 minutes and destain solution overnight (see section 2.2). Samples were supplemented with 10% D20, and protein-RNA complex was formed using a 1 to 1 stoichiometry of protein to RNA with 12nt Arginine pre-tRNA trailer purchased from Dharmacon. HSQC spectra were acquired on a 700MHz Bruker Ascend 700. Spectra were processed using NMR Pipe and NMR Analysis and compared to a solution structure of the C-terminal RRM of human La (La225-334; PDB 1OWX). A heat map of the NMR spectra was generated to visualize the degree of chemical shift perturbation. Firstly, a weak chemical shift perturbation was assigned between 0.015 – 0.0199 ppm. Secondly, a medium strength chemical shift perturbation was assigned between 0.02 – 0.0299 ppm. Finally, a strong chemical shift perturbation was assigned as any difference greater than 0.03 ppm.

3. Results

3.1 Protein purification

To test the C terminal domain (CTD) dependent RNA binding of hLa we generated several mutants encompassing the RRM2 and subsequent C terminal regions. 225-332, 225-339, 225-344, 225-363, 225-375, and 225-408 hLa CTD protein fragments were subcloned into pET28a expression vectors and purified over a pre-packed Histrap Ni-sepharose column. 2µg of purified of recombinant hLa proteins were electrophoresed on a 10% SDS-PAGE gel, which was Coomassie stained to verify sizes and concentrations (Fig.9). Notably, 225-363 hLa CTD mutant was soluble, however prone to rapid degradation following purification. Furthermore, there is also a presence of residual background bands in the 225-375 hLa construct.

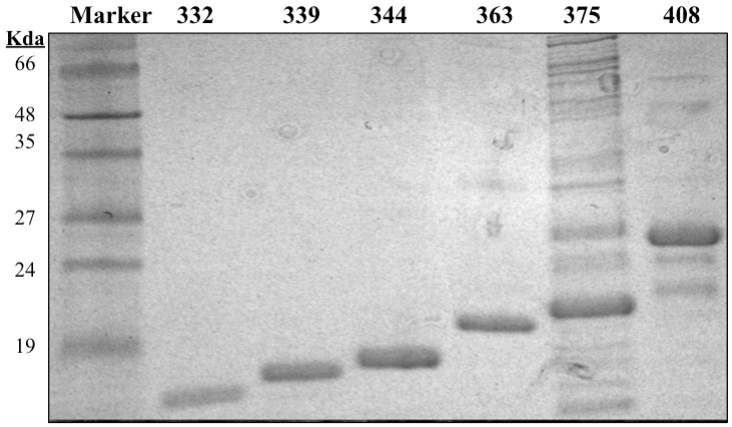


Figure 9. 2µg of purified hLa CTD protein fragments resolved on a 10% SDS-PAGE gel. 225-332, 225-339, 225-344, 225-363, 225-375, and 225-408 hLa CTD protein fragments were subcloned into pET28a expression vector and purified over a pre-packed Histrap Ni-sepharose column. Molecular weight marker in kilo Daltons (kDa) is depicted on the left.

3.2 The RRM2s C terminal extension is required for RNA binding

The truncated hLa CTD fragments were constructed to test their ability to form complexes with RNA. RNA binding via RRM2 has been documented previously and as shown to independent of the La module's activity (Horke et al., 2002; Horke et al., 2004; Martino et al., 2012). More recently, regions outside of the RRM2 have been proposed to assist with RNA binding (Singh et al., 2012; Kuehnert et al., 2015). To evaluate the affinity of hLa CTD fragments we conducted electrophoretic mobility shift assay (EMSA) on a radiolabelled 12 nucleotide (nt) pre-tRNA Arginine ACG (tRNAArgACG) trailer (12nt trailer) that ends with UUU-3'OH. This method is known to measure the affinity of ribonucleoprotein (RNP) complexes. Increasing amounts of hLa were added to trace amount of 32P-labelled RNA. If there is a protein-nucleic acid interaction, it will retard the movement of the complex on a polyacrylamide gel, resulting in a gel shift. We also tested the electrostatic sensitivity of hLa CTD fragments with respect to salt dependence by performing EMSAs at various KCl concentrations. We predict that hLa CTD fragments interact electrostatically with the negatively charged RNA phosphate backbone and this will interaction will be hindered by increased salt (KCl). 50mM KCl was used for low salt conditions and 300mM KCl was used for high salt conditions.

There appears to be no shifting for 225-332 hLa CTD fragment, which lacks an SBM and NLS (Fig.10A, 10B). There is some RNA decay, however this was taken into account when quantitating EMSAs (Tbl.2). This is consistent with the idea that RRM2 cannot bind RNA by itself, similar to what may be expected if the RRM2 of

hLa binds RNA homologous to the RRM2 of p65 (Singh et al., 2013).

Longer CTD constructs (225-339, 225-344) can engage RNA targets, albeit with reduced affinity compared to the full-length CTD (Fig. 11, Fig. 12; Fig.14). When comparing 225-339 and 225-344, mutants that contain incrementally more SBM, we see a two-fold increase in La affinity with the longer constructs (Fig. 11, Fig. 12). In both cases, the proteins favour low salt conditions with shifts starting to occur at 4uM (339) and 2uM (344) respectively. A fully bound RNP complex does not form with higher protein concentrations, which may be indicative of transient binding occurring. Reactions did not reach saturation therefore precise affinities were not determinable (Tbl.2).

There is a significant increase in affinity using the 225-375 construct (Fig. 13). In low salt conditions shifting begins at 500nM and in high salt shifting drops two fold. There is drastic difference in salt dependency for 225-375 in our quantitative analysis , which is less pronounced on the gel (Fig.13A; Fig 13B; Tbl.2). Because a fully bound RNP does not form and the reaction does not proceed to 100% bound, precise affinity quantifications were not determinable. The full-length CTD (225-408) was salt insensitive as binding affinity improved in higher salt conditions. Additionaly, a fully bound RNP complex forms at 1uM in high salt conditions (Fig. 14).

In brief, all constructs, except for the full length CTD have salt dependent binding as we see a drop in affinity from low to high salt. Because of the relatively high number of basic amino acids in regions between 332-375, we predict that hLa

CTD fragments interact electrostatically with the RNA phosphate backbone, which is an indication of binding mediated by salt (Davidovich et al., 2013). In contrast, regions between 375-408 of hLa have been proposed to have a high conservation of hydrophobic (salt insensitive) contacts. As a result, binding becomes stronger and salt-insensitive when regions 375-408 are included.

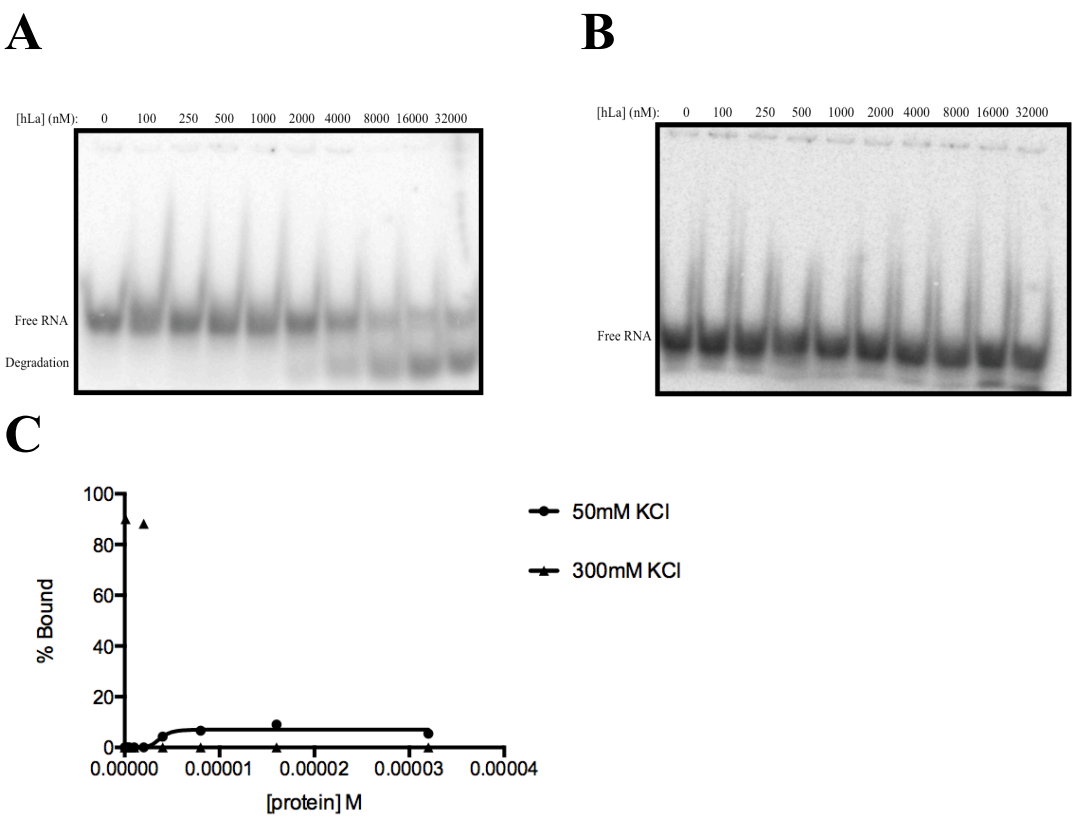


Figure 10. 225-332 EMSA results of hLa CTD fragments binding radiolabelled 12nt trailer performed under different KCl concentrations. EMSA were carried out to evaluate the binding affinity of RNP complex of hLa CTD protein fragments and 12nt trailer. Concentrations of 225-332 are as follows 0nM, 100nM, 250nM, 500nM, 1000nM, 2000nM, 4000nM, 8000nM, 16000nM, and 32000nM. Samples were electrophoresed on a 10% Native gel A) 50mM KCl. B) 300mM KCl. There appears to be no shifting for this construct, which lacks an SBM and NLS. C) Gel images were quantified using the Imagequant TL (Amersham) software, and the equilibrium constants (*K*d) was calculated by measuring the disappearance of free RNA.

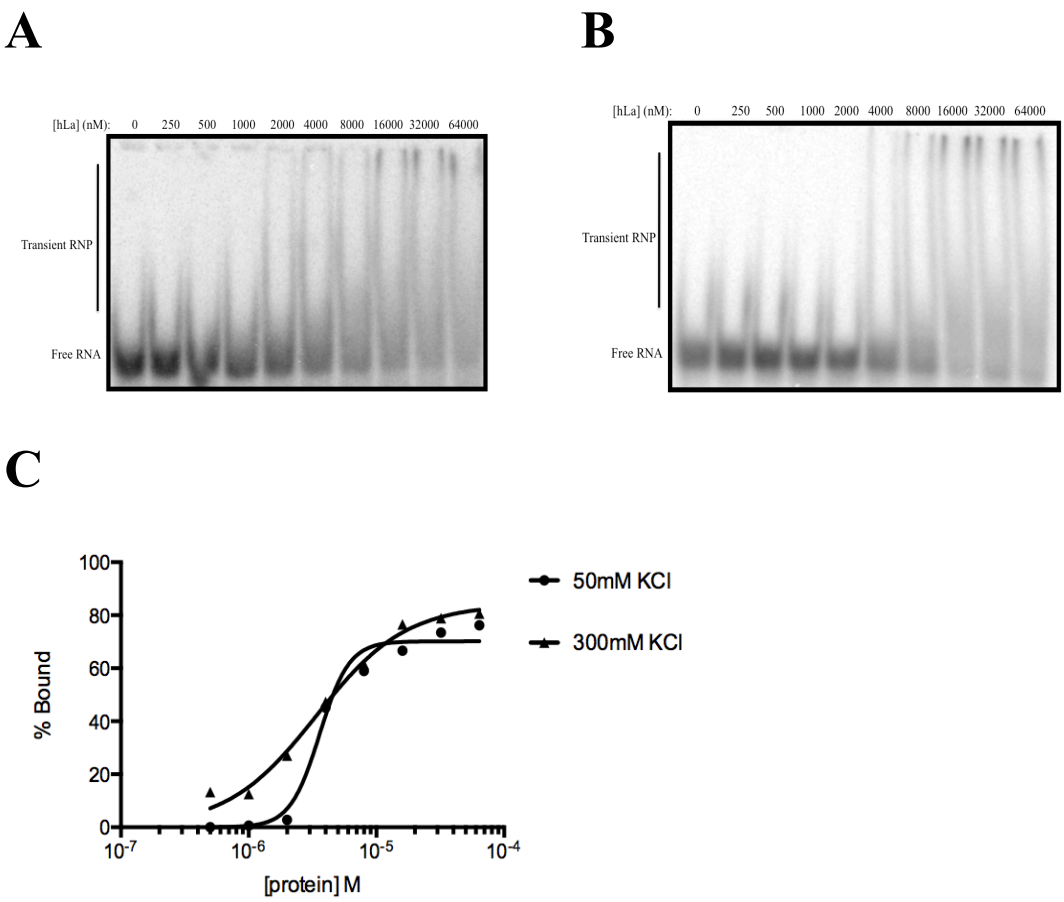


Figure 11. 225-339 EMSA results of hLa CTD fragments binding radiolabelled 12nt trailer performed under different KCl concentrations. EMSA were carried out to evaluate the binding affinity of RNP complex of hLa CTD protein fragments and 12nt trailer. Increasing amounts of hLa were added to trace amount of 32P-labelled RNA. Concentrations of 225-339 are as follows 0nM, 250nM, 500nM, 1000nM, 2000nM, 4000nM, 8000nM, 16000nM, 32000nM, and 64000nM. Samples were electrophoresed on a 10% Native gel. A) 50mM KCl, B) 300mM KCl. 225 – 339: contains only part of the SBM. Shifting occurs at 2uM in low salt conditions and 4uM in high salt conditions. C) Gel images were quantified using the Imagequant TL (Amersham) software, and the equilibrium constants (*K*d) was calculated by measuring the disappearance of free RNA.

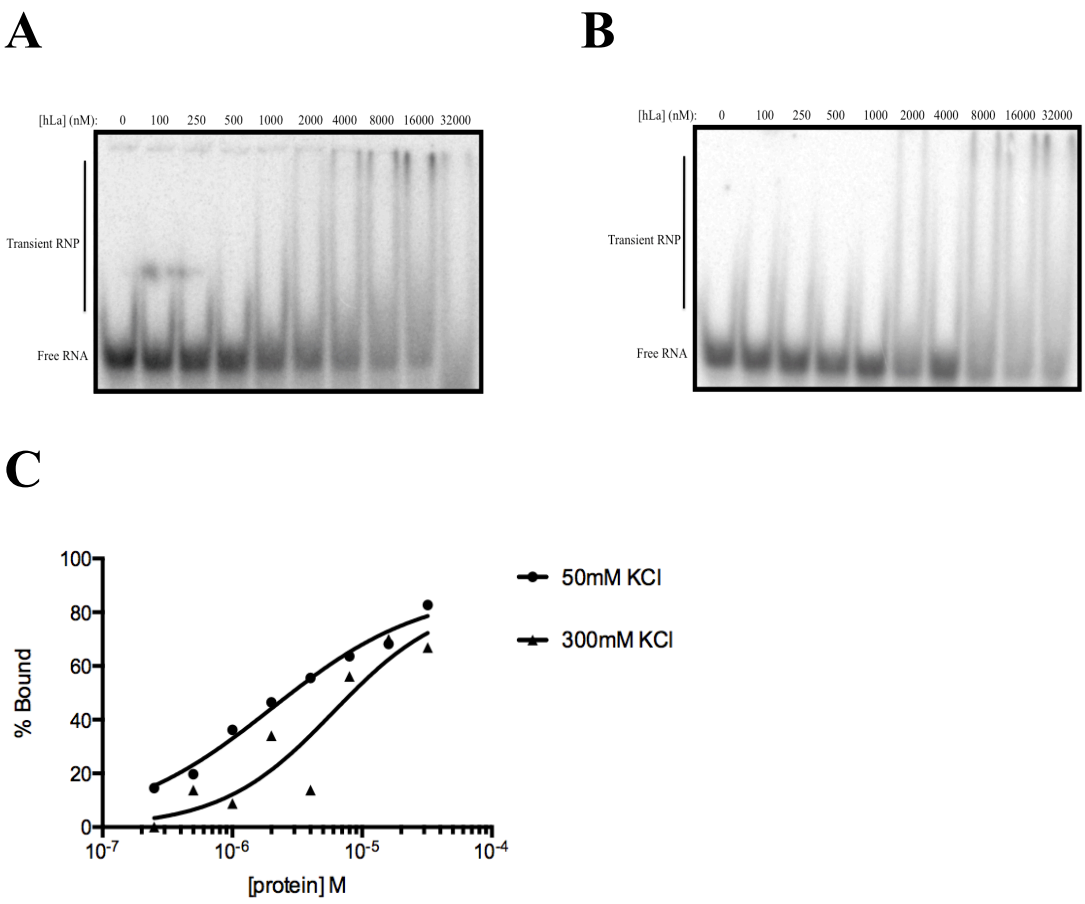


Figure 12. 225-344 EMSA results of hLa CTD fragments binding radiolabelled 12nt trailer performed under different KCl concentrations. EMSA were carried out to evaluate the binding affinity of RNP complex of hLa CTD protein fragments and 12nt trailer. Increasing amounts of hLa were added to trace amount of 32P-labelled RNA. Concentrations of 225-344 are as follows 0nM, 100nM, 250nM, 500nM, 1000nM, 2000nM, 4000nM, 8000nM, 16000nM, and 32000nM. Samples were electrophoresed on a 10% Native gel. A) 50mM KCl, B) 300mM KCl. 225 – 344: containing the RRM2 and whole SBM. Shifting occurs at 1uM in low salt conditions and 2uM in standard and high salt conditions. C) Gel images were quantified using the Imagequant TL (Amersham) software, and the equilibrium constants (*K*d) was calculated by measuring the disappearance of free RNA.

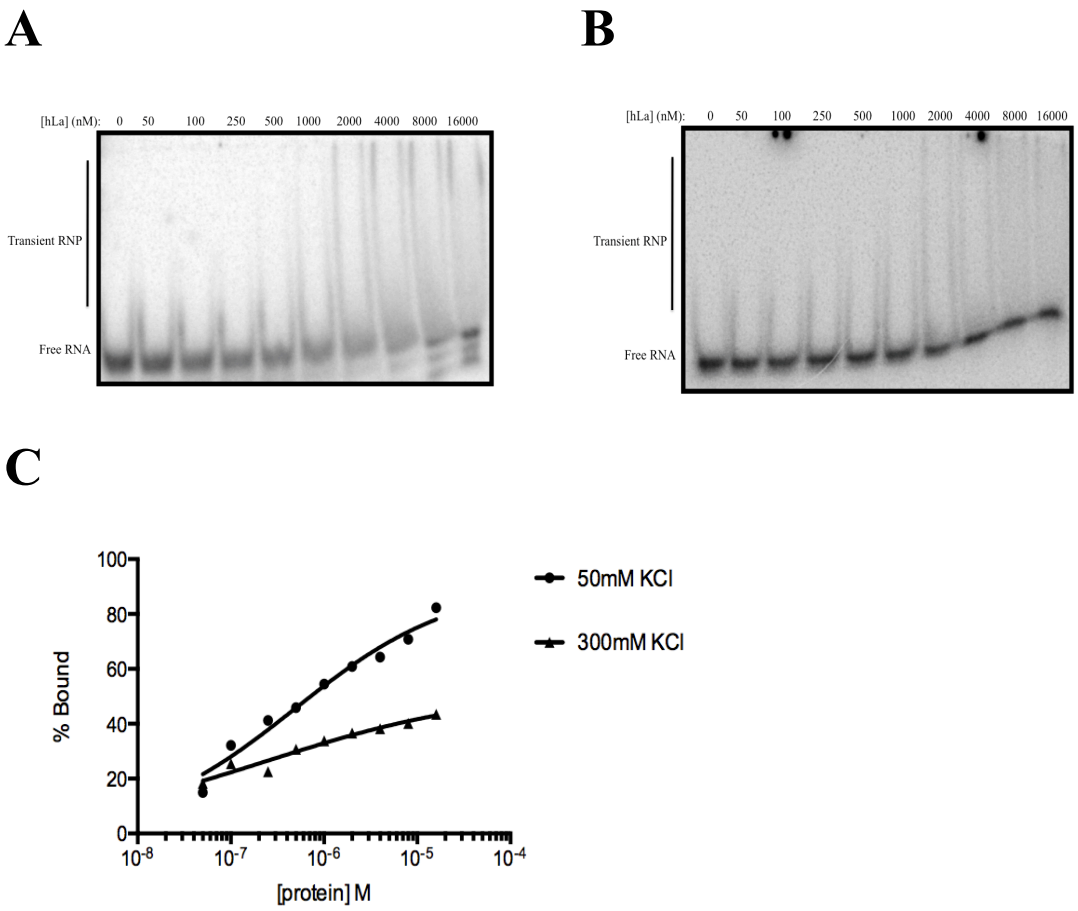


Figure 13. 225-375 EMSA results of hLa CTD fragments binding radiolabelled 12nt trailer performed under different KCl concentrations. EMSA were carried out to evaluate the binding affinity of RNP complex of hLa CTD protein fragments and 12nt trailer. Increasing amounts of hLa were added to trace amount of 32P-labelled RNA. Concentrations of 225-375 are as follows 0nM, 50nM, 100nM, 250nM, 500nM, 1000nM, 2000nM, 4000nM, 8000nM, and 16000nM. Samples were electrophoresed on a 10% Native gel A) 50mM KCl, B) 300mM KCl. 225 – 375: containing the RRM2 and whole SBM, and S366 phosphorylation site. Shifting starts at 500nM and in both low and high salt. C) Gel images were quantified using the Imagequant TL (Amersham) software, and the equilibrium constants (*K*d) was calculated by measuring the disappearance of free RNA.

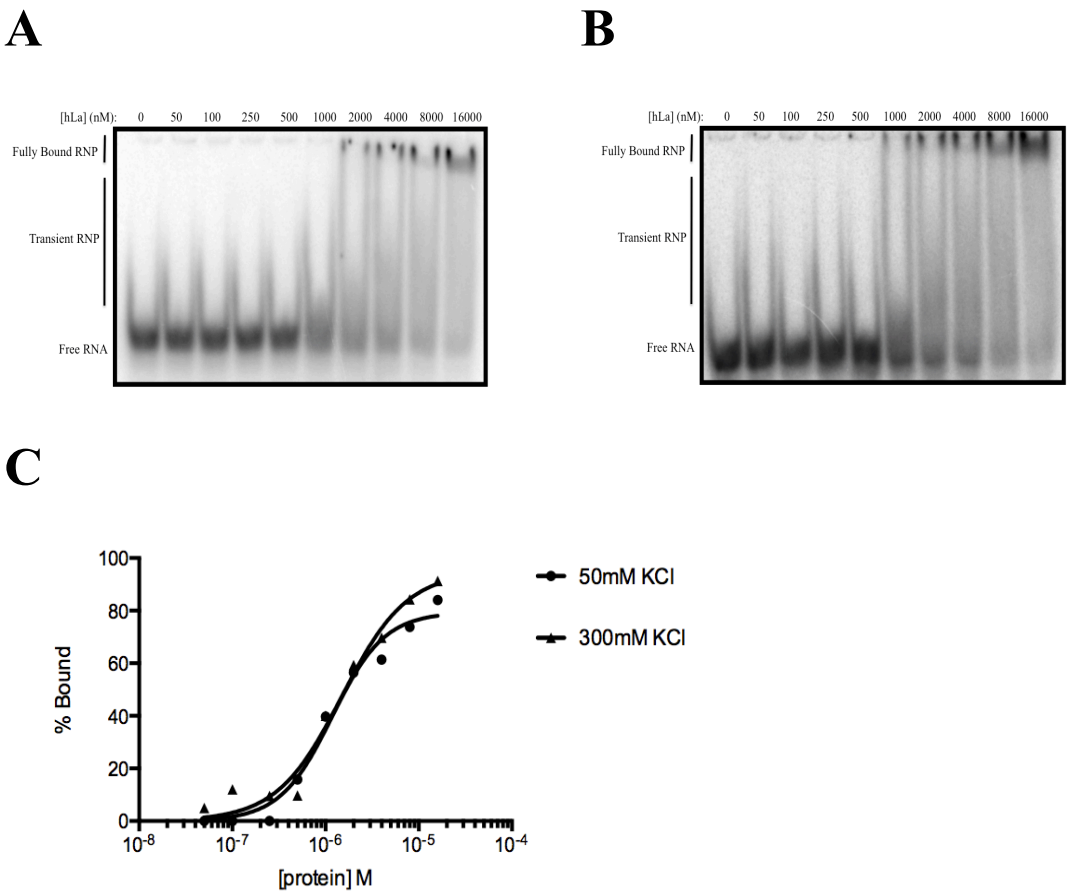


Figure 14. 225-408 EMSA results of hLa CTD fragments binding radiolabelled 12nt trailer performed under different KCl concentrations. EMSA were carried out to evaluate the binding affinity of RNP complex of hLa CTD protein fragments and 12nt trailer. Concentrations of 225-408 are as follows 0nM, 50nM, 100nM, 250nM, 500nM, 1000nM, 2000nM, 4000nM, 8000nM, and 16000nM. Samples were electrophoresed on a 10% Native gel. A) 50mM KCl, B) 300mM KCl.) 225 – 408: Full-length CTD starts to shift RNA at 500nM and completes shifting at 1uM low salt conditions. High salt conditions increases binding affinity by two fold. C) Gel images were quantified using the Imagequant TL (Amersham) software, and the equilibrium constants (*K*d) was calculated by measuring the disappearance of free RNA.

Table 2. Quantitative representation of affinity of hLa mutants. RNA binding was assessed under two different salt conditions, 50mM KCl and 300mM KCl. This table estimates the equilibrium constant values (*K*d) in nM of hLa constructs in low and high salt conditions. Gel images were quantified using the Imagequant TL (Amersham) software, and the equilibrium constants (Kd) was calculated by measuring the disappearance of free RNA.

	*K*d (nM)	
Protein	Binding in 50mM KCl	Binding in 300mM KCl
225-332	<32000	<32000
225-339	3594	3393
225-344	2026	6008
225-375	1948	<32000
225-408	1496	1180

3.3 Increased α-helical content may mediate Protein-RNA interaction

Regions outside the canonical RRM fold have been shown to participate in RNA binding (Afroz et al., 2015). The entropy transfer model postulates that intrinsically disordered regions transition between ordered and unordered states in order to to stabilize the protein/RNA complexes (Tompa & Csermely, 2004). Further evidence in support of this model is p65's RRM2 (Singh et al., 2012). Therefore, in order to determine structural changes in the C terminal disordered SBM of hLa we performed far UV Circular Dicroishm (CD) Spectroscopy. This method relies on detecting the difference in light absorbance of left and right circularly polarized light. CD Spectroscopy is a valuable technique that can estimate the secondary structure of chiral molecules, such as proteins, allowing it to identify the type and amount of the secondary structure present in a protein. CD readings above 240 nm can be used to observe nucleic acids and CD readings between 200nm and 240nm are indicative of

amides of the protein backbone (Greenfield, 2006). Negative ellipticities at 208 nm and 222 nm correspond to α-helical structures and can be a parameter for amount of α-helices present. Positive ellipticities at 212 nm and negative ellipticities at 190 nm can be a parameter for amount of β-turns. Experiments were carried out as described in Fourmann et al, 2013. This method was originally designed to test the changes in RNA structure. Because nucleic acids also make a significant contribution in the 200–240 nm range, CD spectra for the RNP interactions can be obscured by the RNA. As such, we qualitatively described changes in protein structure upon RNA binding.

We tested various RNA substrates including 10mer and 20mer homopolymers of adenine (A), cytosine (C), guanine (G), or uracil (U) ssRNA in addition to a 21mer sequence ssRNA with or without a terminal 3' phosphate and noted that the hLa's RRM2 binds these sequences indiscriminately (data not shown). However, these RNAs obscured our protein signal between 200 and 240nm. We used 12nt trailer because this RNA makes negligible contributions to the signal. The 12nt trailer gave the lowest signal in the CD compared to the other RNAs, as we are mainly interested in structural changes of the protein. Because signal intensity of CD is directly proportional to the concentration of sample, we determined the optimal concentration of hLa constructs for Circular Dichroism experiments to be 2uM to 5uM (Fig. 15).

In our CD experiments we scanned four different conditions for each protein (see section 2.3). In scan three, Protein and RNA were then put in separate compartments to obtain baseline values. In contrast, scan four represents a mixture of the protein and RNA in equal molar ratios. If there were no interaction with between

protein and RNA, spectra would be similar, however if there were interaction between the protein and RNA, the spectra would not correspond. Qualitatively, we see an interaction between protein and RNA in all constructs, except for 225-332 as there is no spectral overlap between scans 3 (No interaction) and scan 4 (Interaction). As indicated by Figure 9, there was a presence of residual background bands in the 225-375 hLa protein purification, and therefore spectra could not be obtained for this sample as the purity of samples affects signal and samples must be at least 95% pure (Greenfield, 2006).

We then determined secondary structure changes in bound versus unbound samples (Fig. 16; Tbl.3). The RNAs baseline spectrum was subtracted from the Protein-RNA interaction's spectrum to account for RNA signal within the 200 to 240 nm region. Secondary structure predictions indicate that 225-332 becomes 1.1% more α helical upon RNA binding (Fig.16A). 225-339 becomes 3.7% more α helical upon RNA binding (Fig.16B,), while 225-344 becomes 5% more α helical upon RNA binding (Fig.16C). The full-length CTD becomes 17% more α helical upon binding to RNA (Fig.16D). Our results therefore indicate that constructs encompassing the SBM (225-339, 225-344, and 225-408) have significantly more α-helical content upon complex formation. Although, these values support our predictions they are not consistent with Jacks et al, 2003, reported structures of hLa 225–334 and 225–408. Therefore, we could not conclude if proteins become more α helical to mediate entropy transfer.

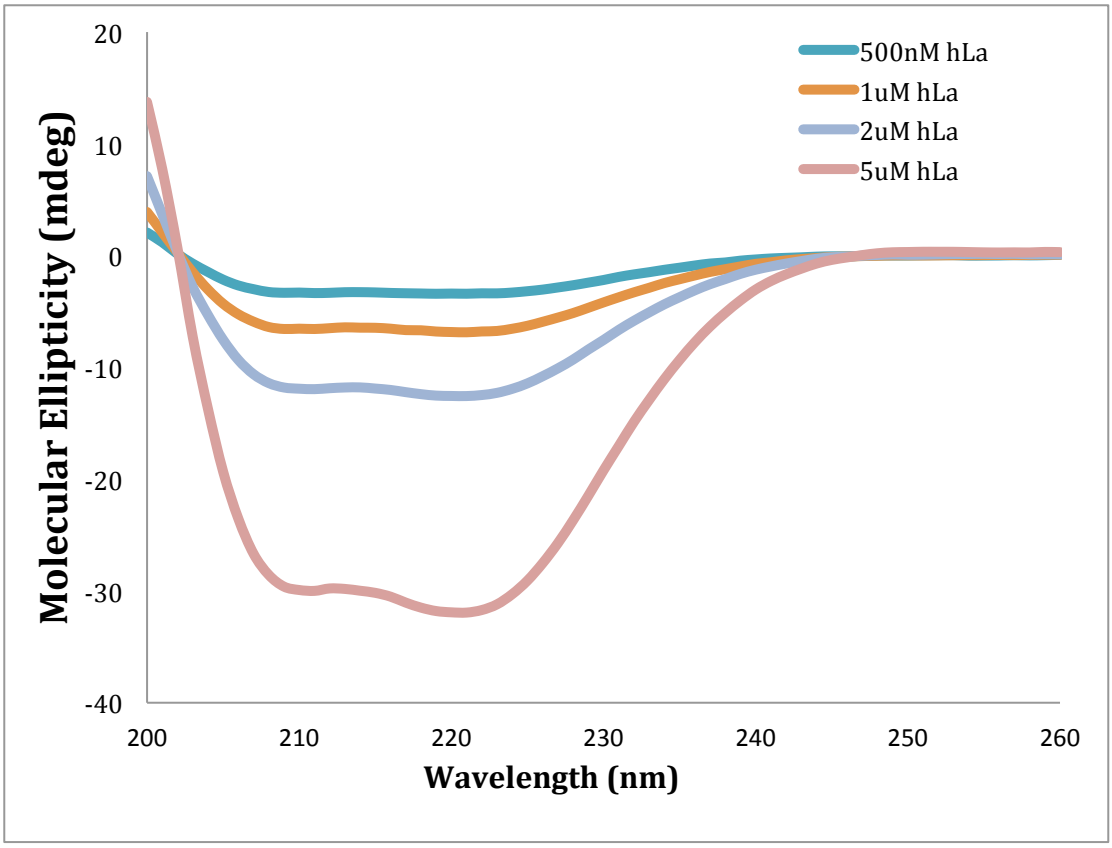


Figure 15. 2uM to 5uM is the optimal protein concentration of hLa 225-332 CTD fragment for Circular Dichroism. 500nm, 1uM, 2uM, and 5uM of 225-332 CTD fragment was used in a circular Dichroism assay. This figure demonstrates that 2uM to 5uM is the optimal concentration to use for CD, as there is a decrease in the two minima (208nm and 222nm) at lower concentrations.

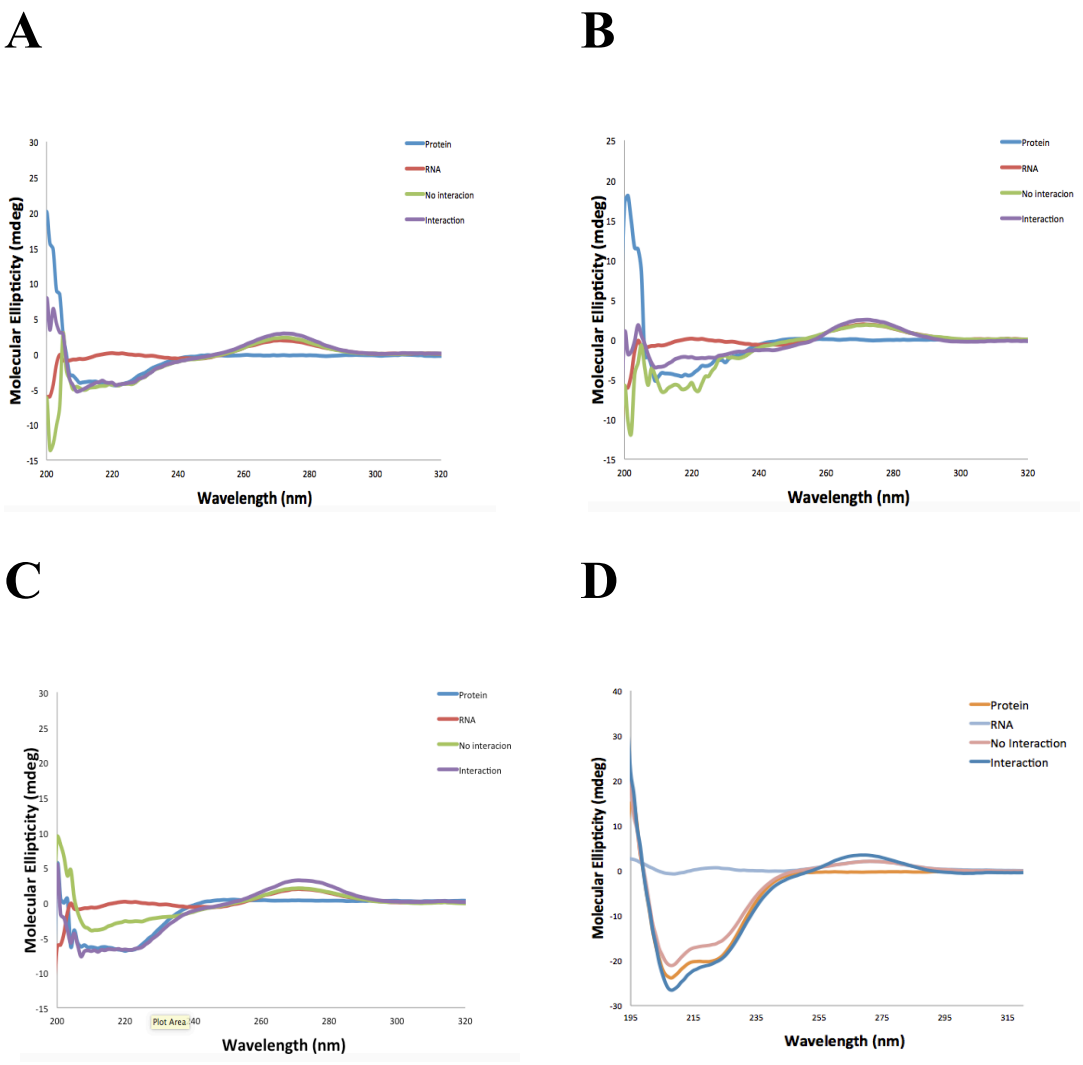


Figure 16. Circular Dichroism spectra of hLa CTD fragments interacting with 12nt trailer. A) 225-332. B) 225-339. C) 225-344. D) 225-408. 4 spectra were acquired at 2uM concentrations of protein and RNA and the final spectra was acquired with 1uM concentrations of protein and RNA in CD Buffer for each figure. 1) First, the protein solution was placed in one compartment and only buffer in the second one. 2) RNA solution was placed in one compartment and only buffer in the second one to obtain base line values. 3) Protein and RNA were then put in separate compartments to obtain a third spectrum. 4) Finally, protein and RNA was mixed at 1uM concentration, and spectra were recorded after 5 minutes. If there were no interaction with between protein and RNA, spectra four would be similar to Spectra 3, however if there were interaction spectra 3 and 4 would not correspond.

Table. 3. Secondary structure analysis from Circular Dichroism spectra of hLa CTD fragments interacting with 12nt trailer. Secondary structure analysis was done using Dichro-Web Online CD Analysis software using CONTIN set 4 (190nm-240nm) optimized analysis program (Whitmore & Wallace, 2004; Whitmore & Wallace, 2008). Spectra 2 (RNA alone) was subtracted from spectra 4 (Protein-RNA interaction) to determine the interactions baseline value and compared to protein baseline value (Spectra 1). Results are presented in % of structural composition.

	225-332		225-339		225-344		225-408	
Structure	-RNA	+RNA	-RNA	+RNA	-RNA	+RNA	-RNA	+RNA
α	0.072	0.083	0.046	0.083	0.052	0.102	0.27	0.454
β	0.1727	0.14375	0.191	0.14375	0.121	0.1535	0.125	0.0057
turn	0.1805	0.215	0.2105	0.215	0.224	0.189	15.7	0.008
unordered	0.3285	0.3305	0.3155	0.3305	0.4285	0.299	28.45	0
Change	0.0105		0.037		0.05		0.176	

3.4 The RRM2 of hLa binds 12nt trailer RNA non-canonically

The RRM2 of hLa is atypical in that it features an additional α-helix and β-sheet, as well as a negative electrostatic surface potential and its mechanism of binding remains unclear (Jacks et al., 2003). The precedent of atypical binding already had been set as p65's (LARP7) RRM2 is homologous to hLa (Singh et al, 2013). Therefore, to structurally characterize the interaction between hLa CTD fragments and 12nt trailer RNA we used Nuclear Magnetic Resonance (NMR) spectroscopy. NMR relies on the magnetic properties of atomic nuclei and can determine 3D structural information like α-helix formation, and binding kinetics. By incorporating isotopes like ^{1}H (Hydrogen), ^{15}N (Nitrogen), and ^{13}C (Carbon) into proteins, we can determine the identity and spatial organization of amino acids within a protein. HLa CTD protein fragments were generated for 1H-15N Heteronucler

Single Quantum Coherence (HSQC) (Fig. 17). Both 225-332 and 225-408 constructs bound to RNA (Fig. 18). However there was severe spectral overlap for the full length CTD (225-408) as indicated by the HSQC, which prevented sequence-specific assignments (Fig. 18B).

Nevertheless, we were able to assign chemical shifts from the 225-332 HSQC spectra. There is an interaction of 225-332 in the NMR with, but this is not seen in the EMSA (Fig. 18A; Fig 10). We use equal molar concentrations of protein to RNA in the NMR, which saturates the interaction to ensure that complex formation can be detected. In the EMSA we use considerably less RNA than protein to quantitatively detect binding affinity. Binding may occur because contact residues in the RRM2 are basic and positively charged and have an electrostatic reaction with the negatively charged RNA. The globular structure of the protein changes upon binding, which can be seen by chemical shift splitting and peak disappearance, however the identity of contact residues remains unclear. In order to determine contact residues we must perform a mutational analysis on hLa. If the residue is necessary for binding, mutating it should abolish an interaction.

We also tested A20 homopolymer and a 21mer ssRNA sequence with or without a terminal 3' phosphate (data not shown). While fragment 225 to 408 bound A20 strongly, preliminary data suggest it did not discriminate significantly between the presence or absence of a terminal phosphate. We show the change in chemical shift perturbations upon ligand binding for 225-332 (Fig. 19). Comparison to the unbound 225-332 shows a global conformational change and qualitatively there is no

peak movement but some line broadening (peak disappearing) in the bound sample versus the unbound control.

We mapped all the perturbations onto a previously solved solution structure of the C-terminal RRM of human La RRM2 structure (La225-334; PDB 1OWX; Jacks, et al., 2003) and assigned the appropriate threshold strength (degree of movement) for each perturbation (Fig. 20). A heat map of the NMR spectra was generated to visualize the degree of chemical shift perturbation. Firstly, a weak chemical shift perturbation was assigned between 0.015 – 0.0199ppm. Secondly, a medium strength chemical shift perturbation was assigned between 0.02 – 0.0299ppm. Finally, a strong chemical shift perturbation was assigned as any difference greater than 0.03ppm.

Structurally the α3 helix, that is obscuring the binding surface is has high degree of movement and amino acids in the RNA binding surface (RNP1 and RNP2; β3 and β1 sheets) change. The highest degree of movement is situated around the β2 sheet, β2β3 loop, and α3 helix, away from the canonical RNA binding surface (β3 and β1 sheets). The structure also indicates that the inner core containing RNP1, Glycine 257, Isoleucine 262, Arginine 266, and Glycine 267 have high threshold strength, which is mapped onto β2 and β2β3 loop. Consequently, The loop region between β2β3 loop3 consisting of a polar Arginine 266 (R266) may be responsible for interacting with the 12nt trailers body. Unexpectedly, movement also occurs around the β4 sheet, which is away from the RNA binding surface. Although this data gives us insight about RNA binding, we must label longer constructs with $^{13}C^{15}N$ to determine if structural changes occur in the SBM upon RNA binding.

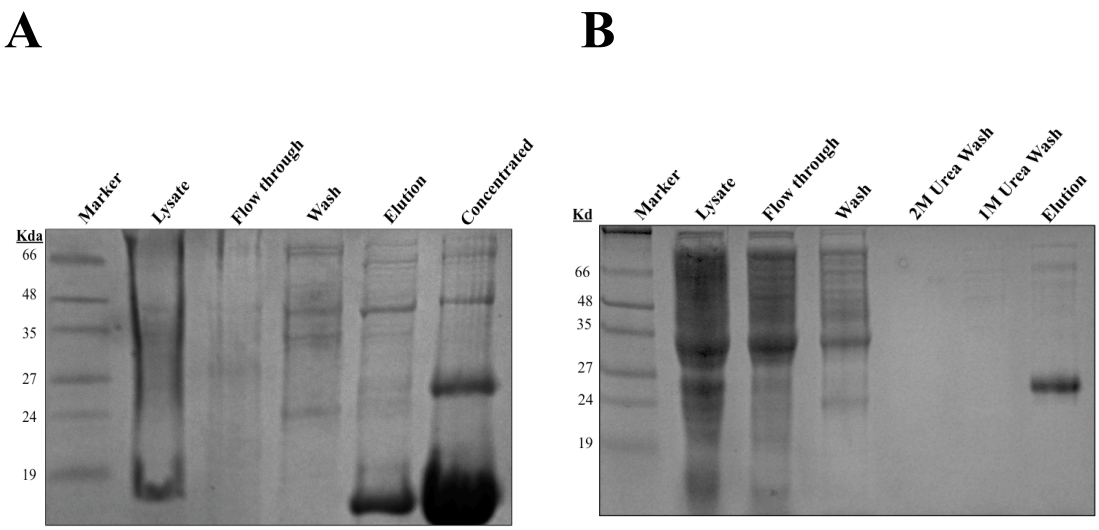
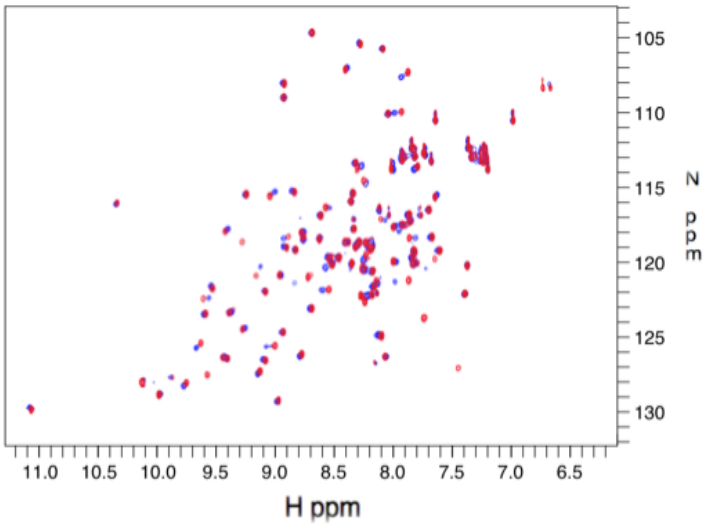


Figure 17. Purified 15N labeled hLa CTD protein fragments resolved on a 10% SDS-PAGE gel. A) 225-332. B) 225-408. HLa CTD protein fragments were subcloned into pET28a expression vector and purified over a Ni-NTA resin followed by sequential washes with 1X T300 (Wash) supplemented with 40mM Imidazole, 1X T300 supplemented with 2M urea and 1M Urea. Samples were analyzed by 10% SDS-PAGE gel and Coomassie stainted. The elutant was FPLC purified and concentrated to 0.1 to 0.15 mM.

A



B

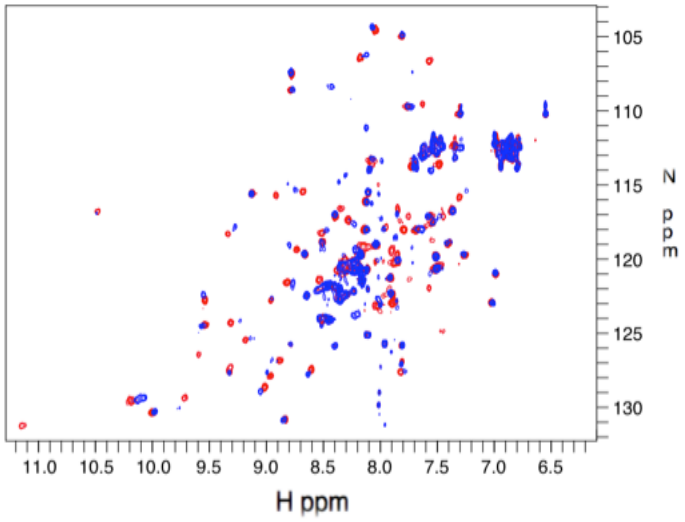


Figure 18. 1H-15N HSQC spectra for hLa 225–332 and 225–408 protein fragments. HLa CTD mutants proteins were generated for 1H-15N heteronuclear single quantum coherence (HSQC) spectroscopy. Bound protein is coloured in blue and unbound protein is coloured in red. A) 0.1mM of 225- 332 construct HSQC B) 0.15mM of 225- 408 construct. Mapping chemical shift perturbations of 225-332. Samples were supplemented with 10% D20 and protein-RNA complex was formed using a 1:1 molar ratio with 12nt Arginine pre-tRNA trailer purchased from Dharmacon. HSQC spectra was acquired on a 700MHz Bruker Ascend 700.

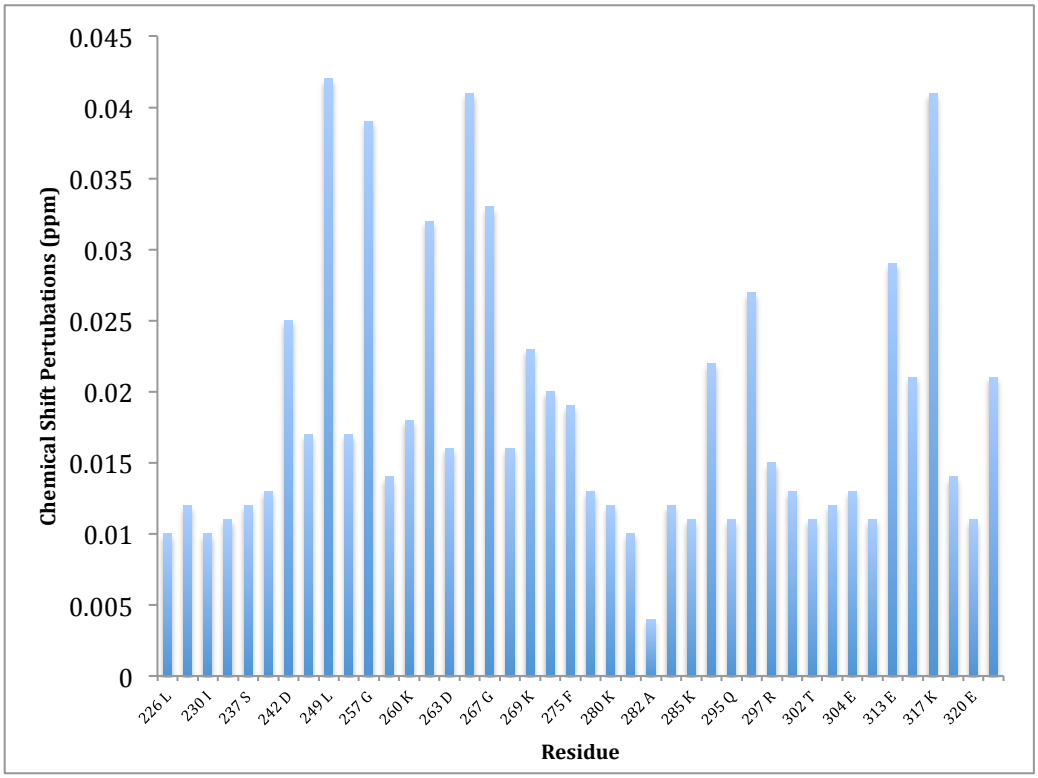
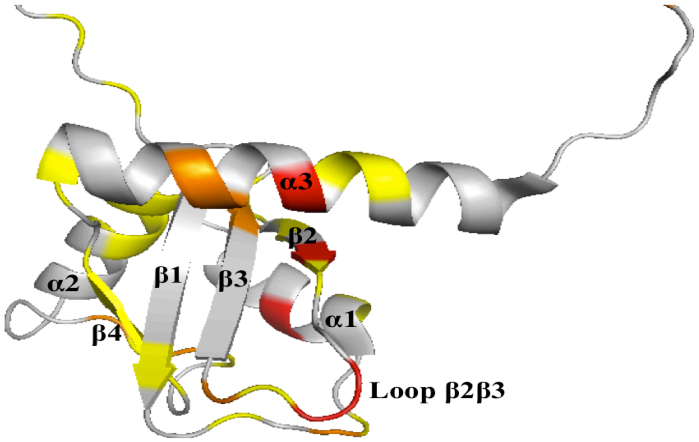


Figure 19. Chemical shift perturbation difference of 1H-15N HSQC spectra for hLa 225–332 between bound and unbound forms. HLa CTD mutants proteins were generated for 1H-15N heteronuclear single quantum coherence (HSQC) spectroscopy. HSQC spectra was acquired on a 700MHz Bruker Ascend 700. Degree of chemical shift perturbation was assigned as follows. Firstly, a weak chemical shift perturbation was assigned between 0.015 – 0.0199ppm. Secondly, a medium strength chemical shift perturbation was assigned between 0.02 – 0.0299ppm. Finally, a strong chemical shift perturbation was assigned as any difference greater than 0.03ppm. All spectra were processed using NMR Pipe and NMR Analysis and compared to a solution structure of the C-terminal RRM of human La (La225-334; PDB 1OWX).

A



B

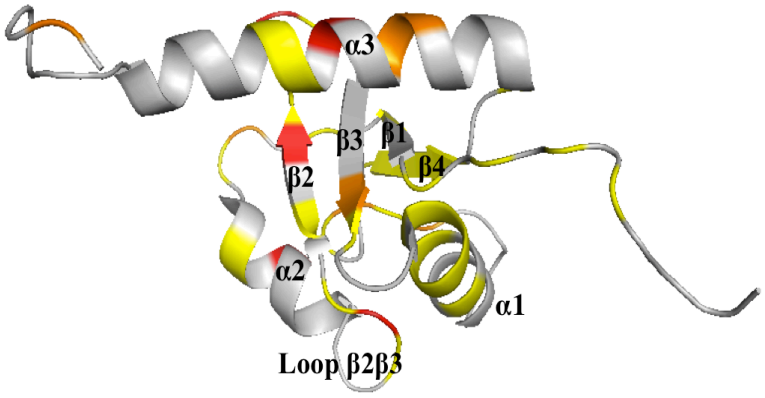


Figure 20. 3D Stereo view of hLa 225-332 protein structure depicting chemical shift perturbations. HLa CTD mutants proteins were generated for 1H-15N heteronuclear single quantum coherence (HSQC) spectroscopy and HSQC spectra was acquired on a 700MHz Bruker Ascend 700. Spectra were processed using NMR Pipe and NMR Analysis and compared to a solution structure of the C-terminal RRM of human La (La225-334; PDB 1OWX). A) Front view of the solvent exposed surface B) top view exposing inner core. Degree of chemical shift perturbation was assigned as follows. A heat map of the NMR spectra was generated to visualize the degree of chemical shift perturbation. Firstly, a weak chemical shift perturbation was assigned between 0.015 – 0.0199ppm (yellow). Secondly, a medium strength chemical shift perturbation was assigned between 0.02 – 0.0299ppm (orange). Finally, a strong chemical shift perturbation was assigned as any difference greater than 0.03ppm (red).

4. Discussion

4.1 A role for the CTD of hLa in RNA affinity

Genuine La proteins are conserved RNA chaperones that primarily target polymerase III transcripts that end in UUU-3'OH (Wolin & Cedervall, 2002; Hussain et al., 2013). All RNAs bound by La make contacts with the conserved La module in the N-terminal domain (NTD), which has been shown to harbour RNA chaperone activity in human La (hLa) through the RRM1 (RNA recognition motif 1) (Naeeni et al., 2012). Moreover, the C-terminal domain (CTD) of hLa harbours both stand displacement and annealing activities (Naeeni et al., 2012). Recent data suggests a role for the non-conserved CTD RRM2 in RNA affinity that is independent of the NTDs chaperone activity (Martino et al., 2012; Kuehnert et al., 2015). Nevertheless, very little is known about the RRM2 dependant chaperone activity of hLa. The objective of this study was to investigate the non-canonical extremities of RRMs in respect to their function in assisting with binding. Additionally, we employed structural analysis during Ribonucleoprotein (RNP) complex formation to elucidate hLa's RRM2 dependant binding. Finally, we investigated structural changes in chaperones that facilitate RNA remodelling.

4.2 The RRM2 alone is not sufficient for binding RNA

The CTD of hLa is comprised of a putative RRM2, and an intrinsically disordered short basic motif (SBM), as well as trafficking elements (Jacks et al., 2003). The full length CTD of hLa (225-408) was previously shown to harbour RNA chaperone activity through FRET assays, albeit with reduced affinity compared to wild type hLa (Naeeni et al., 2012). Interestingly, the RRM1 and RRM2 of hLa both contain an intrinsically disordered C terminus, therefore we and hypothesize that the RNA chaperone activity in the CTD is centered around the RRM2.

To evaluate the affinity of hLa CTD fragments we conducted electrophoretic mobility shift assay (EMSA) on a radiolabelled 12nt long pre-tRNA that ends with UUU-3'OH. The RRM2s C terminal region was shown to be important as binding was abolished with constructs that lack this region (225-332) (Fig.10). This is consistent with the idea that SBM is implicated in RNA binding, indicating that residues after 332 are critical for binding in the context of the C-terminus. Longer CTD constructs (225-339, 225-344, 225-375) can engage RNA targets, albeit with reduced affinity compared to the full-length CTD (Fig. 11 - Fig.14). As such, we can see that between 332 and 408 should be a domain that participates in the RNA binding and residues C terminal to 332 incrementally increase affinity. The loss of SBM abolishes all hLa CTD affinity for 12nt trailer, therefore the RRM2 is not sufficient for binding RNA alone and the minimum length for binding is 225-339.

Binding of the SBM was shown *in vitro* as hLa mutants lacking the SBM, demonstrated a lower affinity for RNAs compared to full-length hLa (1-408) (Fan et

al., 1998). Additionally, The binding capacity of hLa's SBM was assessed *in vivo* through a tRNA mediated suppression assay. Briefly, Intine and Maraia, (2000) showed that the SBM binds the 5'leader of RNA and restricts access to RNase P impeding processing of pre-tRNAs. Ablating the SBM was necessary in order to relive mutated tRNA of obstruction and allowing cells to grow. Moreover, biochemical assays have implicated the C terminal disordered regions of hLa in RNA chaperone activity (Kuehnert et al., 2015). Interestingly, this site is regulated by a novel threonine 389 (T389) phosphorylation site that is controlled by AKT (Kuehnert et al., 2015).

We also investigated hLa's electrostatic sensitivity in respect to salt (KCl) concentration by performing EMSAs at low and high KCl concentrations (50mM and 300mM). We predict that hLa CTD fragments interact electrostatically with the RNA phosphate backbone by virtue of its relatively high number of basic amino acids. All constructs, except for the full length CTD have salt dependent binding as low to high salt we see a drop in affinity. This indicates salt bridges are mediating binding in truncated CTD mutants (Davidovich et al., 2013). Similarly, reports implicate that La's RRM1 affinity for pre-tRNA is salt sensitive as affinity significantly with increased salt. In contrast, regions between 375-408 of hLa have been proposed to have a high conservation of hydrophobic (salt insensitive) contacts, which is in agreement with our results (Bayfield et al., 2009).

In our results a fully bound RNP complex does not form in the context of the isolated CTD. There is also evident smearing at higher protein concentrations, which

may be indicative of weak transient binding. During RNP formation there is steady state equilibrium of bound and unbound complex. In contrast, when the complex is electrophoresed the equilibrium state is lost as the free RNA is migrates faster through the gel. In enzyme kinetics, the equilibrium binding constant ($K_d = k_{off}/k_{on}$) is defined as the dissociation rate constant (k_{off}), divided by the association rate constant (k_{on}). Interactions with a high k_{off} can be disrupted in EMSAs, which underlines a major drawback in EMSA assays (Laniel et al., 2001). Transient interactions are common amongst RNA chaperones (Tompa & Csermely, 2004). With this in mind, a complementary technique that accounts for complex formation with a high dissociation constant (K_{off}) would be FRET or Fluorescence Polarization/Anisotropy (FP), which allows for the quantitation of steady state equilibriums. To understand the structural changes in hLa that account for RNP formation we next turned to NMR.

4.3 RRM2 uses a surface other than the canonical one to bind RNA

Structurally, the RRM2 of hLa contains an additional β4 sheet (β4') and α3 helix, followed by a disordered C terminal region (Birney et al., 1993; Jacks et al., 2003). Furthermore, the RRM2 of hLa is atypical in that it contains a RNA binding surface comprised of charged acidic residues. The precedent of atypical binding already had been set as p65's (LARP7) RRM2 is homologous to hLa (Singh et al, 2013). Because of the atypical RRM fold and electrostatic surface potential we hypothesized that hLa uses a surface other than the canonical one and may be dependent on regions outside the RRM2 fold in binding it's cognate RNA.

As discussed previously, RRMs use several strategies to target their cognate RNAs, including that of conserved RNP1 and RNP2 sequences (Maris et al., 2005). Most commonly, aromatic residue of RNP2 and RNP1 stack between RNA bases, while a positively charged residue of RNP1 forms a salt bridge with the phosphate backbone. However because of the plastic nature of RRMs, there are many different combinations (Maris et al., 2005). The RNP1 sequence of hLa spans residues 265-275 (RGAKEGIILFK). At positions where conserved aromatic residues lie in most RRMs, there is a Glycine (G171) and hydrophobic Isoleucine 173 (I172) (Maris et al., 2005). Additionally, there is also presence of Glutamic Acid 170 (E170), which is not conserved in RNA binding surfaces generally (Jacks et al., 2003). Deletion of the RNP1 sequence in the RRM2 of hLa weakly decreased affinity to pre-tRNAs and to a predicted stem-loop structure in Hepatitis B virus (HBV) RNA (Horke et al., 2002; Horke et al., 2004). The RNP2 sequence of hLa spans residues 235 to 242 (KFSGDLDD). The phenylalanine 236 (F236) aligns to where conserved aromatic residues lie in most RRMs (Maris et al., 2005). The RNP2 sequence was shown to be more important than RNP1 as deletion of RNP2 abolished the interaction with pre-tRNA and to HBV RNA in the context of the full-length protein (Horke et al., 2002; Horke, 2004).

To structurally characterize the interaction between RRM2 and 12nt pre-tRNA we generated 225-332 hLa for 1H-15N Heteronuclear Single Quantum Coherence (HSQC) spectra (Fig. 18A). Generally, there is a significant conformational change upon ligand binding, which can be seen by chemical shift

splitting and peak disappearance (Fig. 19). We mapped all the perturbations onto a previously solved solution structure of RRM2 (La225-334; PDB 1OWX; Jacks et al., 2003) and assigned the appropriate threshold strength via a heat map of the NMR spectra (Fig. 20). Notably, the structure of 225-332/pre-tRNA complex also indicates that the RNP2 is involved in protein-RNA interaction. In contrast, RNP1 seems to be more involved as degree of perturbation was stronger in this region, which is contrary to previous findings.

Remarkably, the non-canonical α3 helix that is obscuring the binding surface has high degree of perturbation. The α3 helix is stabilized by hydrophobic interactions with the β sheets on the RNA binding surface and its role may be to screen the hydrophobic core in its free form (Avis et al., 1996; Jacks et al., 2003). Conceivably, these interactions are displaced in the presence of RNA, which would allow the RNA binding surface to be free from obstruction and engage RNA via the RNP1 and RNP2 sequences analogous to U1A RNA Spliceosomal Protein (U1A) (Oubridge et al., 1994). Furthermore, the high threshold strength of basic residues on the β2 sheet and loop region between β2 and β3 sheets (β2β3 loop3), may be responsible for hydrogen bonding analogous to loop3 in the RRM1 (Naeeni et al., 2012). Unexpectedly, movement also occurs around the β4 sheet, which is away from the RNA binding surface. We predict that this is attributed to global conformational change.

In brief, our results indicate that global conformational changes as well as RNP1 and RNP2 engagement are typical characteristics of RRM binding. In contrast,

we see evidence of the β2 sheet involvement and α3 helix displacement, which is in agreement with the homologous but also atypical p65 RRM2 (Singh et al., 2012; Singh et al., 2013). Binding of p65 to cognate RNA p65 is centered around the β2-sheet and β3-sheet, as RNP1 and RNP2 sequences are absent. Additionally, three aromatic residues on the α3 helix are inserted between RNA base pairs, which form an extension of the α3 helix (α3x) from a C-terminal intrinsically disordered region. As a result, the transitions between ordered and unordered states stabilize the p65/RNA complex. Therefore, these data support our hypothesis of non-canonical RRM2 engagement, however it only gives us preliminary insight about RNA binding. To confirm results we plan to label constructs with $^{13}C^{15}N$ as well as investigate the role of the SBM.

4.4 Increased α-helical content may mediate hLa's chaperone activity

To date, a unified binding domain between RNA chaperones has not been defined and there mechanism of promoting strand annealing and strand displacement activities remains uncertain (Rajkowitsch et al., 2005). While intrinsic disorder is hypothesized to provide RNA destabilization activity through entropy transfer, the mechanism is still poorly understood (Tompa & Csermely, 2004; Rajkowitsch & Schroeder, 2007). The presence of basic residues in intrinsically disordered coils is thermodynamically favourable to form α-helices upon binding (Lomize & Mosberg 1997). Consequently, we hypothesize the disordered SBMs basic residues is the basis for La's chaperone activity.

There is overwhelming evidence of RRM containing proteins using extremities outside the canonical RRM fold to bind RNA (Afroz et al., 2015). The best example that is hypothesized to be homologous to the RRM2 of hLa is the RRM2 of LARP7 family member p65 (Singh et al., 2012). More specifically, deletion of the unstructured region C terminal to xRRM2 of p65 abrogated binding to cognate RNA in the truncated protein, and diminished RNA affinity in the full-length protein (Singh et al., 2013). Likewise, loss of the α3-helix subsequent disordered region in the RRM1 region of hLa decreased strand-annealing activity in FRET assays (Naeeni et al., 2012).

In order to determine structural changes in the C terminal disordered SBM upon RNA binding we performed far UV Circular Dicroishm (CD) Spectroscopy on hLa constructs lacking this region. We monitored negative ellipticities at 208 nm and 222 nm to determine if increased α-helical content in hLa's unstructured region. We predict that increased α-helical content facilitates entropy transfer, which is proposed in our model of hLa's CTD chaperone activity (Fig.21). Qualitatively, we see an interaction in all constructs except for 225-332 as there is no spectral overlap between bound and unbound forms (Fig. 16). Secondary structure predictions indicate that the RRM2 alone (225 -332) becomes slightly more α helical upon RNA binding (Tbl.3). However, constructs encompassing the SBM (225-339, 225-344, and 225-408) have significantly more α-helical content upon complex formation. Although, these values support our predictions they are not consistent with Jacks et al (2003) reported structures of hLa 225–334 and 225–408. Therefore, we could not conclude if proteins

become more α helical to mediate entropy transfer. These results give us some indication of our model, however CD is a generally qualitative study with inherent limitations and a more robust way to confirm our model is NMR.

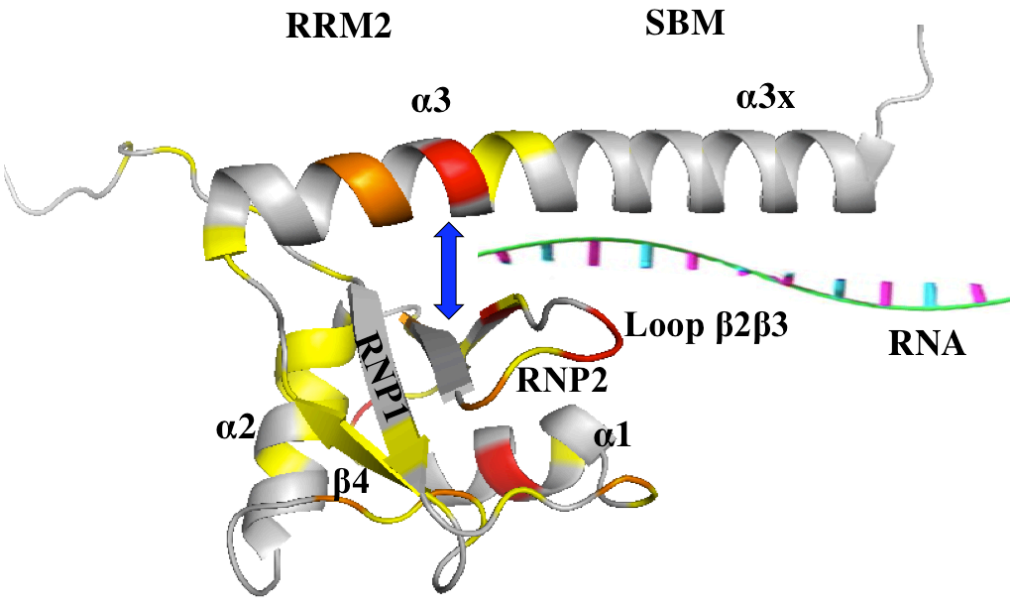


Figure 21. Model of RNA binding by the RRM2 and SBM of hLa. Our results indicate that there is a significant conformational change upon ligand binding. RNP1 and RNP2 engagement are typical characteristics of RRM binding. In contrast, we see evidence of the β2 sheet involvement and α3 helix displacement, which is in agreement with the homologous but also atypical p65 RRM2. Unexpectedly, movement also occurs around the β4 sheet, which is away from the RNA binding surface. In order to accommodate RNA the α3 helix is displaced, which allows for the RNA binding surface to contact RNA (indicated by a blue arrow). The presence of basic residues in intrinsically disordered coils is thermodynamically favourable to form α-helices upon binding, which is hypothesized to provide RNA destabilization activity through entropy transfer. We hypothesize the disordered SBMs basic residues is the basis for hLa's chaperone activity. Hydrogen bonding and electrostatic interactions between the SBM and RNA increases enthalpy. This influences conformational entropy change and formation of an extended α3 helix (α3x) to facilitate RNA remodelling.

4.5 Conclusion

The objective of this study was to employ structural analysis to understand hLa's uncharacterized RRM2 dependant binding. Our work showed that the RRM2 relies on a non-canonical RNA binding surface and displacement of the α3 helix in order to facilitate binding. Overall, constructs containing the disordered SBM exhibit higher affinity for pre-tRNA. Additionally, the RRM2 is salt sensitive and dependent on the SBM, which confirms that RRM containing proteins use extremities outside the canonical RRM fold. Taken together, these data further support the RNA chaperone function of La.

4.6 Future directions

Although the RRM domain exists in all kingdoms of life only a small number have been structurally characterized to date. Therefore it is difficult to predict conserved features of RRMs. Our data gives us preliminary insight about the RRM2 in the CTD of hLa's role in RNA binding, however we must confirm results by labelling longer constructs with $^{13}C^{15}N$ to determine what structural changes occur in the SBM upon RNA binding. Moreover, in order to determine critical residues in RNA affinity we must perform a site directed mutagenesis on hLa's CTD. Additionally, future work will focus on quantitating RNA remodelling kinetics through an RNA chaperone assay, which relies on fluorescence resonance energy transfer (FRET) to determine kinetic parameters and quantitation of steady state

equilibriums. This assay will determine the strand annealing and strand displacement activities for CTD fragments and as a result we can reliably tell if the C-terminal RRM2 has chaperone activity and where this activity maps. Furthermore, we could not determine if increased α-helical content in regions beyond the RRM2 facilitate RNA remodeling through CD. Therefore, we plan on optimizing our CD for to verify our hypothesis. Finally, we plan to probe the function of phosphorylation at Serine 366 (S366) to understand how this affects chaperone activity. Ultimately, we plan on conducting X-ray crystallography on a La optimized construct in order to verify our *in vitro* results. In summary, we aim characterize a conserved role for disordered regions in RNA chaperone activity more generally and be able to extend those findings to other members of the La superfamily as well as other RRM containing proteins.

Supplementary Figures

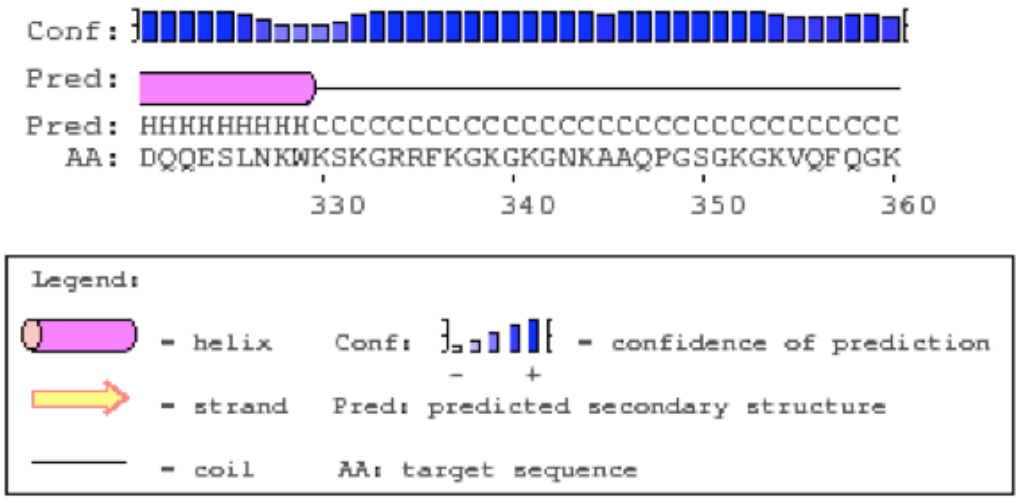


Figure 22. Secondary structure prediction of the short basic motif (SBM) of hLa. Regions C terminal to the RRM2 (330-360) are predicted to form unstructured random coils. Analysis was done using PSIPRED Protein Sequence Analysis Workbench http://bioinf.cs.ucl.ac.uk/psipred/.

References

Afroz, T. Cienikova, Z. Cléry, A. Allain, F. (2015). One, Two, Three, Four! How Multiple RRMs Read the Genome Sequence. Methods in Enzymology. 558: 235–278.

Alfano, C. Sanfelice, D., Babon, J., Kelly, G., Jacks, A., Curry, S., Conte, M.R. (2004). Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein. Nat Struct Mol Biol 11, 323–329.

Allain, F.H., Howe, P.W., Neuhaus, D., Varani, G. (1997). Structural basis of the RNA- binding specificity of human U1A protein. The EMBO Journal, 16(18), 5764–5772.

Ali, N., Pruijn, G.J., Kenan, D.J., Keene, J.D. and Siddiqui, A. (2000) Human La antigen is required for the hepatitis C virus internal ribosome entry site (IRES)-mediated translation. J Biol Chem, 275, 27531-27540.

Bayfield, M.A., Yang, R. and Maraia, R.J. (2010) Conserved and divergent features of the structure and function of La and La-related proteins (LARPs). Biochim Biophys Acta, 1799, 365-378.

Bayfield, M.A. and Maraia, R.J. (2009) Precursor-product discrimination by La protein during tRNA metabolism. Nat Struct Mol Biol, 16, 430-437.

Bayfield, M. A., Kaiser, T. E., Intine, R. V. & Maraia, R. J. (2007). Conservation of a masked nuclear export activity of La proteins and its effects on tRNA maturation. Mol. Cell. Biol. 27, 3303-3312.

Bardwell, J.C.A., Jakob, U. (2012). Conditional disorder in chaperone action. Trends in Biochemical Sciences, 37(12), 517–525.

Bayer, T.S., Booth, L.N., Knudsen, S.M., Ellington, A.D. (2005). Arginine-rich motifs present multiple interfaces for specific binding by RNA. RNA. 11(12):1848-57.

Belisova, A., Semrad, K., Mayer, O., Kocian, G., Waigmann, E., Schroeder, R., Steiner, G. (2005). RNA chaperone activity of protein components of human Ro RNPs. RNA, 11(7), 1084–1094.

Bertrand, E., Houser-Scott, F., Kendall, A., Singer, R. H., & Engelke, D. R. (1998). Nucleolar localization of early tRNA processing. Genes & Development. 12(16), 2463–2468.

Bousquet-Antonelli, C. and Deragon, J. (2009) A comprehensive analysis of the La-motif protein superfamily. RNA.

Biswas, S., Guharoy, M., & Chakrabarti, P. (2008). Structural segments and residue propensities in protein-RNA interfaces: Comparison with protein-protein and protein-DNA complexes. Bioinformation, 2(10), 422–427.

Birney, E., Kumar, S., and Krainer, A.R. (1993). Analysis of the RNA- recognition motif and RS and RGG domains—conservation in meta- zoan pre-messenger-RNA splicing factors. Nucleic Acids Res. 21, 5803–5816.

Brenet, F., Socci, N. D., Sonenberg, N. & Holland, E. C. (2009). Akt phosphorylation of La regulates specific mRNA translation in glial progenitors. Oncogene 28, 128–139.

Broekhuis, C.H., Neubauer, G., van der Heijden, A., Mann, M., Proud, C.G., van Venrooij,W.J., Pruijn, G.J. (2000). Detailed analysis of the phosphorylation of the human La (SS-B) autoantigen. (De)phosphorylation does not affect its subcellular distribution. Biochemistry. 21;39(11):3023-33.

Cai, L., Fritz, D., Stefanovic, L. And Stefanovic, B. (2010). Binding of LARP6 to the conserved 5' stem-loop regulates translation of mRNA encoding type I collagen. Journal of Molecular Biology 395: 309-326.

Callahan, K. P., & Butler, J. S. (2010). TRAMP Complex Enhances RNA Degradation by the Nuclear Exosome Component Rrp6. The Journal of Biological Chemistry, 285(6), 3540–3547.

Chakshusmathi, G., Kim, S. D., Rubinson, D. A., & Wolin, S. L. (2003). A La protein requirement for efficient pre-tRNA folding. The EMBO Journal, 22(24), 6562–6572.

Chakrabartty, A., Kortemme, T., & Baldwin, R. L. (1994). Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. Protein Science : A Publication of the Protein Society, 3(5), 843–852.

Cléry, A., Blatter, M., Allain, F.H. (2008). RNA recognition motifs: boring? Not quite. Curr Opin Struct Biol. 18(3):290-8.

Costa-Mattioli, M., Svitkin, Y., & Sonenberg, N. (2004). La Autoantigen Is Necessary for Optimal Function of the Poliovirus and Hepatitis C Virus Internal Ribosome Entry Site In Vivo and In Vitro . Molecular and Cellular Biology, 24(15), 6861–6870.

Cristofari, G., Darlix, J.L. (2002). The ubiquitous nature of RNA chaperone proteins. Prog Nucleic Acid Res Mol Biol. 72:223-68.

Davidovich, C., Zheng, L., Goodrich, K. J., & Cech, T. R. (2013). Promiscuous RNA binding by Polycomb Repressive Complex 2. Nature Structural & Molecular Biology, 20(11).

Doetsch, M., Schroeder, R., & Fürtig, B. (2011). Transient RNA–protein interactions in RNA folding. The Febs Journal, 278(10), 1634–1642.

Dong, G., Chakshusmathi, G., Wolin, S. L., & Reinisch, K. M. (2004). Structure of the La motif: a winged helix domain mediates RNA binding via a conserved aromatic patch. The EMBO Journal, 23(5), 1000–1007.

Fan, H., Goodier, J. L., Chamberlain, J. R., Engelke, D. R., & Maraia, R. J. (1998). 5′ Processing of tRNA Precursors Can Be Modulated by the Human La Antigen Phosphoprotein. Molecular and Cellular Biology, 18(6), 3201–3211.

Fourmann, J.B., Tillault, A.-S., Blaud, M., Leclerc, F., Branlant, C., & Charpentier, B. (2013). Comparative Study of Two Box H/ACA Ribonucleoprotein Pseudouridine-Synthases: Relation between Conformational Dynamics of the Guide RNA, Enzyme Assembly and Activity. PLoS ONE .8(7): e70313.
Gajiwala, K. S., Burley, S.K. (2000). Winged helix proteins. Curr Opin Struct Biol 10: 110–116.

Gottlieb, E., Steitz, J. A. (1989). Function of the mammalian La protein: evidence for its action in transcription termination by RNA polymerase III. The EMBO Journal, 8(3), 851–861.

Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. Nature Protocols, 1(6): 2876–2890.

Hashimoto C, Steitz JA. (1983). Sequential association of nucleolar 7-2 RNA with two different autoantigens. J Biol Chem. 258:1379–82.

Hendrick, J.P., Wolin, S.L., Rinke, J., Lerner, M.R., Steitz, J.A. (1981). Ro small cytoplasmic ribonucleoproteins are a subclass of La ribonucleoproteins: further characterization of the Ro and La small ribonucleoproteins from uninfected mammalian cells. Mol Cell Biol. 1(12):1138-49.

Horke, S., Reumann, K., Rang, A., Heise, T. (2002). Molecular characterization of the human La protein hepatitis B virus RNA B interaction in vitro. J Biol Chem. 277(38):34949-58.

Horke, S., Reumann, K., Schulze, C., Grosse, F., Heise, T. (2004). The La motif and the RNA recognition motifs of human La autoantigen contribute individually to RNA recognition and subcellular localization. J Biol Chem. 279(48):50302-9.

Hussain, R.H., Zawawi, M., Bayfield, M.A. (2013). Conservation of RNA chaperone activity of the human La-related proteins 4, 6 and 7. Nucleic Acids Res. (18): 8715-25.

Huang, Y., Bayfield, M.A., Intine, R.V. and Maraia, R.J. (2006) Separate RNA-binding surfaces on the multifunctional La protein mediate distinguishable activities in tRNA maturation. Nat Struct Mol Biol, 13, 611-618.

Intine, R.V., Sakulich, A.L., Koduru, S.B., Huang, Y., Pierstorff, E., Goodier, J.L, Phan, L., Maraia, R.J. (2000) Control of transfer RNA maturation by phosphorylation of the human La antigen on serine 366. Mol Cell. 6(2):339-48.

Intine, R.V., Dundr, M., Misteli, T., Maraia, R.J. (2002). Aberrant nuclear trafficking of La protein leads to disordered processing of associated precursor tRNAs. Mol Cell. 9(5):1113-23.

Intine, R.V., Tenenbaum, S.A., Sakulich, A.S., Keene, J.D. and Maraia, R.J. (2003) Differential phosphorylation and subcellular localization of La RNPs associated with precursor tRNAs and translation-related mRNAs. Molecular Cell, 12, 1301-1307.

Jacks, A., Babon, J., Kelly, G., Manolaridis, I., Cary, P.D., Curry, S., Conte, M.R. (2003) Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element. Structure 7:833–843.

Kucera, N. J., Hodsdon, M. E., & Wolin, S. L. (2011). An intrinsically disordered C terminus allows the La protein to assist the biogenesis of diverse noncoding RNA precursors. PNAS. 108(4), 1308–1313.

Kuehnert, J., Sommer, G., Zierk, A.W., Fedarovich, A., Brock, A., Fedarovich, D., Heise, T. (2015) Novel RNA chaperone domain of RNA-binding protein La is regulated by AKT phosphorylation. Nucleic Acids Res. 43(1):581-94.

Kotik-Kogan, O., Valentine, E.R., Sanfelice, D., Conte, M.R. and Curry, S. (2008) Structural analysis reveals conformational plasticity in the recognition of RNA 3' ends by the human La protein. Structure, 16, 852-862.

Krueger, B. J., Jeronimo, C., Roy, B. B., Bouchard, A., Barrandon, C., Byers, S. A., Price, D. H. (2008). LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. Nucleic Acids Research, 36(7), 2219–2229.

Laniel, M.A., Béliveau, A. Guérin, S.L. (2001). Electrophoretic Mobility Shift Assays for the Analysis of DNA-Protein Interactions. Methods Mol. Biol. 148, 13-30.

Long, K. S., Cedervall, T., Walch-Solimena, C., Noe, D. A., Huddleston, M. J., Annan, R. S., & Wolin, S. L. (2001). Phosphorylation of the Saccharomyces cerevisiae La protein does not appear to be required for its functions in tRNA maturation and nascent RNA stabilization. RNA, 7(11), 1589–1602.

Lomize AL, Mosberg HI. (1997). Thermodynamic model of secondary structure for alpha-helical peptides and proteins. Biopolymers. 42(2): 239-69.
Loerch, S., Kielkopf, C.L. (2015). Dividing and Conquering the Family of RNA Recognition Motifs: A Representative Case Based on hnRNP L. J Mol Biol. S0022-2836(15)00349-6.

Maraia, R.J. Lamichhane, T.N. (2011) 3′ processing of eukaryotic precursor tRNAs. RNA 2: 362-375.

Maraia, R.J., & Lamichhane, T.N. (2011). 3' processing of eukaryotic precursor tRNAs. Wiley Interdisciplinary Reviews. RNA, 2(3), 362–375.

Maraia, R.J., Intine, R.V. (2001) Recognition of nascent RNA by the human La antigen: conserved and divergent features of structure and function. Mol Cell Biol 21:367–379.

Maris, C., Dominguez, C., Allain, F.H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J. 272(9): 2118-31.

Martino, L., Pennell, S., Kelly, G., Bui, T.T., Kotik-Kogan, O., Smerdon, S.J., Drake, A.F., Curry, S., Conte, M.R. (2011). 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. 40(3):1381-1394.

Mayer, O., Rajkowitsch, L., Lorenz, C., Konrat, R., & Schroeder, R. (2007). RNA chaperone activity and RNA-binding properties of the E. coli protein StpA. Nucleic Acids Research, 35(4), 1257–1269.

Meerovitch, K., Svitkin, Y. V., Lee, H. S., Lejbkowicz, F., Kenan, D. J., Chan, E. K., Sonenberg, N. (1993). La autoantigen enhances and corrects aberrant translation of poliovirus RNA in reticulocyte lysate. Journal of Virology, 67(7), 3798–3807.

Naeeni, A.R., Conte, M.R. and Bayfield, M.A. (2012) RNA chaperone activity of the human La protein is mediated by a variant RNA recognition motif. J Biol Chem, 287, 5472-5482.

Nykamp, K., Lee, M.H. and Kimble, J. (2008). C. elegans La-related protein, LARP-1, localizes to germline P bodies and attenuates Ras-MAPK signalling during oogenesis. RNA 14:1378-1389.

Ohndorf, U. M., Steegborn, C., Knijff, R. & Sondermann, P. (2001). Contributions of the individual domains in human La protein to its RNA 3'-end binding activity. J. Biol. Chem 276, 27188–27196

Oubridge, C., Ito, N., Evans, P.R., Teo, C.H., Nagai, K. (1994). Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature. 372 (6505): 432-8.

Phizicky, E. M. (2005). Have tRNA, will travel. Proceedings of the National Academy of Sciences of the United States of America, 102(32):11127–11128.

Pudi, R., Srinivasan, P. and Das, S. (2004) La protein binding at the GCAC site near the initiator AUG facilitates the ribosomal assembly on the hepatitis C virus RNA to influence internal ribosome entry site-mediated translation. J Biol Chem, 279, 29879-29888.

Rajkowitsch, L., Schroeder, R. (2007). Dissecting RNA chaperone activity. RNA. 13(12), 2053–2060.

Rajkowitsch, L., Semrad, K., Mayer, O. Schroeder, R. (2005). Assays for the RNA chaperone activity of proteins. Biochem. Soc. Trans. 33:450–455.

Schmidt, K., Xu, Z., Mathews, D. H., & Butler, J. S. (2012). Air proteins control differential TRAMP substrate specificity for nuclear RNA surveillance. RNA, 18(10), 1934–1945.

Schwartz, E., Intine, R.V. and Maraia, R.J. (2004) CK2 is responsible for phosphorylation of human La protein serine-366 and can modulate 5'TOP mRNA metabolism. Mol Cell Biol, 24, 9580-9591.

Simons, F.H., Broers, F.J., Van Venrooij, W.J., Pruijn, G.J. (1996). Characterization of cis-acting signals for nuclear import and retention of the La (SS-B) autoantigen. Exp Cell Res. 224(2):224-36.

Singh, M., Choi, C.P, Feigon, J. (2013). xRRM: a new class of RRM found in the telomerase La family protein p65. RNA Biol. 10(3):353-9

Singh, M., Wang, Z., Koo, B.K., Patel, A., Cascio, D., Collins, K., Feigon, J. (2012). Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. Mol Cell. 47(1):16-26.

Sobel, S. G. & Wolin, S. L. (1999). Two Yeast La Motif-containing Proteins Are RNA-binding Proteins that Associate with Polyribosomes. Mol. Biol. Cell 10, 3849–3862.

Sommer, G., Dittmann, J., Kuehnert, J., Reumann, K., Schwartz, P.E., Will, H., Coulter, B.L., Smith, M.T. and Heise, T. (2011) The RNA-binding protein La contributes to cell proliferation and CCND1 expression. Oncogene, 30, 434-444.

Stefano, J.E. (1984). Purified lupus antigen La recognizes an oligouridylate stretch common to the 3' termini of RNA polymerase III transcripts. Cell. 36(1):145-54.

Tan, R., Chen, L., Buettner, J.A., Hudson, D., Frankel, A.D. (1993). RNA recognition by an isolated alpha helix. Cell. 73(5): 1031-40.

Teplova, M., Yuan, Y.R., Phan, A.T., Malinina, L., Ilin, S., Teplov, A. and Patel, D.J. (2006) Structural basis for recognition and sequestration of UUU(OH) 3' temini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. Mol Cell, 21, 75-85.

Tompa, P., Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. FASEB J. 18(11):1169-75.

Van Horn, D.J., Yoo, C.J., Xue, D., Shi, H. and Wolin, S.L. (1997) The La protein in Schizosaccharomyces pombe: a conserved yet dispensable phosphoprotein that functions in tRNA maturation. RNA, 3, 1434-1443.

Weeks, K,.M. (1997). Protein-facilitated RNA folding. Curr. Opin. Struct. Biol. 7:336–342

Whitmore, L. and Wallace, B.A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. NAR. 32: W668-673.

Whitmore, L. and Wallace, B.A. (2008) Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and Reference Databases. Biopolymers 89: 392-400.

Woodson, S.A. (2010). Taming free energy landscapes with RNA chaperones. RNA Biology, 7(6), 677–686.

Wolin, S.L. and Cedervall, T. (2002) The La protein. Annu Rev Biochem, 71, 375-403.

Xue, D., Rubinson, D.A., Pannone, B.K., Yoo, C.J., Wolin, S.L.(2000). U snRNP assembly in yeast involves the La protein. EMBO J. 19(11):2763.

Yang, R., Gaidamakov, S.A., Xie, J., Lee, J., Martino, L., Kozlov, G., Crawford, A.K., Russo, A.N., Conte, M.R., Gehring, K. and Maraia, R.J. (2011). La-related protein 4 binds poly(A), interacts with the poly(A)-binding protein MILLE domain via a variant PAM2w motif, and can promote mRNA stability. Molecular and Cellular Biology 31: 542-556.

Yoo, C.J., Wolin, S. L. (1994). La Proteins from Drosophila-Melanogaster and Saccharomyces-Cerevisiae - a Yeast Homolog of the La Autoantigen is Dispensable for Growth. Mol. Cell. Biol. 14, 5412-5424.

Zuker, M. (1989). Computer prediction of RNA structure. Methods Enzymol. 180:262–288.