UNIVERSITAT POLITÈCNICA DE VALÈNCIA

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

The RNA binding protein Mip6, a novel cellular partner of Mex67 export factor with implications in mRNA export

Nada Mohamad

Director: Dr. Jerónimo Bravo Sicilia
Tutor: Dr. María Adelaida García Gimeno
Doctoral thesis, September 2017

INSTITUTO DE BIOMEDICINA DE VALENCIA CSIC

Dr. Jerónimo Bravo Sicilia certifies that the doctoral thesis under the title "**The RNA binding protein Mip6, a novel cellular partner of Mex67 export factor with implications in mRNA export**" done by **Ms. Nada Mohamad** under his supervision in the **Institute of Biomedicine of Valencia (IBV- CSIC)**, fulfils the requirements to obtain the PhD in Biotechnology and authorizes the request for the deposition of the final version of thesis.

Valencia, July 2017
Signature:

Acknowledgements

This journey was not without challenges, I'm thankful for every single person on the way that helped me go through it every day.

To Dr. Jerónimo Bravo Sicilia, the director of my thesis, I would like to express my sincere gratitude for accepting me in your lab in the first place, for the continuous support during my PhD, for the trust, the scientific guidance and the knowledge you provided me. Thank you.

To Dr. Maria Adelaida Garcia Gimeno, tutor of my thesis and one of the nicest people I met, thank you for the support, care, help and advice especially toward the end of my thesis.

To Dr. Jose Manuel Pérez Cañadillas, and Dr. Susana Rodriguez Navarro, our collaborators, from whom I learned, for supporting this work and adding to it. Thank you.

To Susana Masiá Adalid, Lidia Orea, Irene Martín, Laura, Adrian (also known as Alejandro), and Noelia. A special thanks to the members of Alberto Marina's lab, Vicente Rubio's lab, to Amparo Almero, Carmen Muñoz Ballester, fellow friends, staff, and colleagues in the institute, former and current, forgive me if I couldn't mention all, it would need many pages to fill all your names, so to all of you I am grateful for the family you've been to me for all these years, in challenging times, long work days, the days where we felt sometimes frustrated, but also for the good times. It was nice working with you all. I love you all.

A very special thanks to my dearest Marcin Węgrecki, Sara Zamora, and Leticia Dominguez for their friendship, all the support, company, advice, and good times, and of course the daily breakfast Marcin! I couldn't have done it without you, not only in thesis terms

but also for opening my mind and life to new perspectives. I owe you love and gratitude. I hope our friendship would last long.

To my family: my parents as well as my sisters Manal, Fatima, Nour and brothers Mohamad, and Ahmad who believed in me all the way through. I love you.

To my companion and best friend Mohamad Abbas, I'm forever grateful for you being there for me in the good but mostly in bad times. Thank you for being the fantastic person you are. This thesis would not have been accomplished and I could have not made it this far without you by my side. Thank you for putting up with me, I am truly thankful for having you in my life.

This was not an easy ride, but certainly was a life changing one. I met many new people, learned many new things. I gained many friendships, knowledge, and experience. I am grateful for it all.

Index

Abstract..........1
Resumen..........5
Resum..........9
1. Introduction..........13
1.1.messenger RNA (mRNA) biogenesis: Interconnecting processes...17
1.1.1.Transcription initiation..........17
1.2. Elongation..........19
1.2.1. Co-transcriptional 5´Capping..........20
1.2.2. Coupling to splicing..........20
1.2.3. 3´ Polyadenylation and transcription termination..........21
1.3. Nucleo-cytoplasmic RNA exports..........22
1.3.1 RNA export factors..........24
Exportins..........24
Crm1/ Xpo1..........25
1.4. mRNA export: From transcription site to the nuclear pore..........26
1.4.1. mRNA export adaptors..........28
1.4.1.1 THO/TREX complex..........29
1.4.1.2 TREX2 complex..........33
1.5. rRNA export..........35
1.6. Nuclear Pore Complex (NPC)..........37
1.6.1. Role in mRNA quality control..........39
1.6.2. mRNA export termination and cytoplasmic mRNPs remodelling..........40
MEX67..........42
MIP6..........46
2. Objectives..........51
3. Materials and Methods..........53
3.1. Cloning..........55
Cloning using Restriction enzymes..........57
Ligase Independent Cloning (LIC)..........58
In-fusion Cloning..........58
3.2. Site directed mutagenesis..........60
3.3. Protein over-expression in *E. coli*..........61
3.3.1. Small-scale expression..........62

3.3.2. Large-scale over-expression........63
3.3.3. Protein expression in insect cells-baculovirus system........64
3.4. Protein purifications........66
3.4.1. Small-scale purification........66
3.4.2. Large-scale protein purification........67
3.4.2.1. Cell lysis and protein extraction........67
3.4.2.2. Affinity chromatography of His-tagged proteins........69
3.4.2.3. Affinity chromatography of GST-tagged proteins........69
3.4.2.4. Affinity chromatography of MBP-tagged proteins........70
3.4.2.5. Affinity chromatography using HiTrap Heparin HD column........70
3.4.2.6. Protein tag digestion........71
3.4.3. Size exclusion chromatography (gel filtration)........71
3.5. SDS-PAGE (Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis)........72
3.6. Protein quantification........72
3.7. Native gel electrophoresis........73
3.8. *In vitro* binding experiments........73
3.8.1. Poly (U) agarose beads binding experiment........73
3.8.2. Pull down assay........74
3.8.3. Bio-layer Interferometry........75
3.8.4. Isothermal Titration Calorimetry (ITC)........76
3.8.5. Protein Cross-linking........77
3.9. Protein crystallization........77
3.9.1. Crystals of Mip6 RRM3$_{(313-389)}$........77
3.9.2. Crystals of Mex67$_{(528-599)}$........78
3.9.3. Crystals from Pes4 RRM3/4........79
3.9.3.1. Crystal structure of Pes4 RRM3........79
3.9.3.2. Crystal structure of Pes4 RRM4........79
3.9.3.3. Crystal structure of RNA-free Pes4 RRM3/4........80
3.9.3.4. Crystal structure of Pes4 RRM3/4 with RNA........81
3.10. Structure surface electrostatic potential........81
3.11. NMR spectroscopy titration........81
4. Results........83

4.1. Mip6 protein expression and purification...............................85
4.2. MIP6 is an RNA binding protein that binds RNA with high affinity..86
4.2.1. Expression and purification of Mip6 (111-480).........................87
4.2.2. Mip6$_{(111-480)}$ binds RNA *in vitro*..88
4.3. MIP6 has fourth RNA recognition motif................................89
4.3.1. Expression and purification of Mip6 RRM1/2 and MIP6 RRM3/4...90
4.3.2. Expression and purification of Mip6 RRM3$_{(313-389)}$...................91
4.3.3. Expression and purification of Mip6 RRM4.............................92
4.3.4. Mip6 RNA recognition motifs independently bind poly-uridylic acid *in vitro*..94
4.3.5. Mip6 RRM4 binds RNA with high affinity...............................86
4.4. Crystal structure of Mip6 RRM3$_{(313-389)}$.......................................99
4.4.1. Crystallization..99
4.4.2. Data collection and processing...100
4.4.3. Mip6 RRM3$_{(313-389)}$ structures reveal different loop conformations..101
4.5. Characterizing the interaction between Mex67 and Mip6.104
4.5.1. Expression and purification of Mex67 C-term......................104
4.5.2. Mex67 binds Mip6 *in vitro* with high affinity........................105
4.5.3. Mex67 interacts with Mip6 through its RRM4......................106
4.6. Crystal structure of Mex67 C-term$_{(528-599)}$................................113
4.6.1. Crystallization and data processing..113
4.6.2. Mex67 C-term structure made of alpha helices.....................116
4.6.3. Evolutionary conservation of the Mex67 C-term UBA domain..117
4.6.4. Mex67 C-term$_{(528-599)}$ electrostatic surface is not positively charged...118
4.7. Mex67 UBA domain is the binding site of Mip6....................120
4.7.1. Expression and purification...120
4.7.2. Removing Mex67 UBA domain disrupts the interaction with Mip6...122
4.7.3. Mex67 UBA domain is the site of binding of Mip6................123

4.7.4. Mex67 UBA domain needs to be in a specific conformation to be able to bind Mip6……124
4.8. Chemical cross-linking……127
4.9. Mip6 RRM4 RNA binding interfere with its interaction with Mex67……129
4.10. Pes4, a Mip6 Paralogue, is also an RNA binding protein..133
4.10.1. Expression and purification of Pes4 RRM3/4……134
4.10.2. Pes4 binds RNA *in vitro* with high affinity……135
4.11. Pes4 RRM3/4 does not interact with Mex67 *in vitro*……136
4.12. Crystal structures of Pes4……138
4.12.1. Crystal structure of Pes4 RRM3……138
4.12.1.1. Crystallization and data processing……138
4.12.1.2. Pes4 RRM3 reveals a canonical RRM structure……140
4.12.2. Crystal structure of Pes4 RRM4……141
4.12.2.1. Crystallization and data processing……141
4.12.2.2. Pes4 RRM4 shows a difference! ……142
4.12.3. Crystal structure of Pes4 RRM3/4……144
4.12.3.1. Crystallization and data processing……144
4.12.3.2. Pes4 RRM3 and RRM4 have no contact between each other……146
4.12.4. Towards the crystal structure of Pes4 RRM3/4 bound to RNA……148
4.13. Mip6 RRM4 model building……151
4.14. NMR titration analysis of Mip6 RRM4 upon complexation with Mex67$_{(528\text{-}599)}$ ……152
4.15. A Mip6 mutant that disrupts the interaction with Mex67……155
4.15.1. Expression and purification of Mip6 W442A mutant……156
4.15.2. Mip6 RRM4 W442A do not bind Mex67$_{(528\text{-}599)}$……157
4.15.3. Mip6 RRM4 W442A mutant disrupts the ability to bind Mex67 but not RNA ……159
5. Discussion……161
5.1. Mip6 and Pes4 protein expression and solubility……163

5.2. Mip6 and Pes4 are RNA binding proteins with multiple RNA recognition motifs..164
5.3. Pes4 RRM3/4 structure and the role of multiple RRMs...............168
5.4. Mip6 RRM4 is the binding surface for Mex67................................171
5.5. Pes4 RRM4 binds RNA but not Mex67..173
5.6. Mip6 RRM4 Mex67 binding affects its ability to bind RNA.......174
5.7. Mex67 UBA domain gets an additional partner............................180
5.8. Mip6 function *in vivo* as a possible novel adaptor protein for Mex67...185
5.9. Mip6 additional role during cellular stress....................................186
5.10. Pes4 evolutionary divergence of function...................................188
5.11. Concluding remarks..189
6. Conclusions..191
7. References...195

Abbreviations

A
Å: Angstroms
ADP: Adenosine diphosphate
APBS: Adaptive Poisson-Boltzmann Solver
ATP: Adenosine triphosphate

B
BME: β-mercapto-ethanol
BSA: Bovine serum albumin

C
CBC: Cap-binding protein complex
CDD: Conserved domain database
Crm1: Chromosome maintenance region 1
CSIYII: Crystal screen I and II
CSP: Chemical shift perturbation
CTD: Carboxy-terminal domain

D
dATP: Deoxyadenosine triphosphate
DEPC: Diethyl pyrocarbonate
DMSO: Dimethyl sulfoxide
DNA: Deoxyribonucleic acid
DTT: Dithiothreitol
dTTP: Deoxythymidine triphosphate

E
***E. coli*:** *Escherichia coli*
EDTA: Ethylenediaminetetraacetic acid
EJC: exon-exon junction complex

F
FF: Fast flow
FG nucleoporins: Phenylalanine-Glycine nucleoporins
FL: Full length
FPLC: Fast protein liquid chromatography
Fwd: Forward
G
GFP: Green fluorescent protein

GST: Glutathione S-transferase
GTase: RNA guanylyltransferase

H
HP: High performance
HSQC: Heteronuclear single quantum coherence

I
InsP6: Inositol hexaisophosphate
IPTG: Isopropyl β-D-1-thiogalactopyranoside
ITC: Isothermal Titration Calorimetry

K
Ka: Association constant
Kb: Kilo base pairs
Kd: Dissociation constant
KDa: Kilo Dalton
KH2PO4: Potassium dihydrogen phosphate

L
LRR: Leucine rich repeat
LB: Lysogeny broth
LIC: Ligase independent cloning

M
MBP: Maltose binding protein
MDa: Mega Dalton
Mex67: mRNA export factor 67
MgSO4: Magnesium sulphate
miRNA: microRNA
Mip6: Mex67 interacting protein 6
MQ water: Milli-Q ‘ultrapure’ water
mRNA: Messenger ribonucleic acid
mRNP: Mature ribonucleoproteins

N
Na2HPO4: Sodium phosphate dibasic
NaCl: Sodium chloride
NaOH: Sodium hydroxide

NCBI: National Center for Biotechnology Information
NES: Nuclear export signal
(NH4)2SO4: Ammonium sulphate
NLS: Nuclear localization signal
NMR: Nuclear Magnetic Resonance
NPC: Nuclear pore complex
NTF2: Nuclear transport factor 2
Nup: Nuclear pore protein

O
OD: Optical density

P
PAP: Poly(A) polymerase
PBS: Phosphate-buffered saline
PCR: Polymerase chain reaction
PDB: Protein data bank
PDB ID: Protein data bank identity document
PEG: Polyethylene glycol
pH: potential of hydrogen
PI: Isoelectric point
Pol I: RNA Polymerase I
Pol II: RNA Polymerase II
Poly A: Poly-adenosine
PP: PreScission protease

R
RBPs: RNA binding proteins
Rev: Reverse
R.M.S.D: root-mean-square deviation of atomic positions
RNA: Ribonucleic acid
RNP: Ribonucleoprotein
Rpm: Revolutions per minute
RRM: RNA recognition motif
rRNA: Ribosomal ribonucleic acid

S
SAP: Shrimp Alkaline Phosphatase

SDS-PAGE: Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis
snRNA: small nuclear RNA
snRNP: small nuclear RNAs and their associated proteins
SR: Serine-Arginine
Sulfo-SMPB: (Sulfosuccinimidyl 4-maleimidophenyl) butyrate
Sxl: Sex letal protein

T
TAFs: TBP-associated factors
TBP: TATA box-binding protein
TE Buffer: Tris-EDTA buffer
TEV : Tobacco etch virus
TLS: Translation libration screwmotion
Tpase: RNA triphosphatase
TREX: TRancription-Export
tRNA: Transfer ribonucleic acid

U
UBA: Ubiquitin associated domain

X
Xpo1: Exportin-1

ΔH: change in Enthalpy
ΔS: change in Entropy

Abstract

Nuclear export of messenger ribonucleic acid (mRNA) is a complex and essential process for a correct gene expression in all eukaryotic cells. The export of mRNA through the nuclear pore complex depends mostly on the crosstalk and coordination of several proteins forming what is known as mRNPs (messenger ribonucleoproteins) that play dynamic, interconnecting roles in the different mRNA biogenesis steps such as pre-mRNA processing, stability, and export.

One key protein in this process is Mex67, conserved from yeast to humans, is the major messenger RNA exporter also involved in ribosomal RNA export. Mex67 interacts with Mtr2 to form an evolutionary conserved heterodimer essential for proper mRNA export and subsequently the survival of the cell. Mex67 have been studied for many years, however due to the complexity and interconnectivity of the different processes in mRNA biogenesis, there is yet to uncover many details on the dynamics of the process and the crosstalk between Mex67 and its many partners.

In this study, using a combination of biochemical, biophysical, and structural analysis, we characterize the interaction between Mex67 and a novel partner protein called Mip6 (Mex67 interacting protein 6). We were able to reconstitute a stable complex *in vitro*, and extensively study the mechanism in which the two proteins interact. We also solved the crystal structure of the C-terminal region of Mex67 that interacts with Mip6 and identified the UBA domain of Mex67, known to bind FG nucleoporins and Hpr1 protein as also the site where Mip6 binds. However, little was known about the structure or function of Mip6 and its paralogue Pes4. Here we proved that Mip6 is an RNA binding protein with four RNA recognition motifs that binds RNA *in vitro* with high affinity. Additionally, its fourth RNA recognition motif was also the site of binding of Mex67. Furthermore, we showed that the Mex67 complex formation with Mip6 RRM4 compromises its ability to bind RNA or vice versa. We also designed a point mutation on Mip6 RRM4 that disrupts its interaction with

Mex67 but not with RNA. Subsequent *in vivo* yeast assays led us to hypothesize a role of Mip6 as an adaptor protein for Mex67 in nuclear export especially upon stress. Additional function of Mip6 was the localization of its bound mRNA to cytoplasmic stress granules in cellular stress conditions.

Moreover, the crystal structures of Mip6 RRM3, Pes4 RRM3, Pes4 RRM4, and Pes4 RRM3/4 were also solved. All RRMs adopted a canonical RRM fold with conserved RNP1 and RNP2 sequences normally involved in RNA binding, except Mip6 RRM3 that was missing the aromatic ring in RNP2. In the structure of RNA-free Pes4 RRM3/4, the tandem RRM domains were connected with a flexible disordered linker and no inter-domain contact between them. Finally, although Pes4 RRM4 was binding RNA *in vitro*, it did not have the ability to interact with Mex67 thus suggesting a separate evolutionary function for Mip6 and Pes4.

Resumen

La exportación nuclear de ácido ribonucleico mensajero (ARNm) es un proceso complejo y esencial para una expresión correcta de los genes en todas las células eucariotas. La exportación de ARNm a través del complejo del poro nuclear depende principalmente de la interacción y coordinación de varias proteínas, que forman lo que se conoce como mRNPs (ribonucleoproteínas mensajeras), que tienen un papel dinámico e interconectado en las diferentes etapas de la biogénesis de ARNm, tales como el procesamiento del pre-ARNm, estabilidad, y exportación.

Una proteína clave en este proceso es Mex67, conservada de levaduras a humanos, que es la principal exportadora de ARN mensajero y también está implicada en la exportación de ARN ribosomal. Mex67 interacciona con Mtr2 para formar un heterodímero conservado evolutivamente esencial para una exportación adecuada de ARNm y la consiguiente supervivencia de la célula. Se ha estudiado Mex67 durante muchos años, sin embargo, debido a la complejidad e interconectividad de los diferentes procesos de biogénesis de ARNm, todavía quedan por descubrir muchos detalles de la dinámica del proceso y las interacciones entre Mex67 y sus muchas proteínas asociadas.

En este estudio, combinando un análisis bioquímico, biofísico y estructural, hemos caracterizado la interacción entre Mex67 y una nueva proteína asociada denominada Mip6 (proteína 6 que interacciona con Mex67). Hemos podido reconstituir un complejo estable *in vitro* y estudiar extensivamente el mecanismo por el cual interaccionan estas dos proteínas. También hemos resuelto la estructura cristalográfica de la región C-terminal de Mex67 que interacciona con Mip6 e identificado el dominio UBA de Mex67, conocido por unirse a nucleoporinas FG y a la proteína Hpr1, así como el sitio por el que se une Mip6. No obstante, se sabía muy poco sobre la estructura o la función de Mip6 y su parálogo Pes4. Hemos

probado que Mip6 es una proteína de unión a ARN con cuatro motivos de reconocimiento de ARN que se unen a ARN in vitro con una afinidad alta. Además, su cuarto motivo de reconocimiento de ARN es también el sitio de unión a Mex67. Posteriormente, demostramos que la formación del complejo de Mex67 con el dominio RRM4 de Mip6 compromete su capacidad para unir ARN o viceversa. También diseñamos una mutación puntual en el RRM4 de Mip6 que rompe la interacción con Mex67 pero no con el ARN. Los ensayos posteriores *in vivo* en levaduras nos permitieron establecer una hipótesis sobre el papel de Mip6 como proteína adaptadora para Mex67 en la exportación nuclear, especialmente en condiciones de estrés. Una función adicional de Mip6 era la localización del ARNm que se unía a ella en gránulos de estrés en condiciones de estrés celular.

Además, hemos resuelto las estructuras cristalográficas del RRM3 de Mip6, RRM3 de Pes4, RRM4 de Pes4 y los RRM3 y 4 de Pes4. Todos los RRMs adoptaron una conformación canónica RRM con secuencias RNP1 y RNP2 conservadas generalmente implicadas en la unión a ARN, excepto el RRM3 de Mip6 que carecía del anillo aromático en RNP2. En la estructura sin ARN de los RRM3 y 4 de Pes4, los dominios RRM tándem estaban conectados por una región flexible desordenada y no había un contacto inter-dominio entre ellos. Finalmente, aunque el RRM4 de Pes4 se unía a ARN *in vitro*, no presentaba la capacidad de interaccionar con Mex67 lo cual sugiere una divergencia evolutiva de la función de Mip6 y Pes4.

Resum

L´exportació nuclear d´àcid ribonucleic missatger (mRNA) es un procés complex i essencial per a una correcta expresió gènica en totes cèl·lules eucariotes. L´exportació del mRNA a través del complex del porus nuclear depén principalment de la interacció i coordinació de diverses proteïnes, que formen el que es coneix com mRNPs (ribonucleoproteïnes missatgeres), que tenen un paper dinàmic i interconnectat en les diferents etapes de la biogènesi d´ARNm, com el processament del pre-ARNm, estabilitat, localització i exportació.

Una proteïna clau en aquest procés és MEX67, conservada de llevats fins a humans, que és la principal exportadora de ARN missatger i també està implicada en l´exportació de ARN ribosomal. Mex67 interacciona amb Mtr2 per a formar un heterodímer conservat evolutivament essencial per a una exportació adequada d´ARNm i la consegüent supervivència de la cèl·lula. S´ha estudiat Mex67 durant molts anys, però degut a la complexitat i interconectivitat dels diferents processos de biogènesi d´ARNm, encara queden per descobrir molts detalls de la dinàmica del procés i les interaccions entre Mex67 i les seues moltes proteïnes associades.

En aquest estudi, combinant l´anàlisi bioquímic, biofísic i estructural, hem caracteritzat la interacció entre Mex67 i una nova proteïna associada anomenada Mip6 (proteïna 6 que interacciona amb Mex67). Hem pogut reconstituir un complex estable *in vitro* i estudiar extensivament el mecanisme pel qual interaccionen estes dos proteïnes. També hem resolt l´estructura cristal·logràfica de la regió C-terminal de Mex67 que interacciona amb Mip6 i identificat el domini UBA de Mex67, conegut per unir-se a nucleoporines FG i a la proteïna Hpr1, així com ser el lloc pel que s´uneix Mip6. No obstant, se sabia molt poc sobre l´estructura o la funció de Mip6 i el seu paràleg Pes4. Hem comprobat que Mip6 es una proteïna d´unió a ARN amb quatre motius de reconeixement d´ARN que s´uneixen a ARN *in vitro* amb una afinitat alta. A més, el seu quart motiu de reconeixement d´ARN és també el lloc d´unió a Mex67. Posteriorment, demostràrem que la formació del complex de Mex67 amb el domini RRM4 de Mip6 compromet la seua capacitat per a unir ARN o viceversa. També vam dissenyar una mutació puntual en el RRM4 de Mip6 que trenca la interacció amb Mex67 però no amb l´ARN. Els assajos posteriors *in vivo* en llevats ens van permetre establir una hipòtesi sobre el paper de Mip6 com a proteïna adaptadora per a Mex67 en l´exportació nuclear, especialment en condicions d´estrès. Una funció adicional de Mip6 era la localització de l´ARNm que s´unia a ella en grànuls d´estrès en condicions d´estrès cel·lular.

A més, hem resolt les estructures cristal·logràfiques del RRM3 de Mip6, RRM3 de Pes4, RRM4 de Pes4 i els RRM3 i 4 de Pes4. Tots els RRMs adoptaren una conformació canònica RRM amb seqüències RNP1 i RNP2 conservades generalment implicades en la unió a ARN, excepte el RRM3 de Mip6 que mancava del anell aromàtic en RNP2. En la estructura sense ARN dels RRM3 i 4 de Pes4, els dominis RRM tàndem estàven conectats per una regió flexible desordenada i no hi havia un contacte interdomini entre ells. Finalment, encara que el RRM4 de Pes4 es unia a ARN *in vitro*, no presentava la capacitat d´interaccionar amb Mex67, la cual cosa sugerix una divergencia evolutiva de la funció de Mip6 y Pes4.

1. Introduction

[In the beginning, there was no life, and then nature said,
“Let there be RNA”, and the world came alive.
All forms of life came to be.]

Nucleic acids, DNA (Deoxyribonucleic acid) and RNA (Ribonucleic acid), are the fundamental codes and building blocks of life, all forms of life on earth. In order to grow, function, and replicate, cells in all living beings require DNA, RNA, and proteins, in which the function of each one depends on the other.

For decades, scientists searched for the origin of life. Nowadays, the widely accepted hypothesis among scientists is an “RNA-first world”, as RNA being a putative first precursor of life. For RNA, like DNA, is a genetic information storehouse and like proteins it has the ability to self-fold into secondary structures or even tertiary structures, thus acting as biocatalyst, self-replicating (Robertson & Joyce, 2014), self-cleaving ribozymes (Jimenez, Polanco & Lupták, 2015), and/or in regulating gene expression.

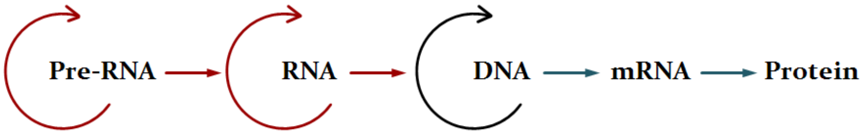


Figure 1.1. RNA world. The hypothesis that RNA preceded DNA in early cells.

Whether RNA was central to early biochemistry, or proceeded by an earlier primitive form of genetic code, RNA molecules are definitely central and fundamental to the function of all modern cells. However, in contemporary biology, the genetic material is copied, transcribed, and translated by complex protein machineries.

Nevertheless, messenger RNA (mRNA) is never alone. Through the short life of mRNA, starting from its transcription and processing in the nucleus to its translation and decay in the cytoplasm, mRNA is always coated and protected by an assembly of multiple RNA-binding proteins (RBPs) (Müller-McNicoll & Neugebauer, 2013). These RBPs are loaded on the mRNA transcripts during all the different interconnecting processes, to ensure the protection, fidelity, regulation and proper execution of all the steps during the mRNA cycle in the correct order of events, which is crucial for proper and efficient gene expression. Although individual RBPs can have several functions on their own, usually they form part of large multi-protein complexes, where all its components interplay, recruiting additional proteins or dissociating others, thus interconnecting the different processes in a dynamic, efficient manner.

In recent years, and due to increasingly evolving advances in both molecular and structural biology, we are having a better understanding of the mechanism of action of these dynamic processes, and a clearer view on the involved molecular players, their interactions, crosstalk, and functions.

However though, as we advance our knowledge on crucial events like mRNA export in which the survival of eukaryotic cells depend on, we are being faced with a far more complex, dynamic, and overlapping processes than previously anticipated that we have yet to uncover. Therefore, shedding light and studying this critical cell process have been a field of interest not only to improve our understanding of how eukaryotic cells function, but also to apply our knowledge on developing future therapeutic strategies for disease associated with mRNA export defects or even targeting cancer cells.

Since the yeast species *Saccharomyces cerevisiae* have been a very useful eukaryotic system, and many cellular processes including the mRNA biogenesis and export have been better studied and understood using yeast, it will be the main focus here bearing in mind the general conservation of this mechanism from yeast to humans.

1.1. Messenger RNA (mRNA) biogenesis: Interconnecting processes

To ensure a proper gene translation, prior mature mRNA transcripts have to be properly synthesized and processed in the nucleus in order to be competent for export to the cytoplasm where it can get translated into functional proteins. This involves various extensive processing steps, and the coupling between these steps is thought to be indispensable for precise and efficient gene expression (Fujiwara, Shiki, & Masu, 2012).

1.1.1. Transcription initiation:

Transcription is the first step of translating genetic information encoded in DNA, and its initiation is a very complex process. In prokaryotes, this process is carried out by a single RNA polymerase species that transcribes all types of RNA, however in eukaryotes, different types of molecular machines called nuclear RNA polymerases; RNA polymerase I (Pol I), Pol II, and Pol III are required. While Pol I transcribes the ribosomal RNA (rRNA) precursor, Pol III transcribes small non-coding RNAs such as transfer RNAs (tRNA). Pol II, on the other hand, is the transcribing enzyme responsible for the transcription of all protein coding genes to produce mRNAs, some other few small nuclear RNAs, and most non-coding RNA genes, thus a major controller of gene expression.

Pol II cannot recognize the promoters of target genes alone (Fujiwara *et al.*, 2012; Woychik & Hampsey, 2002); hence the transcription of these protein coding genes requires the assembly of an evolutionary conserved transcription-initiation complex. This consists of the assembly of Pol II with general transcription factors TFIIB, TFIID, TFIIE, TFIIF and TFIIH at a gene promoter to form the pre-initiation complex (PIC) (Sainsbury, Bernecky, & Cramer, 2015; Woychik & Hampsey, 2002). This complex, especially Pol II activity, is regulated and orchestrated by the not-so conserved Mediator multi-protein complex associated with unphosphorylated carboxy-terminal domain (CTD) region of Pol II (Allen & Taatjes, 2015).

This process is initiated by the binding of TFIID transcription factor that harbours the TATA box-binding protein (TBP) and several other TBP-associated factors (TAFs). TBP, the initiation factor, is conserved and essential for transcription for all promoters by any of the three eukaryotic RNA polymerases. TBP binds directly to the minor groove in DNA with nano-molar affinity, bending the helix by ~90 degrees. TFIIA, the auxiliary factor, works to stabilize this TBP-DNA complex at this stage.

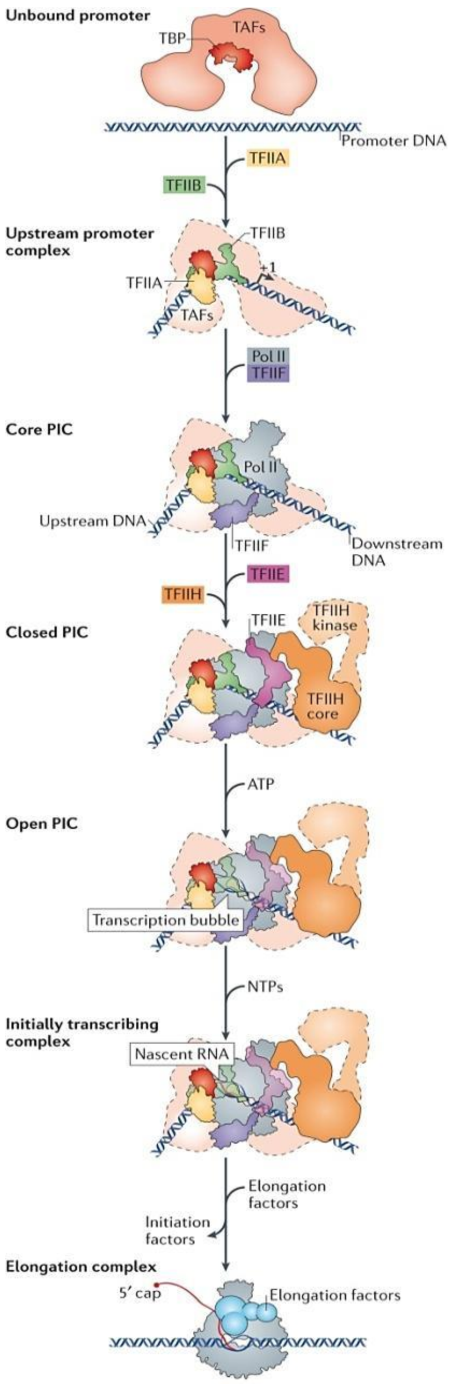

Unbound promoter
TBP
TAFs
Promoter DNA
TFIIA
TFIIB
Upstream promoter complex
TFIIB
+1
TFIIA
TAFs
Pol II
TFIIF
Core PIC
Pol II
Upstream DNA
TFIIF
Downstream DNA
TFIIE
TFIIH
Closed PIC
TFIIE
TFIIH kinase
TFIIH core
ATP
Open PIC
Transcription bubble
NTPs
Initially transcribing complex
Nascent RNA
Elongation factors
Initiation factors
Elongation complex
5′ cap
Elongation factors

Figure 1.2. A schematic representation of Pol II transcription initiation. A model of the stepwise assembly of the general transcription factors and Pol II on promoter DNA (Sainsbury *et al.*, 2015).

Another transcription factor, TFIIB, interact with both TBP and DNA facilitating the DNA bending but importantly recruits Pol II to the promoter. This classical model, of preformed TFIIB-TBP-DNA complex binding to Pol II-TFIIF complex is called the core initiation complex, and is conserved for the transcription systems. TFIIE and TFIIH binding to the later core initiation complex are required for DNA opening. TFIIH, harbouring DNA helicases, melts the DNA in an ATP dependent manner, thus exposing the template strand and thereby allowing the RNA synthesis to begin (Sainsbury *et al.*, 2015).

The later stimulates the phosphorylation of Pol II CTD, releasing its contact with the Mediator and PIC thus leaving the promoter and transitioning from initiation to productive elongation (Allen & Taatjes, 2015).

1.2. Elongation:

Once RNA synthesis starts, biogenesis of eukaryotic mRNA requires that the transcripts produced by RNA polymerase II mature through several post transcriptional dynamic processing steps that must be completed accurately before it can be exported. These processes include 5′capping, splicing, and 3′ end processing/ polyadenylation. The transcribed pre-mRNA undergoes these three major processing steps co-transcriptionally (figure 1.3).

The accurate execution of each of these processes is a large determinant of the fate of a transcript, in which a tight coordination and mechanistic coupling of the consecutive events is required to produce and export functional mRNAs.

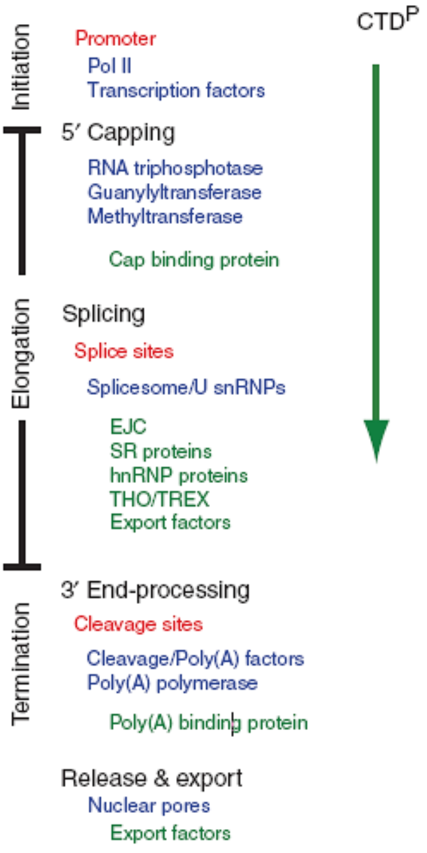

CTD^P
Initiation
Elongation
Termination
Promoter
Pol II
Transcription factors
5′ Capping
RNA triphosphotase
Guanylyltransferase
Methyltransferase
Cap binding protein
Splicing
Splice sites
Spliceosome/U snRNPs
EJC
SR proteins
hnRNP proteins
THO/TREX
Export factors
3′ End-processing
Cleavage sites
Cleavage/Poly(A) factors
Poly(A) polymerase
Poly(A) binding protein
Release & export
Nuclear pores
Export factors

Figure 1.3. mRNA processing steps coupled to transcription. The phosphorylated CTD of Pol II critical for the elongation progression and mRNA processing is shown on the right (Hocine, Singer, & Grünwald, 2010).

1.2.1. Co-transcriptional 5′Capping:

During the first processing event, the eukaryotic RNA polymerase transcripts are modified by the addition of a 7-methyl guanosine cap (m^{7}GpppN) on its 5′ -end. This process is carried out co-transcriptionally by the enzymatic action of RNA triphosphatase (Tpase), RNA guanylyltransferase (GTase), and guanine-N7 methyltransferase (guanine-N7 MTase) which associates with Pol II phosphorylated CTD (Figure 1.3).

Once the nascent pre-mRNA is capped, the nuclear cap-binding protein complex (CBC) binds with high affinity to the 5′cap. CBC, a heterodimer composed of two subunits, cap binding proteins CBP20 and CBP80, seems to be the first protein complex to assemble with the pre-mRNAs (Andrés Aguilera, 2005).

This assembly is essential for the subsequent steps of mRNA processing. mRNA capping is critical, not only because it protects the pre-mRNA from 5′-3′exonuclease cleavage, but also for the recruitment of the splicing machinery through the nuclear CBC, polyadenylation, and ultimately nuclear export (Ramanathan, Robb, & Chan, 2016).

1.2.2. Coupling to splicing:

The precise removal of non-coding intervening sequences, or introns, from the pre-mRNA transcripts is crucial for proper gene expression (Hocine, Singer, & Grünwald, 2010). In high eukaryotes the vast majority of nascent RNA transcripts undergo this crucial editing process to remove introns. The CBC enhances the recruitment and assembly of the spliceosome, an evolutionary conserved MDa complex composed of 5 small nuclear RNAs and their associated proteins (snRNP) to the 5′splice site. Subsequent transesterification reactions carried out by the snRNP leads to the excision of the introns and the ligation of the exon sequences in order to generate a mature mRNA transcript.

An important splice-dependent mark is the deposition of the multiprotein exon-exon junction complex (EJC) to the spliced junctions by the spliceosome, 20-24 nucleotides upstream of exon-exon junctions (Le Hir *et al.*, 2000). This dynamic post splicing complex serves as a binding platform for the proteins involved in the later mRNA processing as it accompanies mRNAs to the cytoplasm. EJC plays a role in the recruitment of the export machinery to the spliced transcripts, thus enhancing the efficiency of nucleo-cytoplasmic export of spliced mRNA. It does so through proteins associated with the EJC, such as ALYREF (Yra1 in S. cerevisiae) a component of the TRancription-EXport (TREX) complex, that serves as an adaptor protein and a binding site for the co-transcriptional recruitment of the export factor Mex67 to the maturing transcripts(Le Hir, Gatfield, Izaurralde, & Moore, 2001; Strässer *et al.*, 2000).

In humans, Few shuttling members of the serine-arginine (SR) proteins family involved in splicing also play a role in mRNA export, coupling splicing to export, by serving as adaptor proteins for the export factor, TAP (Mex67 human homolog), in mRNA export (Huang, Gattoni, Stévenin, & Steitz, 2003).

1.2.3. 3′ Polyadenylation and transcription termination:

Next, 3′end processing of the spliced, protein coding, transcripts, consists of an endonucleolytic cleavage at the 3′end followed by polyadenylation where a poly-adenosine tail (Poly-A tail) of approximately 70-90 nucleotides (250-300 in humans) is added. This process is required to ensure the maturity and stability of the transcribed gene and for efficient export and translation later on.

The newly synthesized mRNA contains several sequenced elements (*cis* elements) recognized by cleavage factor complexes to define the cleavage site (figure 1.4.).

Once the transcript have been cleaved, It is released from Pol II transcription machinery, and the poly(A) polymerase (PAP) starts adding the adenosines that comprise the poly (A) tail aided by poly A binding proteins (PABPs). This process paves the way to the release the mature mRNA from the transcription site. And like all co-transcriptional processes, it is tightly coupled with mRNA export.

In yeast, poly (A) binding proteins such as Nab2 are also an export adaptor that exports the nucleus with the mature messenger ribonucleoproteins (RNPs). Its function is required for efficient polyadenylation and nuclear export as it links the two processes and any loss in its expression would lead to hyperadynylation and nuclear accumulation of poly(A)$^+$ RNA (Hector *et al.*, 2002; Kuhlmann, Valkov, & Stewart, 2014).

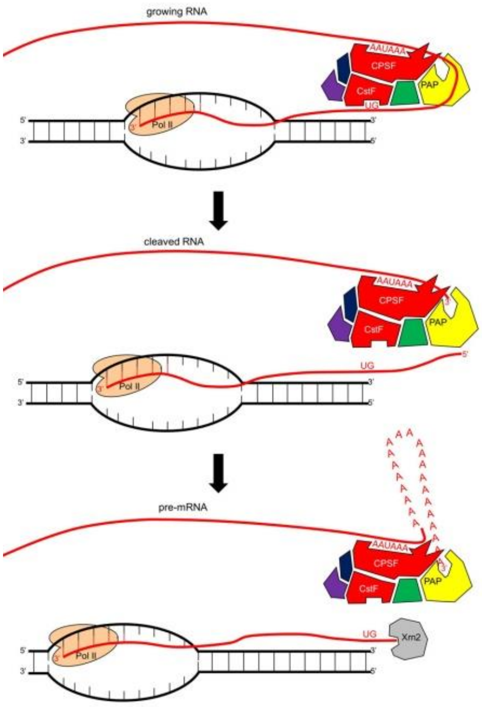


Figure 1.4. 3´ Polyadenylation and transcription termination. Cleavage and polyadenylation specificity factor (CPSF), the core component of the 3´end processing complex, specifically recognizes the core sequence element AAUAAA in the polyadenylation site and catalyses cleavage. Whereas the cleavage stimulating factor (CSTF) recognizes the U/GU-rich elements downstream the AAUAAA sequence. Other cleavage factors are involved in this process.

1.3. Nucleo-cytoplasmic RNA export

Eukaryotic cells are defined by their compartmentalization into different organelles that serve as the sites for different cellular activities (Jun Katahira, 2012). One of the most important features of this definition is the physical separation of the nucleus and the cytoplasm by a nuclear envelope, which ensures and controls molecular exchanges between the nucleus and the cytoplasm through a specialized and evolutionary conserved structure, the nuclear pore complex (NPC). This nuclear pore complex acts as a two-way gate, allowing macromolecules such as RNA, proteins, or ribonucleoprotein (RNP) particles to be transported in and out from the nucleus to cytoplasm and vice versa (Kabachinski & Schwartz, 2015; Knockenhauer & Schwartz, 2016).

While proteins are synthesized in the cytoplasm, the nucleus is the storage site of the DNA and the site of transcription and processing of the different types of RNA. Therefore, the export of RNA molecules from the nucleus to the cytoplasm is fundamental for gene expression. The RNA molecules exported through the NPC include not only messenger mRNA, but also ribosomal rRNA, transfer tRNA, microRNA (miRNA), small nuclear RNA (snRNA). While the biogenesis of some types of RNA (snRNA) requires the transport in and out of the nucleus, others like mRNA and tRNA move exclusively in one way direction out of the nucleus. Their journeys are aided by mobile transport factors, and is mediated almost solely through the NPC (Fujiwara *et al.*, 2012; Knockenhauer & Schwartz, 2016; Köhler & Hurt, 2007).

This dynamic transport of cargoes through the NPC requires the two major transport systems in eukaryotes: The nuclear transporters of the importin/exportin-β super family, commonly known as karyopherins, which are further categorized into importins and exportins, and the nuclear exporter Mex67-Mtr2 (Nxf1-Nxt1/Tap-p15 in humans) heterodimer (figure 1.5.).

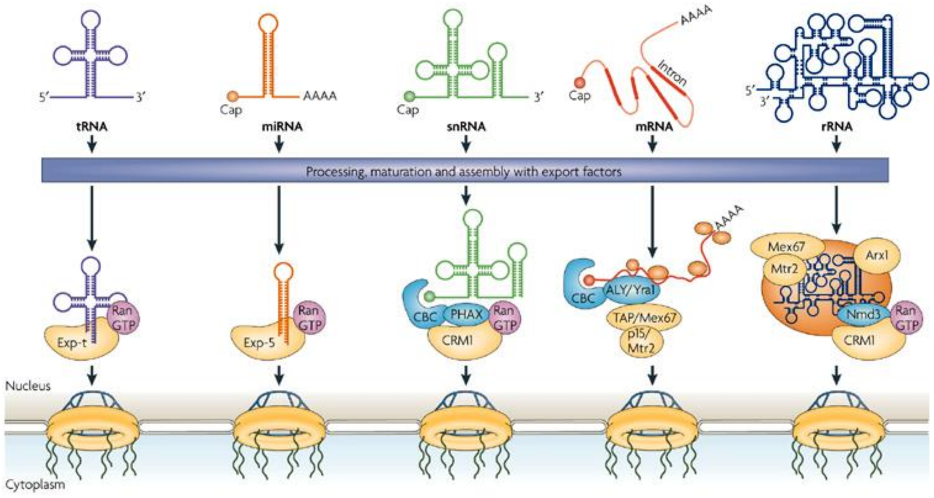

5′
3′
tRNA
Cap
AAAA
miRNA
Cap
3′
snRNA
Cap
Intron
AAAA
mRNA
5′
3′
rRNA
Processing, maturation and assembly with export factors
Ran GTP
Exp-t
Exp-5
CBC
PHAX
CRM1
ALY/Yra1
TAP/Mex67
p15/Mtr2
AAAA
Mex67
Mtr2
Arx1
Nmd3
Nucleus
Cytoplasm

Figure 1.5. Major RNA nuclear export routes. Major export factors for each type of RNA with their adaptor proteins are shown in the figure (Köhler & Hurt, 2007).

1.3.1 RNA export factors

Exportins:

Most nucleo-cytoplasmic transport is signal-mediated. Karyopherins, or importin β family members, responsible for the major part of that transport, are a conserved super-family of transport nuclear factors. They function by the recognition/binding of short peptide signals on cargo proteins, in a guanosine triphosphate GTP-bound form of the guanine nucleotide binding protein Ran (Ran-GTP) gradient dependent manner. They can recognize nuclear localization signals (NLS) and nuclear export signals (NES) in proteins to direct them, through the NPC, in or out of the nucleus respectively (Chook & Süel, 2011; Köhler & Hurt, 2007).

That is to say, karyopherins include nuclear import factors (importins) that recognize the NLS in the cargo protein, and nuclear export factors (exportins) that recognize the NES sequence.

Exportins recognize and bind, chiefly, the NES sequences which are usually a hydrophobic, classically leucine rich, short peptide sequences in cargo proteins. The export of some classes of RNA similar to that of proteins follows the general paradigm of exportins-mediated protein export, though each might require different exportins. Processed and mature tRNA is mainly exported by exportin-t (Xpot, Los1 in yeast) by binding directly to the tRNA in a RAN-GTP dependent manner, while the non-coding, gene regulating, miRNA is mostly exported by exportin-5 (Xpo5 or Exp5) in a similar simple conserved export route (Köhler & Hurt, 2007; Okamura, Inose, & Masuda, 2015). On the other hand, exportin-1 (Xpo1), or better known as Crm1 (Chromosome maintenance region 1) plays a major role in snRNAs and rRNA export (figure 1.5).

Crm1/ Xpo1

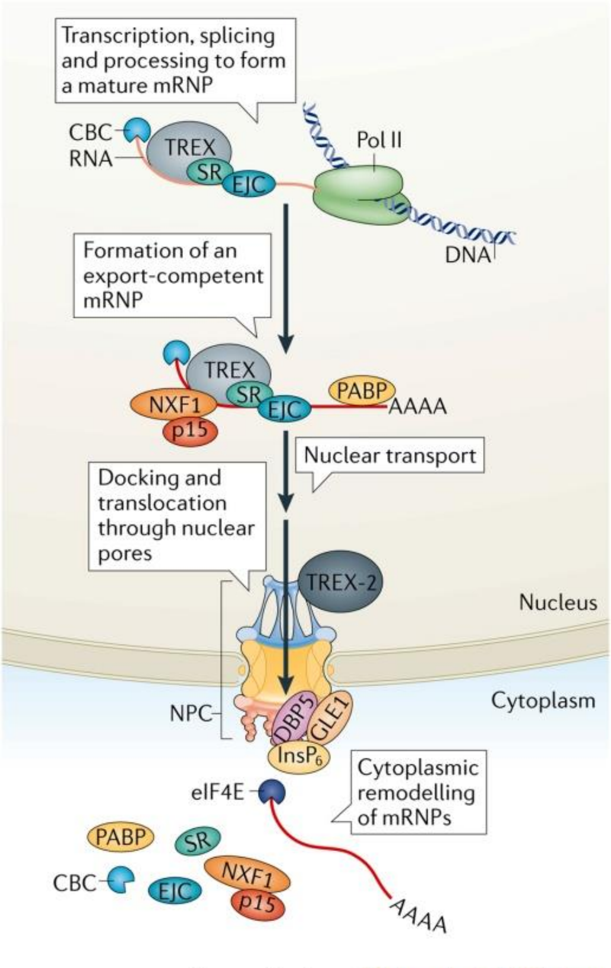
Crm1 or Xpo1, first characterized as an exporter binding NES sequence of the viral HIV-1 Rev protein (Fornerod, Ohno, Yoshida, & Mattaj, 1997), is one of the most conserved RanGTPase-driven nuclear export factors that is structurally well characterized (X. Dong *et al.*, 2009; Güttler *et al.*, 2010; Monecke *et al.*, 2009). Like other exportins, the directionality of Crm1 exported cargoes is driven and regulated by the asymmetric distribution of Ran-GTP (nucleus) and Ran-GDP (cytoplasm). The cargo-exportin-RanGTP complexes in the nucleus drive the export through the nuclear pore to the cytoplasm, where the cargo is released upon the hydrolysis of Ran-GTP by Ran-GTPase activating protein (Ran-GAP).

The Importance of Crm1 comes from the fact that it exports a wide range of substrates from the nucleus, including ribosome, mRNPs, snRNA, and many regulatory proteins. It was also suggested that Crm1 also plays a role in the nucleo-cytoplasmic shuttling of mature microRNAs (Castanotto, Lingeman, Riggs, & Rossi, 2009), and is essential for replication of viruses such as HIV. But unlike other exportins such as Xpot and Xpo5, Crm1 cannot bind directly to its targeted RNA; instead it requires adaptor proteins bound to RNA. It is an export factor for leucine rich NESs (Fornerod *et al.*, 1997), where it binds physically to the NES sequences in the associated adaptor RNPs. however, it is also known for Crm1 to recognize not only short peptides but also large portion of the export cargo (Monecke *et al.*, 2009). The mechanism of recognition in which Crm1 is able to recognize its wide range of substrates and its ability to carry multiple cargoes is yet to be better understood.

1.4. mRNA export: From transcription site to the nuclear pore

Like others RNAs, mRNA is transcribed in the nucleus, but unlike the export of many RNAs, mRNA export is exclusively unidirectional, mechanistically different process. Although the classical Crm1/Xpo1 exportin was demonstrated to be involved in the export of only small groups unspliced or partially spliced of mRNAs (Cullen, 2003; S. Dong *et al.*, 2007; Gallouzi & Steitz, 2001), it doesn´t have a major role in mRNA export (Neville & Rosbash, 1999; Reed, Clouse, Luo, & Zhou, 2001). Instead, Mex67/Mtr2 heterodimer in yeast (TAP/P15 or NFX1/NFT1 in metazoan) is the general nuclear exporter of mRNA in the form of mature ribonucleoproteins (mRNP) particle. This heterodimer is evolutionary conserved from yeast to humans (J. Katahira *et al.*, 1999). It is structurally unrelated to the classical karyopherins/exportin transport system, and does not obey the classical RanGTP-RanGDP gradient directionality paradigm (Reed *et al.*, 2001). Instead, mRNA export is known as a facilitated and energy-dependent process (Dargemont & Kühn, 1992). Although, like karyopherins, Mex67/Mtr2 interacts directly with the FG repeats of the NPC nucleoporins (Askjaer *et al.*, 1999; Bayliss, Littlewood, & Stewart, 2000; Sträßer, Baßler, & Hurt, 2000).

Mature ribonucleoprotein (RNP) complex formation is a hallmark of the mRNA export, not required for the export of other types of RNA. Like rRNA, mRNA is never alone and definitely doesn´t travel alone. Prior to export, nascent transcripts must be 5´capped, spliced, and polyadenylated correctly as mentioned previously. RBPs associate with the nascent pre-mRNA transcripts early on during transcription and elongation/maturation processes ensuring the tight coupling between mRNA biogenesis and export. RBPs are usually components of multi-protein complexes recruited to the mRNA transcripts in a carefully orchestrated series of processing steps, which are all kept under tight surveillance before being exported to the cytoplasm. These RBPs usually contain RNA recognition/binding domains to interact with the mRNA along the way to produce a mature, exportable mRNP particle (Kurshakova, Georgieva, & Kopytova, 2016; Tutucci & Stutz, 2011). Although the exact mechanism of packaging of mRNP is not yet fully understood, and many more RNA-binding proteins that play a role in its formation are yet to be discovered, we know that the composition and formation of an export-competent mRNP complex, is highly dynamic where its proper packaging is critical for the regulation and export


Transcription, splicing and processing to form a mature mRNP
CBC
RNA
TREX
SR
EJC
Pol II
DNA
Formation of an export-competent mRNP
TREX
SR
NXF1
p15
EJC
PABP
AAAA
Nuclear transport
Docking and translocation through nuclear pores
TREX-2
Nucleus
Cytoplasm
NPC
DBP5
GLE1
InsP6
eIF4E
Cytoplasmic remodelling of mRNPs
PABP
SR
CBC
EJC
NXF1
p15
AAAA

Nature Reviews | Molecular Cell Biology

Figure 1.6. Steps of mRNP Export from transcription to the cytoplasm. All processes are coupled and overlapped (Wickramasinghe & Laskey, 2015)

of the mRNA from the nucleus. RBPs recruit the Mex67/Mtr2 heterodimer to the mRNP.

Although Mex67, unlike Crm1, can itself bind with mRNA (S. Aibara, Valkov, Lamers, & Stewart, 2015), it operates with many of the adaptor RNA-binding proteins that act as bridges, mediating and further stabilizing the interaction between the mRNA and its Mex67/Mtr2 heterodimer export factor. But before an mRNP can enter the NPC, a tight quality control is performed during the whole, pre-export, process to ensure that aberrant RNAs are degraded, and only mature mRNPs are exported. This nuclear quality control is, mostly, carried out co-transcriptionally by members of the TRAMP (trf4-Air2-Mtr4p polyadenylation) and exosome protein complexes, which survey mRNA, and identify and degrade any defective mRNPs (Chlebowski, Lubas, Jensen, & Dziembowski, 2013; Kilchert & Vasiljeva, 2013). Mature mRNPs are then escorted to the nucleoplasmic side of the NPC, or more specifically, the nuclear basket by the TREX and TREX2 (transcription –export complexes) (Fischer *et al.*, 2002; Sträßer *et al.*, 2002). Once mature mRNP is assembled and targeted to the NPC, Mex67 promotes the docking of mRNP to the nuclear basket of the NPC and the translocation through the central channel of the NPC via its physical interaction with the phenylalanine-glycine (FG repeats) of the nucleoporins. mRNA transcripts are always exported in a 5′-3′direction.

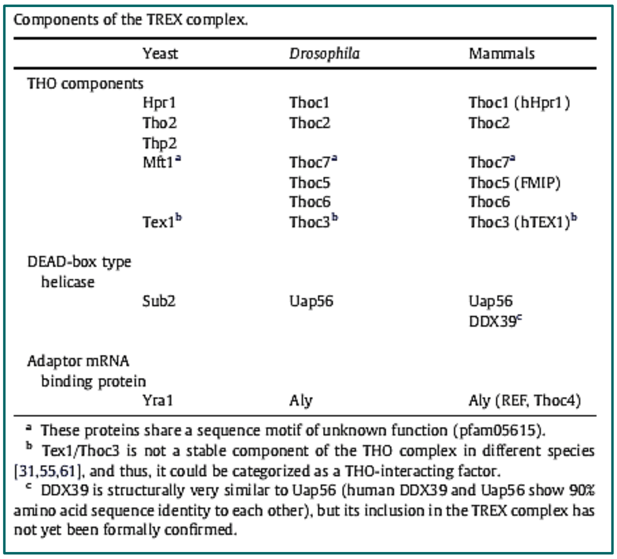
1.4.1. mRNA export adaptors

Although mRNP biogenesis and export steps are often described separately for simplicity, they are tightly coupled and interdependent, with some factors participating in multiple stages. Importantly, the early requirement of proteins not only allows efficient pre-mRNA processing and mRNP assembly but also protects the mRNA from degradation through continuous competition with the nuclear surveillance machineries (Tutucci & Stutz, 2011).

Among the big number of adaptor proteins/complexes that play a role in mRNA export, many have been identified, but it is likely that more will be uncovered among the huge number of the predicted RBPs in eukaryotes. Due to the limited space we have, here we are going to discuss, in brief, the key adaptor proteins/complexes in the mRNA export machinery.

1.4.1.1 THO/TREX Complex

The THO complex, a nuclear protein complex with a key role in the co-transcriptional formation of the export-competent mRNPs, is conserved from yeast to humans. The yeast THO has a stable core of tetrameric sub-complex formed by Tho2, Hpr1, Mft1, and Thp2 associated with a fifth protein named Tex1 (figure 1.7). In metazoan, THO contains 3 more subunits THOC5, THO6, and THOC7 that does not appear to have yeast homologs (Rehwinkel *et al.*, 2004).



Components of the TREX complex.

	Yeast	*Drosophila*	Mammals
THO components			
	Hpr1	Thoc1	Thoc1 (hHpr1)
	Tho2	Thoc2	Thoc2
	Thp2		
	Mft1[a]	Thoc7[a]	Thoc7[a]
		Thoc5	Thoc5 (FMIP)
		Thoc6	Thoc6
	Tex1[b]	Thoc3[b]	Thoc3 (hTEX1)[b]
DEAD-box type helicase			
	Sub2	Uap56	Uap56 DDX39[c]
Adaptor mRNA binding protein			
	Yra1	Aly	Aly (REF, Thoc4)

[a] These proteins share a sequence motif of unknown function (pfam05615).

[b] Tex1/Thoc3 is not a stable component of the THO complex in different species [31,55,61], and thus, it could be categorized as a THO-interacting factor.

[c] DDX39 is structurally very similar to Uap56 (human DDX39 and Uap56 show 90% amino acid sequence identity to each other), but its inclusion in the TREX complex has not yet been formally confirmed.

Figure 1.7. Components of the THO/TREX complex involved in coupling co-transcriptional elongation to mRNA export (Jun Katahira, 2012).

THO has a key role in the mRNA co-transcriptional elongation during mRNPs biogenesis by Pol II. However, recently it was also shown to enhance transcription initiation and elongation by RNA Pol I (Yinfeng Zhang, French, Beyer, & Schneider, 2016). Loss of any of the

THO subunits impairs transcription elongation, genome stability, and mRNA export.

Expected to be one of the first players to act co-transcriptionally, helping recruiting other adaptor proteins to ensure a correct packaging of mRNPs and further recruiting the export machinery required for its export, THO was also shown crucial to prevent the formation of DNA-RNA hybrids (R-loops) by binding to chromatin and to the nascent mRNA (Huertas & Aguilera, 2003). Hpr1, being the first known key component in the THO complex, was first identified in genetic screen for hyper-recombinant mutants in *S. cerevisiae* (A Aguilera & Klein, 1988). Further studies found a role for Hpr1 in transcription elongation and mRNA export as it recruits the export factor Mex67 to the mRNP through the ubiquitin-dependent interaction with Mex67 C-terminal ubiquitin associated (UBA) domain (Gwizdek *et al.*, 2006; Hobeika *et al.*, 2009).

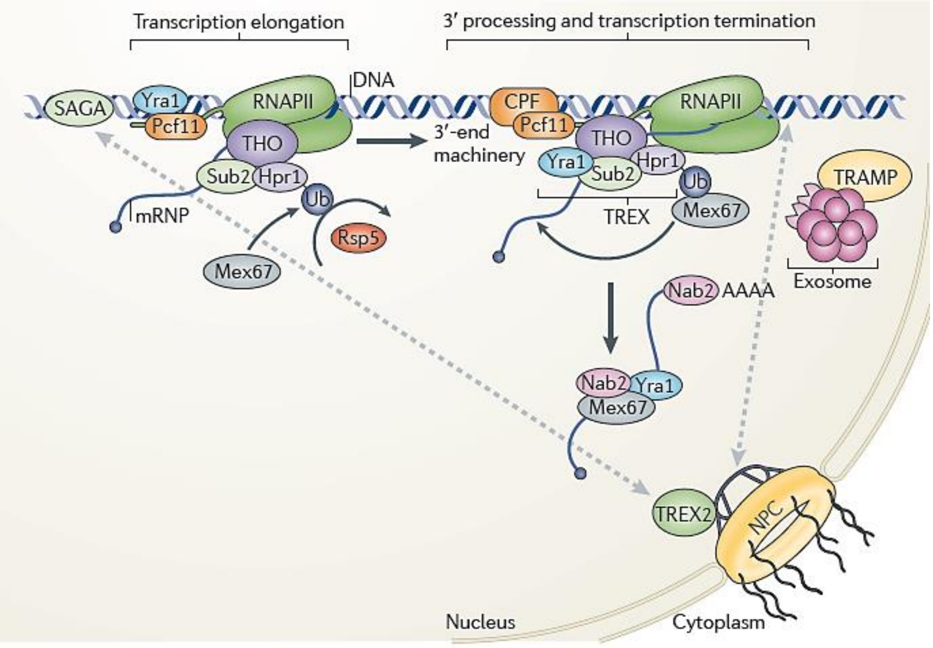

Transcription elongation
3′ processing and transcription termination
DNA
SAGA
Yra1
Pcf11
RNAPII
THO
Sub2
Hpr1
Ub
mRNP
Rsp5
Mex67
CPF
Pcf11
3′-end machinery
RNAPII
THO
Yra1
Sub2
Hpr1
Ub
TREX
Mex67
TRAMP
Exosome
Nab2
AAAA
Nab2
Yra1
Mex67
TREX2
NPC
Nucleus
Cytoplasm

Figure 1.8. THO/TREX complex co-transcriptional arrangement and its role in the recruitment of the mRNA export machinery (Tutucci & Stutz, 2011).

Two additional adaptor proteins named Yra1 (ALY in humans) and Sub2 (Uap56 in humans) are then associated with the THO to form the conserved TRanscription-EXport (TREX) complex (Sträßer *et al.*, 2002) which, as the name indicates, couples transcription to export. Sub2, a dead-box ATPase required for spliceosome assembly process (Libri, Graziani, Saguez, & Boulay, 2001), is recruited to the transcription machinery via the THO complex. Once bound, Sub2 assumes a semi-open conformation stimulating its ATPase activity, allowing the recruitment of the adaptor protein Yra1 where they interact and cooperatively bind to RNA (figure 1.9.) (Ren *et al.*, 2017; Sträßer & Hurt, 2001). Previous studies have reported that Sub2 is dispensable for Yra1 recruitment; instead yra1 is recruited by interacting with Pcf11 (the Pol II CTD binding subunit of cleavage-polyadenylation factor CF1A) (Johnson, Cubberley, & Bentley, 2009; Saguez & Jensen, 2009). Yra1 is also associated with cap-binding and the EJC during splicing as mentioned earlier.

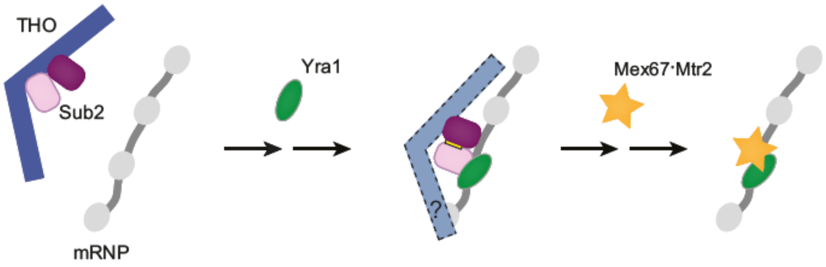


Figure 1.9. A model of TREX-mediated mRNP remodelling prior to nuclear export (Ren *et al.*, 2017).

The THO complex assembly with Yra1 and Sub2, is ATP-dependent which is also conserved with the human homologs (Dufu *et al.*, 2010; Ren *et al.*, 2017). This TREX complex packing, important for a stable mRNP formation and export (Zenklusen, Vinciguerra, Wyss, & Stutz, 2002), subsequently recruits and assists in loading the export factor Mex67/Mtr2 to the mRNPs, through the direct interaction between Yra1 and Mex67 (Strässer *et al.*, 2000). However, Sub2 and Mex67 compete for the same binding site on Yra1 (figure 1.9.). That is to say, after Yra1 is recruited to TREX by sub2, it dissociates from Sub2 as Mex67 binds (Sträßer & Hurt, 2001).

While yeast THO/TREX is associated transcriptionally with all actively transcribed genes since most genes are intronless, human TREX requires 5´capping and Its assembly is coupled to transcription, indirectly, through splicing (Masuda *et al.*, 2005), suggesting a regulation mechanism where Mex67 is exclusively recruited to spliced transcripts, thus making sure only correctly spliced/processed mRNPs can export the nucleus. However, some studies indicated that splicing-independent mechanisms also seem to operate in higher eukaryotes (Chung *et al.*, 2016).

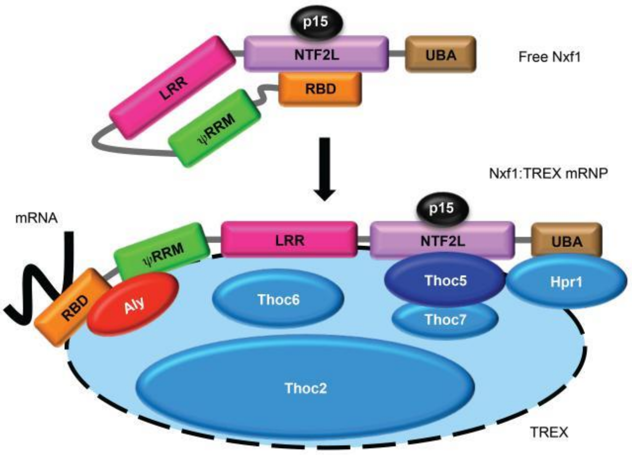


Figure 1.10. The conformational changes in TAP (NXF1) induced by TREX binding prior to mRNA export (Viphakone *et al.*, 2012).

In humans, Mex67 in its free form exhibits a closed conformation. It´s recruitment to the mRNP by the TREX complex, and its contact with TREX subunits remodels Mex67 to an open conformation, exposing its RNA binding domain (Viphakone *et al.*, 2012). The TREX adaptor proteins thus handover the mRNA to Mex67 and enhances its RNA-binding ability essential for its export function (Hautbergue, Hung, Golovanov, Lian, & Wilson, 2008) (figure 1.10.). Whether the yeast counterpart utilizes a similar mechanism to bind mRNA is not so clear.

This recruitment of Mex67/Mtr2 to the mRNP via TREX brings Mtr2/Mex67/mRNP closer to the nuclear pore that will then facilitate the mRNP transverse through the NPC.

Other major known adaptors of Mex67 involved in mRNA export is the poly (A) RNA binding protein Nab2 mentioned earlier, that is capable of interacting spontaneously with Mex67 and Yra1, where Yra1 enhances the Mex67/Nab2 interaction (Iglesias *et al.*, 2010). However, unlike Yra1 that dissociates from Mex67 prior to nuclear export, Nab2 travels with the mRNP and Mex67 through the NPC where it is released in the cytoplasm (Tran, Zhou, Corbett, & Wente, 2007). Another multifunctional protein that participates in a variety of RNA-related processes, including transcription, splicing, mRNA export, and translation is Npl3 (Santos-Pereira, Herrero, Moreno, & Aguilera, 2014). Npl3 is a conserved mRNA binding export adaptor that packs co-transcriptionally and shuttles with mRNPs (Lei, Krebber, & Silver, 2001).

Bearing that in mind, the precise mechanism of the dynamic and complex crosstalk between mRNA and all the different proteins in the TREX-mediated mRNP assembly awaits further understanding.

1.4.1.2 TREX2 complex

Close to the nuclear pore, another complex named TREX2 has been recently described to further couple Transcription and mRNA export. TREX2 is an evolutionary conserved complex composed of five subunits Sac3, Sus1, Thp1, Cdc31, and Sem1 in yeast, and their homologs in human.

First identified as a sac3/Thp1 complex, it was proposed that this sac3/Thp1 complex functions in docking the export machinery to the nuclear basket of the NPC (Fischer *et al.*, 2002). That was demonstrated through the binding ability of sac3 to Mex67 on one hand, and its ability to interact with Nup1 and Nup60 nucleoporins at the nuclear face of the NPC on the other hand. In yeast, like THO mutants, deletion of any of TREX-2 subunits results in mRNA export defects. Moreover, recently a study showed that defects in THO and TREX-2 assembly causes accumulation of mRNP granules (Eshleman, Liu, McGrath, Parker, & Buchan, 2016). Further studies have confirmed the positioning of the TREX-2 at the nuclear pore (Rodríguez-Navarro *et al.*, 2004). However, a recent study suggested that TREX2 is a stably associated NPC complex, possibly with other functions independent of transcription (Umlauf *et al.*, 2013). Sac3 was also shown to interact with Sub2 thus additionally proposing an involvement in coordinating co-transcriptional processing with export.

The TREX-2 complex is based on a Sac3 scaffold to which Thp1, Sem1 associate with the middle part whereas Cdc31 and two copies of Sus1 bind to its C-terminal domain (CID) (Shintaro Aibara, Bai, & Stewart, 2016) (figure. 1.11). The Sac3/Sem1/Thp1 association results in a structural conformation necessary for mRNA binding (Marius Boulos Faza *et al.*, 2009). Additionally, the FG like repeats in the N-terminus of Sac3 binds the Mex67/Mtr2 exporter (Dimitrova *et al.*, 2015), Therefore escorting the export machinery Mtr2/Mex67/mRNP to the NPC.

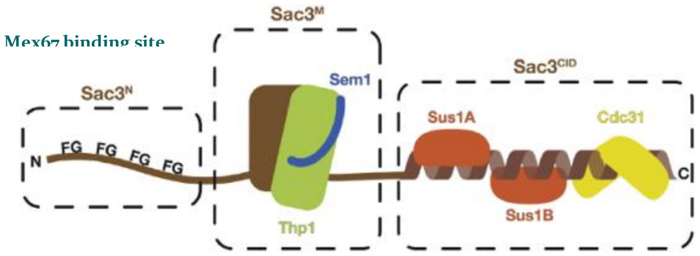


Figure 1.11. Sac3 serves as a scaffold for which the components of TREX-2 bind and Mex67 binds. (Shintaro Aibara *et al.*, 2016).

Interestingly, Sus1 is also a component of the SAGA complex involved in histone modifications (acetylation and de-ubiquitination) and transcription activation (Rodríguez-Navarro *et al.*, 2004), thus linking transcription initiation to export. A recent study emphasised this relation by showing that TREX-2 furthermore directly interacts, with the mediator complex associated with Pol II during transcription initiation (Schneider *et al.*, 2015). Thus connecting the transcription site to the export machinery and bringing transcribed genes closer to the NPC facilitating mRNP export to the cytoplasm, especially for actively transcribed genes. However, how TREX-2 regulates this gene expression machinery and its exact mechanism of action is yet to be uncovered.

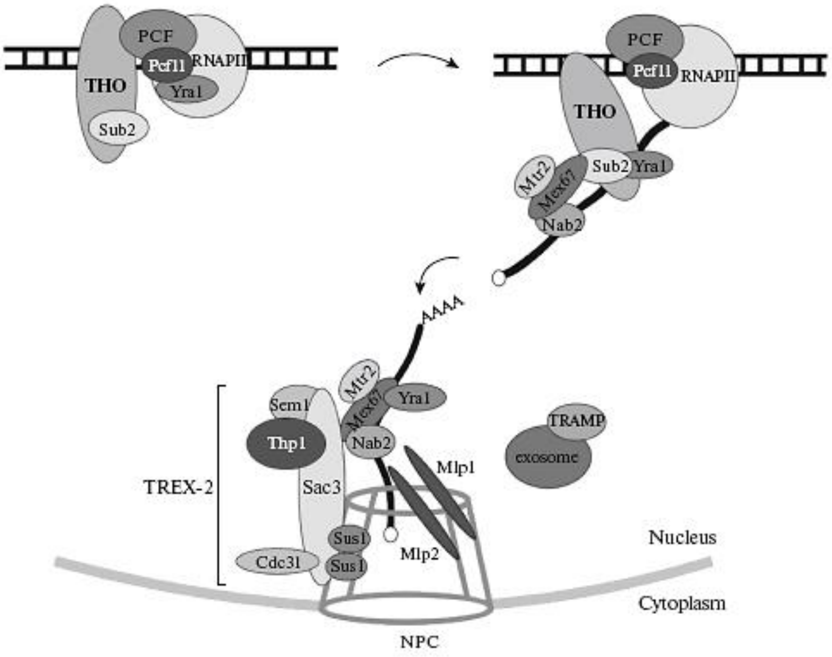

PCF
Pcf11
RNAPII
THO
Yra1
Sub2
PCF
Pcf11
RNAPII
THO
Mtr2
Mex67
Sub2
Yra1
Nab2
AAAA
Sem1
Thp1
Mtr2
Mex67
Yra1
Nab2
TRAMP
exosome
TREX-2
Sac3
Mlp1
Sus1
Sus1
Cdc31
Mlp2
Nucleus
Cytoplasm
NPC

Figure 1.12. Mex67-dependent mRNA nuclear export. The recruitment of Mex67/Mtr2 by THO/TREX complex, and later the interaction of the mRNP particle with the TREX-2 complex positioned at the nuclear face of the NPC (Kurshakova *et al.*, 2016).

1.5. rRNA export

The large pre-60S and the small pre-40S ribosomal subunits are assembled in the nucleus before being transported to the cytoplasm to contribute to the polysomic formation. Like mRNA, the ribosomal subunits export the nucleus in the form of RNPs. Due to the fact that the ribosomal subunits are one of the largest RNA-containing particles to pass through the NPC channel, studies have shown that its export involves several export factors that can work cooperatively, but independently to ensure efficient export of the ribosomal subunits and prevent any disassembled ribosomal RNP or backsliding to the nucleus. Notably, unlike the case of mRNA, the export machinery is recruited to the pre-ribosomal particles at a late stage after it had already undergone complicated steps of processing and maturation.

Subsequently, the mature pre-60S and pre-40S subunits follow separate export routes. Currently, we better understand the pre-60S subunit export but many aspects about the pre-40S subunit export are not yet well understood. What's known so far is a role for three export receptors: exportin Crm1/Xpo1, Mex67/Mtr2, and Arx1 (figure 1.13.).

Crm1 requirement for the export of both pre-60S and pre-40S rRNA in a RanGTP dependent manner is conserved from yeast to humans, however as mentioned earlier it requires different adaptor proteins in each case since it cannot bind the rRNA on its own. Nmd3 is well known to be the NES-containing adaptor protein required by Crm1 in the case of pre-60S export (Sengupta *et al.*, 2010). On the other hand, Arx1, can bind and accompany the pre-60S through the NPC but whether it has a role to play in the pre-40S export is not yet known (Bradatsch *et al.*, 2007).

Like Crm1, Mex67/Mtr2 heterodimer have been also recently described to be involved in the export of both pre-60S and pre-40S rRNA (Marius B. Faza *et al.*, 2012; Yao *et al.*, 2007). Both Mex67 and Arx1, unlike Crm1, accompany the rRNA through a direct contact with the ribosomal subunits. Mex67 has been demonstrated to bind the pre-60S and 5S ribosomal RNA in vitro (Yao *et al.*, 2007). All three exporters translocate the rRNA through direct interaction with the FG nucleoporins lining the NPC channel, although none of them is structurally related. Whether TAP, the human homolog of Mex67 also functions in rRNA is unclear. Additionally, whether unknown Mex67 adaptor proteins contribute in this process is not yet well understood. Moreover, Npl3 known as an export adaptor for Mex67 in the mRNP export machinery have also been characterized to mediate the export of pre-60S ribosomal subunit independently of Mex67 (Hackmann, Gross, Baierlein, & Krebber, 2011).

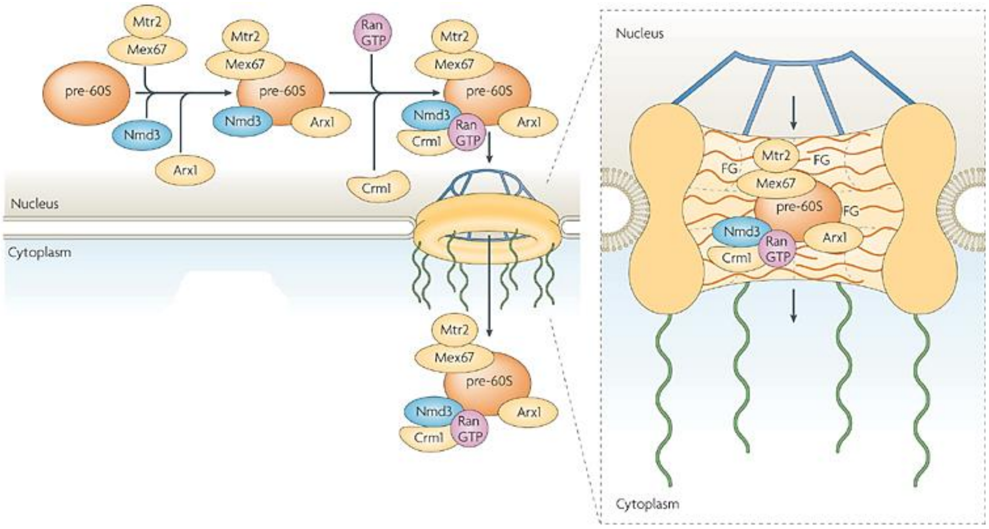


Figure 1.13. The Pre-60S Export in yeast (Köhler & Hurt, 2007).

1.6. Nuclear Pore Complex (NPC)

After the mature mRNP is properly packaged and docked to the NPC nuclear face, the clear next step is for the mRNP to pass through the NPC irreversibly to the cytoplasm.

Despite the dynamic protein composition of the NPC in different species, the structure and function of the NPS in eukaryotes is remarkably evolutionary conserved (Knockenhauer & Schwartz, 2016; Lin *et al.*, 2016). It is, almost, the sole mean of communication between the nucleus and cytoplasm. The NPS are vast protein assemblies, embedded in the nuclear envelope, acting like a channel or a gate that regulates the selective and massive transport of macromolecules between these cellular compartments (Knockenhauer & Schwartz, 2016; Niño, Hérissant, Babour, & Dargemont, 2013). These compartments need to communicate extensively between each other, and therefore molecules need to be exchanged frequently, reliably, and efficiently across membranes. Thus, NPC serves not only as a nucleoplasmic mediator, but also as a regulator of gene expression (Lin *et al.*, 2016).

The NPC is one of the largest protein assemblies in the cell, with an estimated size of ~50 MDa in yeast NPC to ~120 MDa for that of vertebrates. Due to its massive size and flexibility, scientists had hard time determining its structure. Our most accurate current view of the NPC today comes from biochemical and structural studies using x-ray crystallography for separate sub complexes in addition to cryo-EM (Hoelz, Glavy, & Beck, 2016). The work from Lin D. *et al.* recently published in the journal *Science* have combined previous structural studies with additional work to reconstitute a high resolution, fairly accurate, biochemical interaction map of the symmetric core of the nuclear pore (Lin *et al.*, 2016, Figure 1.14).

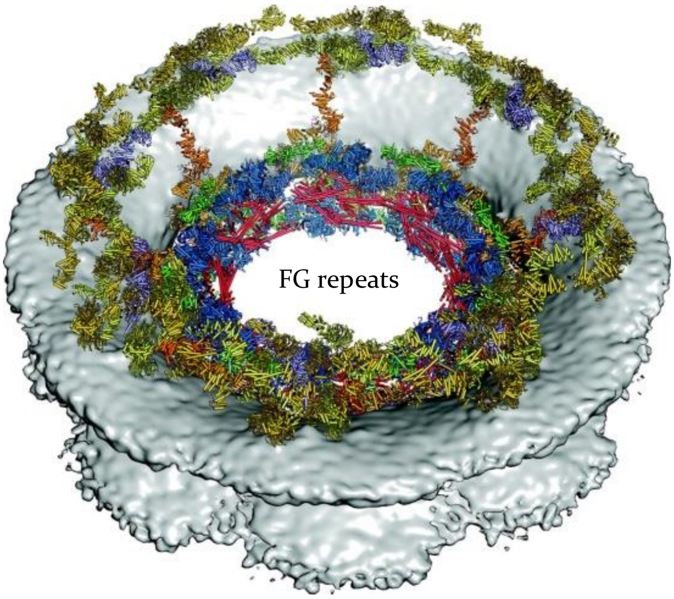


Figure 1.14. Composite structure of the NPC symmetric core (Lin *et al.*, 2016).

It is a cylindrical channel composed of multiple copies of more than 30 different nuclear pore proteins (nucleoporins) divided into three main categories. Most nucleoporins arrange in an eight fold symmetrical core structure perpendicular to the nuclear envelope, while others extend asymmetrically on the nuclear and cytoplasmic faces (forming the nuclear basket and the cytoplasmic filaments respectively). Many nucleoporins contain disordered repetitive sequences called FG repeats, due to their high content in phenylalanine and glycine residues. These FG repeats wrap the central channel to form a selective barrier. While ions, metabolites, and macromolecules below 40 KDa can pass freely through the NPC by passive diffusion; larger macromolecules, and RNA molecules in the form of RNP complexes requires ,tightly regulated active, transport mediated by transport factors.

After the escorting/docking of mature mRNPs to the nuclear basket of the NPC, Mex67 then interacts physically with the FG repeats of the central channel nucleoporins through its C-terminal UBA domain and NTF2 central domain, pulling the mature mRNPs further into the NPC and out to the cytoplasm.

1.6.1. Role in mRNA quality control

Practically every step of mRNA biogenesis involves rigorous quality control to detect possible errors in transcription, mRNA processing or export. In addition to the role of the nuclear exosome/TRAMP in the quality control of mRNA mentioned earlier, anchored to the NPC on the nuclear basket of the NPC, are the filamentous myosin-like proteins (Mlp1/Mlp2 in yeast/ Tpr in vertebrates). Mlps act as “gate keepers” or the final mRNA quality control check points, making sure that only mature, correctly spliced mRNPs are exported from the nucleus. Thus, preventing the escape of any aberrant, non-spliced, or misfolded mRNP missed by the nuclear surveillance machinery and therefore, blocking the translation of improperly synthesized mRNAs. This is achieved by their ability to associate with RNP complexes on their journey exporting the nucleus through their C-terminal globular domain (Green, Johnson, Hagan, & Corbett, 2003). It is still unclear how exactly Mlps recognize malformed mRNPs, but it was shown that Mlp1 interact with several proteins of the export machinery such the Poly (A) binding protein Nab2, Npl3 (Fasken, Stewart, & Corbett, 2008), and Sac3 of TREX-2 (Oeffinger *et al.*, 2007) (Figure 1.15.).

1.6.2. mRNA export termination and cytoplasmic mRNPs remodelling

Once the mRNP make it through the NPC and to the cytoplasmic side of the NPC, it requires a termination step to release the mRNA from its export factors in order to be further translated by the ribosome. Dead-box family enzymes have been known to play role in many aspects of RNA metabolism, but importantly for their function as ATP-dependent helicases in RNA-duplexes unwinding and/or the removal of RNA associated proteins (Fairman *et al.*, 2004; Jankowsky & Bowers, 2006). Dbp5, a highly conserved DEAD-box protein helicase (DDX19 in humans) that shuttles between the nucleus and cytoplasm plays a major role in this compositional re-arrangement step. Dbp5 associates with nascent transcripts in the nucleus and plays an essential role in mRNA export (Snay-Hodge *et al.*, 1998), but it´s ATPase activity in mRNP remodelling was suggested to be only activated in the cytoplasm, where the conformational changes leads to the dissociation of the mRNP bound export receptor Mex67 and other export factors such as Nab2 (Lund & Guthrie, 2005; Tran *et al.*, 2007). This is due to the fact that Dbp5 requires the function of another NPC-associated cytoplasmic protein called Gle1 and its co-factor InsP6 (inositol hexaisophosphate) to activate it (Weirich *et al.*, 2006).

The current model (figure 1.15) proposes that, at a steady-state Dbp5 localizes to the NPC through a direct interaction with Nup159 (Nup214 in humans) a conserved nucleoporin attached to the cytoplasmic face of the NPC. However, another cytoplasmic nucleoporin Nup42, adjacent to Nup159, interacts with Gle1-InsP6, which stimulates ATP binding and thus activation of Dbp5. The resulting Dbp5-ATP/Gle1-InsP6 complex is crucial as it changes Dbp5 conformation that allows its binding to mRNA (Montpetit *et al.*, 2011). The consequent ATP-hydrolysis then triggers a conformational change that results in the release of the mRNA and bound export factors, and hence mRNP remodelling (Montpetit *et al.*, 2011). However, other studies suggested that the ADP bound form Dbp5 and not the ATP-bound Dbp5 is crucial to remove export factors such as Nab2 from mRNA (Tran *et al.*, 2007). Furthermore, the resulting Dbp5-ADP binds Nup159 promoting the release of its bound ADP and thus aiding Dbp5 recycle, whereas export factors are re-imported into the nucleus (Noble *et al.*, 2011).

It is now known that this ATP-dependent remodelling mechanism of Dbp5 is responsible for providing directionality to the exported mRNPs by capturing the Mex67 bound mRNPs at the cytoplasmic face of nuclear pore, where its instant mRNP remodelling function generates an irreversible release in the cytoplasm, thus preventing its regression to the NPC central channel (Lund & Guthrie, 2005).

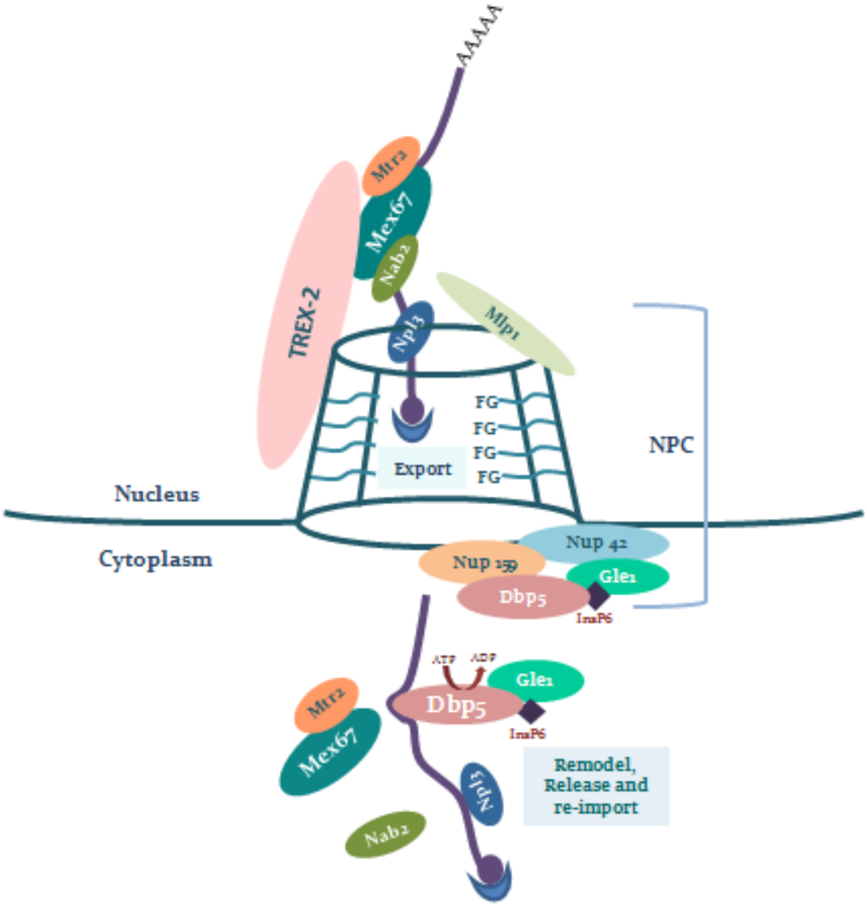

AAAAA
Mtr2
Mex67
Nab2
TREX-2
Npl3
Mlp1
FG
FG
FG
FG
Export
NPC
Nucleus
Cytoplasm
Nup 42
Nup 159
Gle1
Dbp5
InsP6
ATP
ADP
Gle1
Dbp5
InsP6
Mtr2
Mex67
Npl3
Remodel,
Release and
re-import
Nab2

Figure 1.15. Exported mRNP capturing and remodeling by Dbp5 at the cytoplasmic face of the NPC.

A recent study, showed a novel role of Dbp5 in rRNA export, proposing a distinct export/cytoplasmic remodelling mechanism for the pre-ribosomal subunit. Besides demonstrating that Dbp5 is also required for the nuclear export of pre-ribosomal subunits (pre-40S; pre-60S), it suggested that unlike its function in mRNP remodelling, Dbp5 ATPase cycle stimulated by Gle1 is dispensable in this case. Additionally, ATPase deficient dbp5 mutants that selectively inhibit mRNA export do not affect ribosomal transport. That is to say, whereas ATPase dependent activity of Dbp5 is required to dissociate bound export receptor Mex67 from the emerging mRNA, it does not displace Mex67 from the pre-ribosomal RNA. Instead, Mex67 remains bound to ribosomal subunits upon transit to the cytoplasm where it is detectable on translating ribosomes (Neumann, Wu, Hackmann, & Krebber, 2016).

This suggest that different RNPs may engage in different transport mechanisms both involving Dbp5, where its direct contact with Mex67 might aid in capturing the ribosomal subunits upon their cytoplasmic appearance at the NPC preventing any backsliding (Neumann *et al.*, 2016). When and how the Mex67 is released from the ribosomal subunits, and whether additional adaptor proteins are involved in this process is currently unknown. And why Dbp5 acts differently with different RNP cargoes is yet to be investigated.

MEX67

Mex67, a 67.35 KDa protein, was first identified in yeast (*S. cerevisiae*) through its genetic interaction with nucleoporin Nup85, where a synthetic lethal screen with a mutant allele nup85Δ, that exhibits a poly (A) $^+$ RNA export defect, led to the identification of Mex67 as a novel and essential mRNA export factor (Segref *et al.*, 1997). Regardless of the poor sequence homology to the yeast Mex67, TAP/NFX1 was characterized as the human homolog of Mex67, as an evolutionary conserved nuclear export factor, both in structure and function, that shuttles between the nucleus and the cytoplasm, capable of interacting with nucleoporins, and binding to poly (A) $^+$ RNA both *in vivo* and *in vitro* (J. Katahira *et al.*, 1999; Segref *et al.*, 1997).

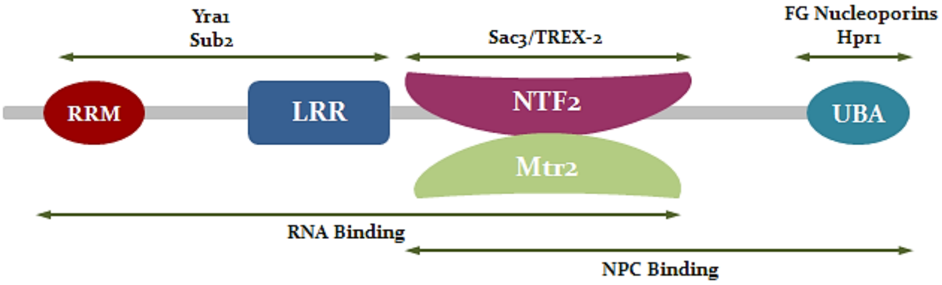


Figure 1.16. Mex67 domain distribution and the binding sites organization. Mex67 contains four major ordered domains. The N terminal region containing the RNA-recognition motif (RRM) and the leucine rich repeat (LRR) domain is the expected site of binding of adaptor proteins Yra1 and Sub2. Whereas the nuclear transport factor 2-like (NTF2) domain is the site of interaction with Mtr2 forming the Mex67/Mtr2 heterodimer and also the binding site of Sac3 from the TREX-2 complex. Additionally, the N-terminal region is involved in RNA binding, while the C-terminal ubiquitin associated (UBA) domain and the NTF2 domain are directly involved in binding FG nucleoporins of the nuclear pore complex. The C-terminal UBA domain is also the binding site of Hpr1 protein of the THO complex.

Mex67 belongs to the nuclear export factors (NFX) family, and is the major export receptor for mRNAs in the form of mRNPs, and also have a role in rRNA export. It also plays an essential role in mRNP release and directionality of transport (Smith *et al.*, 2015). It contains four modular domains: an N terminal RNA-recognition motif (RRM), a leucine rich repeat (LRR), a nuclear transport factor 2-like (NTF2) domain, and a C-terminal ubiquitin associated (UBA) domain (figure 1.16.).

Through its NTF2 domain, Mex67 forms the functional heterodimer with Mtr2 (p15 in humans) that is crucial for the nuclear mRNA export (Santos-Rosa *et al.*, 1998) (figure 1.17). Mutants where Mex67 and Mtr2 can no longer interact show inhibition of nuclear mRNA export and cell growth (Santos-Rosa *et al.*, 1998). Additionally, over expression of the human TAP/p15 can partially rescue the lethal *Mex67/Mtr2* knockout strain in *S. cerevisiae* (J. Katahira *et al.*, 1999).

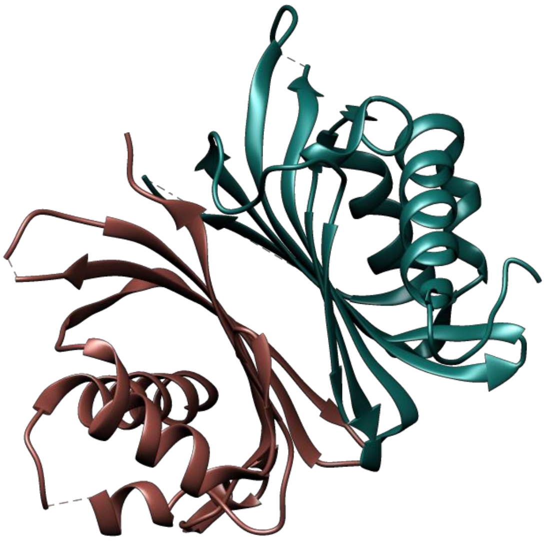


Figure 1.17. Crystal structure of the Mex67 NTF2 domain (Green)/Mtr2 (pink) heterodimer at 2.8 A° resolution. PDB ID: 1OF5.

Both the NTF2 domain, as well as the UBA domain of Mex67 contain binding sites for the FG repeats of the NPC nucleoporins in the context of strengthening the interaction (Fribourg, Braun, Izaurralde, & Conti, 2001; Sträßer *et al.*, 2000) thus both contributing to the NPC association/translocation process. Moreover, Mex67/Mtr2 also interacts with Sac3 of the TREX-2 complex through the NTF2 domain of Mex67 (Dimitrova *et al.*, 2015). The flexible RRM domain and LRR domain of Mex67 are known to have the ability to bind RNA on their own, but in yeast the structural orientation of the LRR domain relative to the NTF2/Mtr2 regions creates a large positively charged surface that have been seen to extend the surface of RNA binding and to augment the contribution made to RNA binding by the RRM and LRR domains (S. Aibara *et al.*, 2015; Teplova, Wohlbold, Khin, Izaurralde, & Patel, 2011). The N-terminal domain of Mex67 is also involved in the binding with Yra1.

The C-terminal UBA domain of Mex67, however, have been proven necessary and sufficient for cargoes export through the nuclear pore since it is a direct site of interaction with the all types of FG repeats (FG, GLFG or FXFG core sequences where X is any residue) (Bachi *et al.*, 2000; Grant, Neuhaus, & Stewart, 2003; Hobeika *et al.*, 2009; Sträßer *et al.*, 2000). Interestingly, it is also the binding site of Hpr1 of the THO.TREX complex, where it synchronizes the recruitment of the export machinery with transcription, and plays a role in Hpr1 ubiquitination dependent turnover (Gwizdek *et al.*, 2006). Structural studies have showed that both Hpr1 and the FG repeats compete for the same binding site in the UBA domain (Hobeika *et al.*, 2007, 2009) which suggests they don't bind simultaneously especially that both Hpr1 and the FG repeat show similar binding affinity for the UBA domain. However, the binding of Hpr1 and not the FG repeats appears to promote UBA interaction to tetra-ubiquitin (Hobeika *et al.*, 2007, 2009).

Mex67 is usually recruited primarily to mature, spliced mRNPs, but it was also shown to be implicated in the export of un-spliced or partially spliced forms of viral RNA (Teplova *et al.*, 2011). Recently, Mex67/Mtr2 heterodimer have been also shown in *S. Cerevisiae* to bind and export the pre-60S ribosomal subunit in vivo as well as binding 5S rRNA in vitro, where loop insertions in the NTF2 domain of Mex67 and Mtr2 are crucial structural requirements for this function. It was suggested that an electrostatic interaction could occur between the positively charged Mex67 loop and an exposed negatively charged rRNA surface on the pre-60S subunit (Yao *et al.*, 2007). Importantly, the predicted site on NTF2 for FG repeats interaction is located on the opposite face of the loop-confined surface. Remarkably, the same loops were later shown to recruit and bind Mex67/Mtr2 to the pre-40S subunits (Marius B. Faza *et al.*, 2012). But surprisingly, these loop insertions are missing from the NTF2 region of the TAP/p15 human homolog. Therefore, whether the human Mex67 is also involved in the export of pre ribosomal subunits is still to be investigated. However, in higher eukaryotes, several Mex67 isoforms of the NXF family members might have other specialized transport functions and perhaps compensate this pre ribosomal subunits export function (Herold *et al.*, 2000).

Another important role has been very recently uncovered for Mex67 in yeast. The new study suggested that under cellular heat stress, and unlike their role in mRNA export under normal conditions, Mex67 adaptor proteins dissociate from regular mRNAs thus dissociating Mex67 from its regular cargo. Instead, it was suggested that the now freely available Mex67 selectively binds and rapidly exports heat shock transcripts, without the need of adaptor proteins, bypassing the quality control system. It was also proposed that in this case, Mex67 interacts directly with the RNA Pol II largest subunit (Rbp1) thus enabling rapid export of heat shock mRNAs, speeding up translation and guarantying cell survival under extreme cases (Zander *et al.*, 2016).

MIP6

The first mention of the yeast protein Mip6 was in the same paper where Mex67 was first characterized. It was identified through a two-hybrid interaction screen against a yeast genome library using Mex67 as bait in an attempt to find any possible Mex67 partners (Segref *et al.*, 1997). It was described as an uncharacterized open reading frame (ORF) in yeast, and was named Mex67 interacting protein 6 (Mip6). Additionally, it was suggested that an intact Mex67 carboxy-terminal domain which includes part of the P/Q/G domain, as stated, is required for the two-hybrid interaction with Mip6 (Segref *et al.*, 1997).

Although Mex67 have been extensively studied afterwards, Mip6 structure, function, and mechanism of action remain unknown. Software predictions predicted Mip6 as a putative RNA-binding protein with predicted three putative RNA recognition motifs (RRMs), but whether Mip6 is really capable of binding RNA or if together with Mex67 is involved in the mRNA export mechanism was never investigated.

On the other hand, it was suggested that Mip6 have high homology for Pes4, another yeast putative RNA binding protein which was found as a suppressor of DNA-polymerase ε subunit. Little else is known about Pes4 structure or function (Segref *et al.*, 1997).

However though, a recent study on protein concentration dependent cellular toxicity, suggested that over expression of the yeast dosage sensitive Mip6 protein causes the protein to localize, reversibly, to cytoplasmic RNP granules causing a reduction in the global translation rates and thus triggering cellular toxicity and growth impairment when strongly over expressed (Bolognesi *et al.*, 2016).

2. Objectives

While Mex67 have been well studied before, more is yet to be learned regarding its structure and the dynamics of its interactions and function. Additionally, considering the insufficient literature concerning Mip6 or its Pes4 paralog structurally and functionally and the very few evidence in which the interaction between Mip6 and Mex67 and the subsequent function and implications of this complex formation is based on, our main objectives initially set were as follows:

1. *In vitro* expression and purification of Mip6, Mex67, and Pes4 constructs.
2. Crystallizing and solving the structure of Mip6 and Pes4.
3. Elucidating and confirming the RNA binding properties of Mip6.
4. Studying and characterizing the interaction between Mip6 and Mex67.
5. Defining the minimum interaction boundaries required by the Mip6/Mex67 interaction and studying its properties.
6. Checking whether Pes4 retained the ability to bind Mex67.
7. Reconstituting the Mip6/Mex67 complex *in vitro* and using it for crystallization trials in an attempt to solve the complex structure.
8. Studying the role of nucleic acid binding in the complex formation between Mip6 and Mex67.
9. Speculating on possible *in vivo* functions for Mip6 and the Mip6/Mex67 complex regarding mRNA export.

3. Materials and Methods

3.1. Cloning

Plasmids used in this study were either commercially available or were obtained from lab members. Mip6 FL pMAL-c2X construct was received from Dr. Susana Rodriguez Navarro lab (Centro de Investigación Príncipe Felipe (CIPF); Valencia, Spain), Pes4 initial constructs were a gift from Dr. Jose Manuel Pérez Cañadillas lab (Institute of physical chemistry Rocasolano, CSIC, Madrid, Spain), and pMBPKI and pGKI vector were a gift from Dr. Marcin Węgrecki engineered by him in our laboratory from pET28-NKI/LIC 6His/3C vector originally obtained from Dr. Anastassis Perrakis group (NKI, Amesterdam).

The constructs used for this project were modified from previously obtained constructs and cloned into different available plasmids in the lab. The gene sequences coding for all of proteins of interest used in this study: Mip6, Mex67, and Pes4 were that of *S. Cerevisiae* yeast species proteins. The genes coding for proteins of interest or for the deletions of the proteins were amplified by PCR (Polymerase Chain Reaction) using primers listed in table 3.1. The Primers overhangs varied depending on the cloning method and plasmids used.

Primer	Sequence
Mip6 NKI Fwd	5´cagggacccggtatgccaaactctcATGGTAATGTATTGAATAAT 3’
Mip6 NKI Rev	5´cgaggagaagcccggTTAAACTGTAAATCCGAGGCTTTCAGCA 3´
Mip6 pOPINM Fwd	5´aagttctgtttcagggcccgATGCCAAACTCTCATGGT 3´
Mip6 pOPINM Rev	5´atggtctagaaagctTTAAACTGTAAATCCGAGGC 3´
Mip6 390 pOPINM Fwd	5´aagttctgtttcagggcccgatg AAGCCTGTTCATAATCAA 3´
Mip6 111 BamHI Fwd	5´gacacggatccatgAATAGTTTGTTTATAG 3´
Mip6 279 Xho Rev	5´ccggcctcgagtcaAGTTCTTACTTCTTT 3´
Mip6 313 BamHI Fwd	5´gacacggatccatgAAAACAATATTAGTCA 3´
Mip6 389 Xho Rev	5´ccggcctcgagtcaATCTTTACCAGGACCA 3´
Mip6 480 Xho Rev	5´ccggcctcgagtcaATTCTTTGACCATGATGCCT 3´
Mip6 401 BamHI Fwd	5´gtacctggatccatg AAAACAAAGGTTTAT 3´
Mip6 390 BamHl Fwd	5´gtacctggatccatgAAGCCTGTTCATAAT3´
Mip6 390 NKI Fwd	5´cagggacccggtAAGCCTGTTCATAATCAAATTGGAACG 3´
Mip6 495 NKI Rev	5´cgaggagaagcccggttaGTTGTTACCATCGTCATAGT 3´
Mip6 480 NKI Rev	5´cgaggagaagcccggttaATTCTTTGACCATGATGCCT 3´
Mex67 528 BamHI Fwd	5´gtacctggatccatggtcATGGCTCCCACT 3´
Mex67 NotI Rev	5´gacgcttgcggccgcTTAGAACTGCACAAATGC 3´
Mex67 545 BamHI Fwd	5´gtacctggatccatgAATCCCGTACAACTCGAG 3´

Mex67 544 NotI Rev	5´gacgcttgcggccgcttaTAGCCGCGACTGGACATC 3´
Mex67 481 BamHI Fwd	5´gtacctggatccatgGCAATAGCACAGCCTCCA 3´
Pes4 RRM3/4 NKI Fwd	5´cagggacccggtatgAATTCTATATTCATCAAAAACCTTCCA 3´
Pes4 RRM3/4 NKI Rev	5´cgaggagaagcccggttaATTATTTTGTCTTTCCCAAGATGT 3´

Table 3.1. Primers used for cloning the constructs used in this work. The sequences that include the overhangs and the extra nucleotides not part of the gene are shown in lower case letters.

The PCR reaction mix had a final volume of 25 µl and was prepared as shown in the following table 3.2:

Component	25 µl reaction	Final concentration
5x Kapa Hifi buffer	5 µl	1x
10 mM Kapa dNTPs mix	0.75 µl	0.3 mM each
10 µM forward primer	0.75 µl	0.3 µM
10 µM reverse primer	0.75 µl	0.3 µM
Pure DMSO	1.25 µl	
Template DNA	1 µl or as required (10-100 ng)	As required
1 U/µL KAPA HiFi (Hifidelity) DNA Polymerase (Kapa Biosystems)	0.5 µl	0.5 U
MQ (Milli-Q) water	15 µl (up to 25 µl)	N/A

Table 3.2. PCR reaction mixture preparation for Kapa Hifi Hot start Polymerase.

The PCR reaction was then performed with the following cycling protocol setup:

setup	Temperature	Duration	Cycles of Amplification
Initial denaturation	95°C	3 minutes	1
Denaturation	98°C	20 seconds	30
Annealing	60-70°C depending on desired Tm (melting temperature)	15-20 seconds	
Extension	72 °C	30 seconds/kb	
Final extension	72 °C	5 minutes (1 minute/kb)	1

Table 3.3. PCR reaction amplification conditions setup.

PCR amplified products were then visualized on 1-2% agarose gels (depending on the size of the expected DNA band) stained with 1x GelRed (Biotium) and run at 100V for the required duration. The molecular weight marker used to assess the visualized band size was DNA ladder 1kb (Nippon Genetics). The PCR products of interest were then purified using commercially available kits (SPEEDTOOLS PCR CLEAN UP kit from BIOTOOLS; E.Z.N.A Cycle Pure Kit and MicroElute Gel Extraction Kit from OMEGA bio-tek). DNA concentration obtained was afterwards quantified with NanoDrop (Thermo Scientific). The resulting DNA fragments were then incorporated into the corresponding plasmids depending on the required method as follows:

- **Cloning using Restriction enzymes:** complementary compatible ends were generated by digesting both the insert and vector with the same desired restriction enzymes purchased from New England Biolabs (NEB). The reaction was prepared as such:

	Insert PCR product	**Plasmid**
Sample	30 μl	15 μl
Restriction enzyme (NEB)	2 μl	1.5 μl
10x Buffer (NEB)	4 μl	2 μl
100x BSA (10mg/ml)	0.5 μl	0.5 μl
MQ water	Up to 40 μl	Up to 25 μl
Incubation temperature	37 °C	37 °C
Incubation time	2 hours	2 hours

The NEBuffer used depended on the restriction enzymes used and the quantity of enzyme was also variable depending of the amount of DNA used. Additionally, 1 μl of SAP (Shrimp Alkaline Phosphatase) purchased from Fermentas was added to the plasmid sample after 1 hour 30’ of incubation at 37 °C before being further incubated for an additional half an hour. The insert and vector samples were then visualized on 1% agarose gel and the digested insert and vector samples were purified from directly sample or from gel. The ligation was then carried out using the Quick ligation kit (NEB) where 50 ng of digested vector were mixed with 3-fold molar excess of insert and the final volume was adjusted to 10 μl with MQ water. 10 μl of Quick ligation reaction buffer was then added to the mixture with a final 1 μl of Quick T4 DNA ligase and incubated at room temperature for 5 minutes before being chilled on ice and transformed into *E. coli* (*Escherichia coli*) DH5α bacterial cells.

- **Ligase Independent Cloning (LIC):** The vectors were linearized by KpnI restriction enzyme (NEB) and further purified using a commercial PCR clean kit, then both the vectors and insert were treated separately with T4 DNA polymerase (Thermo scientific) in a reaction mix of a final volume of 20 µl as following: 15.3 µl of DNA (~ 0.1 pmol of vector or insert) was mixed with 4 µl of 5x T4 DNA polymerase buffer (Thermo Scientific), 0.5 µl of 100 mM Deoxyadenosine triphosphate (dATP) (for insert) or 0.5 µl of 100 mM Deoxythymidine triphosphate (dTTP) (for vector), and 0.2 µl of T4 DNA polymerase (5U/µl). The reaction was incubated for 5-15 minutes at room temperature and then for 10 minutes at 70 °C to inactivate the enzyme. Vector and insert were then mixed in equal volumes (2 µl each) and left to incubate 5 minutes at room temperature. Then 1 µl of EDTA (50 mM) was added and the mixture was further incubated for additional 5 minutes at room before being transformed into *E. coli* DH5α cells.

- **In-Fusion Cloning:** was performed using the commercial Clontech In-Fusion kit. The vector pOPINM was linearized through KpnI and HindIII enzymes according to instructions provided by Oxford Protein Production Facility (OPPF, UK) that produced the pOPIN vectors suite. The linearized vector was then mixed with purified PCR fragment of insert (50-100 ng each) and 1 µl of 5x In-Fusion HD Enzyme Premix. MQ water was added up to 5 µl of total volume. The reaction was incubated for 15 minutes at 50 °C, and then placed on ice to stop the reaction. 10 µl of TE (Tris-EDTA) buffer were immediately added to dilute the reaction before being transformed into *E. coli* DH5α cells.

The constructs used in this study are summarized in Table 3.4 with the corresponding cloning method used.

Construct	Primers	vector	Cloning method
Mip6 FL pNKI	Mip6 NKI Fwd Mip6 NKI Rev	pET28-NKI/LIC 6His/3C	LIC
Mip6 FL pMBPKI	Mip6 NKI Fwd Mip6 NKI Rev	pMBPKI	LIC
Mip6 FL pGKI	Mip6 NKI Fwd Mip6 NKI Rev	pGKI	LIC
Mip6 FL pOPINM	Mip6 pOPINM Fwd Mip6 pOPINM Rev	pOPINM	In-Fusion
Mip6 111-480 pGEX-6P-2	Mip6 111 BamHI Fwd Mip6 480 Xho Rev	pGEX-6P-2	Restriction enzymes
Mip6 RRM1/2 pGEX-6P-2	Mip6 111 BamHI Fwd Mip6 279 Xho Rev	pGEX-6P-2	Restriction enzymes
Mip6 RRM3/4 pGEX-6P-2	Mip6 390 BamHl Fwd Mip6 480 Xho Rev	pGEX-6P-2	Restriction enzymes
Mip6 390 pOPINM	Mip6 390 pOPINM Fwd Mip6 pOPINM Rev	pOPINM	In-Fusion
Mip6 RRM3 pGEX-6P-2	Mip6 313 BamHI Fwd Mip6 389 Xho Rev	pGEX-6P-2	Restriction enzymes
Mip6 390-495 pNKI	Mip6 390 NKI Fwd Mip6 495 NKI Rev	pET28-NKI/LIC 6His/3C	LIC
Mip6 390-480 pNKI	Mip6 390 NKI Fwd Mip6 480 NKI Rev	pET28-NKI/LIC 6His/3C	LIC
Mip6 401-480 pGEX-6P-2	Mip6 401 BamHI Fwd Mip6 480 Xho Rev	pGEX-6P-2	Restriction enzymes
Mex67 528 pGEX-6P-2	Mex67 528 BamHI Fwd Mex67 NotI Rev	pGEX-6P-2	Restriction enzymes
Mex67 UBA pGEX-6P-2	Mex67 545 BamHI Fwd Mex67 NotI Rev	pGEX-6P-2	Restriction enzymes
Mex67 ΔUBA pGEX-6P-2	Mex67 481 BamHI Fwd Mex67 544 NotI Rev	pGEX-6P-2	Restriction enzymes
Pes4 RRM3/4 pNKI	Pes4 RRM3/4 NKI Fwd Pes4 RRM3/4 NKI Rev	pET28-NKI/LIC 6His/3C	LIC

Table 3.4. Constructs used in this work. Primers are shown in table 3.1.

After successful incorporation methods, the mix was transformed into bacterial DH5α *E. coli* strain cells using heat shock. 100 µl of chemically competent cells were thawed on ice for 5 minutes. The ligation mix was then introduced and further incubated on ice for 20-30 minutes. The cells were then heat shocked by placing them at 42 °C for 1.5 minutes and then placed on ice for 5 minutes. 500 µl of fresh LB media was then added to the cells and cells were incubated in a shaker at 37°C for about 1 hour. After 1 hour, the cells were plated on

LB-agar containing required antibiotics depending on the resistance of each plasmid, and then incubated at 37°C for around 18 hours. Colonies from subsequent plates were used to perform colony PCR to scan for positive colonies according to the following conditions:

setup	Temperature	Duration	Cycles of Amplification
Initial denaturation	95°C	3-5 minutes	1
Denaturation	95°C	30 seconds	30-35
Annealing	50-70°C depending on desired Tm (melting temperature)	30 seconds	
Extension	72 °C	1 min/kb	
Final extension	72 °C	5-10 minutes	1

Positive colonies of choice were then inoculated in 5ml of LB culture media supplemented with the appropriate antibiotic (Final concentration of 33 µg/ml for Kanamycin or 100 µg/ml for Ampicillin) depending on the plasmid antibiotic resistance and incubated at 37°C over night. The cells were then collected by centrifugation (4000 rpm for 15-20 minutes) and then the plasmid was extracted using QuickGene Plasmid Kit S II (Fujifilm) according to the instruction manual. The DNA concentration of the resulting plasmids was measured using NanoDrop (Thermo Scientific) and then sent to the sequencing service at the Institute of Biomedicine of Valencia for sequence analysis to assure all constructs were well done.

3.2. Site directed mutagenesis

The mutation of interest in the sequence of the gene was generated following the QuickChange site-directed mutagenesis protocol (Kapa Biosystems) using forward and reverse primers where the mutation is introduced in the central part of the primer. These primers are designed to amplify the whole plasmid, generating amplified plasmid copies that contain the mutation. Kapa Hifi Hot start polymerase was used to set the mutagenesis as the following:

setup	Temperature	Duration	Cycles of Amplification
Initial denaturation	95°C	2 minutes	1
Denaturation	98°C	20-30 seconds	16 cycles
Annealing	60 °C depending on desired Tm (melting temperature)	15 seconds	
Extension	72 °C	30 sec/kb	
Final extension	72 °C	1-5 minutes	1

The obtained PCR product was then visualized on a 1% agarose gel. Further incubation of the PCR product with 1.5 µl of DpnI enzyme (NEB; 20,000 U/ml) was carried on for 1-2 hours at 37°C to digest the template plasmid not containing the mutation. The mix was then chilled on ice to stop the reaction and transformed by heat shock into DH5α *E. coli* bacterial cells as previously described and some colonies inoculated in LB with the correct antibiotics and left at 37 °C overnight. The subsequent extracted plasmid was sequenced to confirm a successful mutagenesis.

3.3. Protein over-expression in *E. coli*

Proteins were transformed into the Bl21 (DE3) or Bl21 codon plus (DE3)-RIPL chemically competent bacterial *E. coli* cells (Stratagene) for expression. 1 µl of DNA (5-100 ng/µl) was added to 100 µl of the chemically competent cells and incubated 20-30 minutes on ice. The cells were then heat shocked by placing them at 42 °C for 45 seconds and then back on ice for 5 minutes. 500 µl of fresh LB media was added to the cells and further incubated for 1 hour at 37 °C shaking. 100 µl of the cells were plated in LB-agar containing the appropriate antibiotics. The plates were placed in a 37 °C incubator overnight.

The expression of newly tested proteins was assessed by using a small-scale expression system. If soluble protein is obtained, then the protein was further expressed and purified on a large scale.

3.3.1. Small-scale expression

From the LB-agar plate of the E. coli (DE3) BL21 codon plus RIPL cells, two colonies were picked to be checked for expression: one in LB medium using IPTG expression, and the second in ZY medium using the Auto-induction system.

- **IPTG expression test:** one colony was inoculated in 10 ml of LB media with the appropriate antibiotics depending on the plasmid resistance and the type of cells used to express the protein (Chloramphenicol at a final concentration of 33 µg/ml added in case of Bl21 (DE3) codon plus). The culture was left growing at 37 °C for 5-6 hours until OD (optical density) of 0.6-0.7 and then the expression of protein was induced by adding 0.5-0.8 mM of IPTG (Isopropyl β-D-1-thiogalactopyranoside). The culture was then left at 20 °C overnight.
- **Auto-induction:** A culture of 9.3 ml of ZY media, 10 µl $MgSO_4$ (1M), 200 µl 5052 (50x), and 500 µl NPS (20x) plus required antibiotics were inoculated with the other colony and were incubated at 37 °C shaking until an OD of 1. The culture was then changed to 20 °C and left there overnight. In cases of Kanamycin is required as antibiotic, double the normally used quantity at 33 µg/ml is required.
  Recipes: NPS 20X (0.5 M $(NH_4)_2SO_4$, 1M KH_2PO_4, 1M Na_2HPO_4; 5052 50x (25% glycerol, 2.5% glucose, 10% α-lactose); ZY (1% Tryptone, 0.5% yeast extract).

The next day, the cultures from both expression systems were centrifuged and the cells were collected at 4000 rpm speed at 4 °C for 20 minutes. The supernatants were then discarded and the cells re-suspended with 1 ml PBS 1x (Phosphate-Buffered Saline) buffer and transferred to 1.5 ml Eppendorf tubes, then further centrifuged, supernatants discarded and pelleted cells preserved for further small scale purification to assess the solubility of the protein. Additionally, depending on the solubility, different temperatures were also used for expressing the proteins.

3.3.2. Large-scale over-expression

The proteins that showed good solubility after the small-scale expression followed by small-scale purification were then expressed on a large scale for large-scale protein purifications. The expression system and media used for the large scale expression depended on which small scale expression system showed more protein solubility, and accordingly some proteins were expressed using the IPTG expression system while others using Auto-induction.

- **IPTG expression:** 50 ml of a pre-culture of LB media with antibiotics was inoculated with a colony from the agar plate following bacterial cells transformation into a strain suitable for expression, and left to grow at 37°C overnight. The next morning, the overnight pre-culture, was added into a 950 ml of fresh LB with antibiotics and left shaking at 37 °C until an OD of 0.6-0.8 was reached. The culture was then put at 20°C (or lower depending on the protein) and induced with 0.5-0.8 mM of IPTG and left expressing overnight (Mip6 RRM4 W442A mutant was expressed at 12 °C for 72 hours before cells were collected). After expression, the cells were collected and centrifuged at 3500 rpm for 45-55 minutes at 4°C. The supernatants were then discarded and pellets washed with PBS and further centrifuged at 4000 rpm for 30 minutes at 4 °C, supernatants discarded and pelleted cells frozen at -80 °C for future use.
- **Auto-induction:** An isolated colony was inoculated in a 10 ml pre-culture of 9.3 ml ZY media, 10 µl $MgSO_4$ (1 M), 200 µl Glucose (40%), 500 µl NPS (20x) plus antibiotics and left growing overnight at 37 °C. The

next morning, 1 litre of fresh Auto-induction culture was prepared: 930 ml of ZY media, 1 ml $MgSO_4$ (1 M), 20 ml 5052 (50x), 50 ml NPS (20x) with antibiotics and separated into two, 2 Litre, flask containing 500 ml litre each. 5 ml of the overnight pre-culture was then inoculated in each flask and left at 37°C to grow until an OD of 0.8-1. The cultures were then moved to a lower temperature (20 °C or less depending on the protein) and left there for at least 18 hours depending on the temperature used for expression. For proteins with no problems in solubility, the expression was carried out at 18-20 °C for 18-23 hours. As for Mip6 FL for example, the protein was expressed at 16 °C for 48 hours due to low solubility of the protein. The cells were centrifuged at 3500 rpm for 45-55 minutes at 4 °C, then supernatants discarded and pellets re-suspended in PBS 1x and further centrifuged at 4000 rpm for 30 minutes, and pellets were frozen at 80 °C for future purification.

3.3.3. Protein expression in insect cells-baculovirus system

Mip6 FL cloned in pOPINM (N terminal 6xHis tag followed by MBP tag), produced by In-fusion cloning following the OPPF (Oxford Protein Production Facility, UK) insect expression protocol, was used for Sf9 cells transfection. 2 ml of Sf9 (*Spodoptera frugiperda*) cells were disposed on a six-well plate (1x106 of cells/well) and the plate was left at 27 °C for 1 h. The transfection mix was then prepared containing 200 µl of Sf900II media (Life Technologies), 1 µl of Ian Jones bacmid (750 ng) (Zhao, Chapman, & Jones, 2003), 1 µl of Mip6 FL pOPINM plasmid (500-1500 ng) and 4 µl of FugeneHD (promega) for each well and then the mix left for 30 minutes at room temperature. After that, the mix was added to each well containing Sf9 cells and the plates were left at 27 °C. After 5-7 days after transfection, the supernatants containing Po virus were collected and the cells collected by centrifugation (1000 x g for 15 minutes at 4 °C) were then lysed and lysates centrifuged at 4 °C for 15 minutes. The soluble fraction in the supernatants and the insoluble fraction were then analyzed on SDS-PAGE and western blot to detect soluble protein.

We also tried the Bac-to-Bac expression system (Life Technologies) with Mip6 FL protein cloned into pFastBacHTb vector and transformed into DH10Bac *E. coli* competent cells where obtained positive colonies were used to extract the recombinant bacmid. 5 µl of bacmid DNA (1-2 µg) with 4 µl fugeneHD transfection reagent (Promega) were transfected into Sf9 insect cells in Sf900II medium to a final volume of 2 ml and plates were left at for 1 week at 27 °C with daily checking. 500 µl of transfected cells were then inoculated in 30 ml of 1.5x106 Sf9 cells and further left for one week at 27 °C. A serial dilution of the virus was then done in a 96-well plate and the fluorescence of GFP (Green Fluorescent Protein) expressing cells was checked with confocal microscope after 5 days to assess the success of the transfection. High five insect cells at a density of 2x106 in Express five medium (Gibco) were then infected with the recombinant virus and incubated at 27 °C for 72 hours. The cells were then harvested by centrifugation for 15 minutes at 1100 rpm, washed with PBS 1x and centrifuged at 600 g for additional 30 minutes. The supernatants were discarded and the pelleted cells were stored at -80 °C for further use.

Table 3.5 bellow summarizes the expression system and purification method used for every protein used in this study.

Protein	**Residues**	**Tag (Cleavage site)**	**Host**	**Expression method**	**Affinity Chromatography Purification**
Mip6 FL	1-659	6xHis (3C)-Nterm GST (3C) Nterm MBP (3C) Nterm	Bl21; BL21 C$^{+;}$ sf9	IPTG induction; Auto-induction; Baculovirus	5ml HisTrap; 5ml GSTrap; 5ml MBP trap; 5ml amylose resin packed gravity column
Mip6$_{(111-480)}$	111-480	GST (3C) Nterm	BL21 C^{+}	Auto-induction	5ml GSTrap
Mip6 RRM1/2	111-279	GST (3C) Nterm	BL21 C^{+}	Auto-induction	5ml GSTrap

Mip6 RRM3/4	313-480	GST (3C) Nterm	BL21 C^{+}	Auto-induction	5ml GSTrap
Mip6 390 pOPINM	390-659	MBP (3C) Nterm	BL21 C^{+}	Auto-induction	5ml MBP trap
Mip6 RRM3	313-389	GST (3C) Nterm	BL21 C^{+}	Auto-induction	5ml GSTrap
Mip6 RRM4$_{(390-495)}$	390-495	6xHis (3C)-Nterm	BL21 C^{+}	Auto-induction	5ml HisTrap
Mip6 RRM4$_{(390-480)}$	390-480	6xHis (3C)-Nterm	BL21 C^{+}	Auto-induction	5ml HisTrap
Mip6 RRM4$_{(401-480)}$	401-480	GST (3C) Nterm	BL21 C^{+}	Auto-induction	5ml GSTrap
Mip6 W442A	390-482	TRX-His (TEV) Nterm	BL21	IPTG induction	5ml HisTrap
Mex67$_{(528-599)}$	528-599	GST (3C) Nterm	BL21 C^{+}	Auto-induction	5ml GSTrap
Mex67 UBA	545-599	GST (3C) Nterm	BL21 C^{+}	Auto-induction	5ml GSTrap
Mex67 ΔUBA	481-544	GST (3C) Nterm	BL21 C^{+}	Auto-induction	5ml GSTrap
Pes4 RRM3/4	303-473	6xHis (3C)-Nterm	BL21 C^{+}	Auto-induction	5ml HisTrap

Table 3.5. The proteins purified on a large scale for this study and the type of affinity chromatography used. All purifications were followed by size exclusion chromatography purification. The host column represents the host used to express the protein (Bl21: *E. coli* (DE3) strain; BL21 C+: *E. coli* (DE3) codon plus RIPL strain; Sf9: *Spodoptera frugiperda* cells).

3.4. Protein purifications

All buffers used for the purification are summarized in table 3.6.

3.4.1. Small-scale purification

The cells obtained from the small-scale expressions produced in *E. coli* (section 3.3.1) were purified on a small scale to assess the expression and solubility of the protein. Depending on the tag fused to each protein, different resins were used for the small-scale purification: High Density Nickel beads (Agarose Bead Technologies) for His tagged proteins, Glutathione Sepharose 4B beads (GE Healthcare) for GST tagged proteins, and Amylose Resin (NEB) for MBP tagged proteins. The purification protocol was as follows:

- Pelleted cells were re-suspended in 300 µl of lysis buffer.
- The samples were then sonicated in a Bioruptor UCD-200 (Diagenode) at medium intensity (200 W) with 30 seconds ON/OFF intervals for 30 minutes at 4 °C.
- After sonication, samples were centrifuged at 4 °C for 20 minutes at maximum speed (16000-21000 x g).
- The soluble fraction in the supernatants was then added to 70-100 µl of beads previously washed with water and equilibrated by 1 ml of Buffer A.
- The samples were then left mixing with the beads on a rotator for 30-60 minutes at 4°C.
- The beads were then precipitated by centrifugation for 5 minutes at 4000 rpm. The non-bound fraction in the supernatants was then discarded and beads were washed three to four times with 1 ml of fresh buffer A.
- Finally, 50 µl of elution buffers were added, or simply 50 µl of SDS-PAGE sample loading buffer (6x), before being boiled for 10 minutes.

Aliquots from all the steps of the purification in addition to the elution samples were then analyzed on 10% SDS-PAGE. A molecular weight Blue Star Pre-stained Protein Marker (Nippon) was used to assess the molecular weight corresponding to the visualized bands. The acrylamide gel was run at 200 V for the time needed, and stained with a Coomassie Blue solution (1g Coomassie Blue, 40% methanol, 10% acetic acid, 50% distilled water) for 5-10 minutes and then de-stained in a de-staining solution (10% Methanol, 10% acetic acid, 80% distilled water).

3.4.2. Large-scale protein purification

3.4.2.1. Cell lysis and protein extraction

The cells collected from the large-scale expressions were first left to thaw on ice, before being re-suspended in cold lysis buffer

(buffer composition depends on the tag of the protein, refer to table 3.6) on ice, supplemented with one pill of Complete EDTA-free Protease Inhibitor Cocktail (Roche). The cells were then lysed by sonication 15-30 minutes (1 second ON 1 second OFF pulses) at 25% amplitude using Bioblock scientific vibra cell W75042 sonicator. The duration of soniction changed depending on the volume of cells and type of cells to be lysed. The lysed cells were then centrifuged at 15 000 rpm for 45 minutes in Sorvall RC6, SS-34 rotor at 4 °C. The resulting supernatants were then filtered using a 0.45 µm syringe filter (VWR) and loaded into the desired affinity chromatography column for purification. Aliquots from both supernatants and pellets were taken to be visualized later on acrylamide gel.

Affinity Purification Protocol

Buffer	HisTrap FF	GSTrap HP
Lysis Buffer	50 mM Hepes pH 7.5 500 mM NaCl 10 mM Imidazole 5% Glycerol 0.1% Triton X-100 1 mM BME*	50 mM Hepes pH 7.5 500 mM NaCl 5% Glycerol 0.1% Triton X-100 1 mM BME
Buffer A	50 mM Hepes pH 7.5 500 mM NaCl 10 mM Imidazole 5% Glycerol 1 mM BME	50 mM Hepes pH 7.5 500 mM NaCl 5% Glycerol 1 mM BME
Elution Buffer	50 mM Hepes pH 7.5 500 mM NaCl 500 mM Imidazole 5% Glycerol 1 mM BME	50 mM Hepes pH 7.5 500 mM NaCl 20 mM reduced gluthatione 5% Glycerol 1 mM BME

Affinity Purification protocol

Buffer	MBP Trap HP	HiTrap Heparin
Lysis Buffer	50 mM Hepes pH 7.5 500 mM NaCl 5% Glycerol 0.1% Triton X-100 1 mM DTT 1 mM EDTA	N/A
Buffer A	50 mM Hepes pH 7.5 500 mM NaCl 5% Glycerol 1 mM DTT	50 mM Hepes pH 7.5 100 mM NaCl 5% Glycerol 1 mM BME

	1 mM EDTA	
Elution Buffer	50 mM Hepes pH 7.5 500 mM NaCl 20 mM Maltose 5% Glycerol 1 mM DTT 1 mM EDTA	50 mM Hepes pH 7.5 1 M NaCl 5% Glycerol 1 mM BME
SE (size-exclusion) Buffer	50 mM Hepes pH 7.5 150 mM NaCl 5% Glycerol 1 mM BME	50 mM Hepes pH 7.5 150 mM NaCl 5% Glycerol 1 mM BME

Table 3.6. Buffers used in protein purifications depending on the type of purification protocol used. *ß-Mercaptoethanol (BME). The affinity columns used are obtained from GE Healthcare. HP stands for high performance, while FF stands for fast flow columns.

3.4.2.2. Affinity chromatography of His-tagged proteins

A 5 ml HisTrap FF (GE healthcare) was used for the IMAC (Immobilized metal ion affinity chromatography). The column was washed with 10 column volumes of MQ water, and then equilibrated with 10 column volumes of Buffer A before the protein filtered lysate was loaded on it with a Gilson peristaltic pump- MINIPULS 3. After loading the protein sample, the column was washed with additional 10 column volumes of Buffer A to remove any proteins non-specifically bound to the column. The FPLC ÄKTA purifier system along with UNICORN control software from GE healthcare was then used to carry out the elution step. A linear gradient of increasing concentrations of the elution buffer was used to identify at what elution buffer concentration the protein of interest elutes from the column as pure as possible. Fractions of 5 ml were collected and the fractions were then analyzed and visualized on 10 or 15% SDS-PAGE for checking which aliquots contained protein and the state of purity of the protein.

3.4.2.3. Affinity chromatography of GST-tagged proteins

Filtered protein lysate was loaded on a 5ml GSTrap HP column (GE Healthcare) that was already washed with 10 column volumes of MQ water and equilibrated with 10 column volumes of Buffer A all using a Gilson peristaltic pump MINIPULS 3. Once the sample was all loaded into the column, the column was washed with 30 ml of Buffer A to remove any unbound, and the column bound protein was eluted with 100% elution buffer directly. The samples were then visualized on 10% SDS-PAGE.

3.4.2.4. Affinity chromatography of MBP-tagged proteins

The same procedure followed for the GST-tagged proteins was used for purifying MBP-tagged proteins but on MBP HP columns (GE Healthcare). MBP tagged Protein was eluted with the elution buffer containing 20 mM Maltose.

Additionally, in other cases an amylose resin (NEB) packed gravity column was used instead of the commercially packed MBP HP columns. However, the same protocol was followed but manually without the use of a peristaltic pump.

3.4.2.5. Affinity chromatography using HiTrap Heparin HP column

In some cases where a protein of interest co-purified with nucleic acids, an additional step of purification using a HiTrap Heparin column was used. The column was washed with 10 column volumes of MQ water, and equilibrated with 10 column volumes of Buffer A. The protein sample to be loaded on the Heparin column was first dialyzed against a low salt buffer (Buffer A for HiTrap Heparin purification, Table 3.6). The elution of the column was done using the FPLC ÄKTA purifier system and managed using UNICORN control software (GE Healthcare) using a linear gradient of increasing concentrations of elution buffer. The absorbance at 260 and 280 nm was constantly monitored in order to check which protein fractions were nucleic acid-free. All purification samples were then checked on 10% SDS-PAGE gel.

3.4.2.6. Protein tag digestion

The proteins where the digestion of their fused tag was required were dialyzed against Buffer SE overnight using a protein Slide-A-Lyzer Dialysis Cassette (Thermo Scientific) with the molecular weight cutoff size depending on the molecular weight of the protein to be dialyzed. PreScission protease (PP) (human rhinovirus 3C protease and GS), consisting of human rhinovirus 3C protease fused to GST, was added to the sample depending on the quantity of protein (1 µg/ml) and left to digest the Tag overnight at 4 °C. For proteins that had a TEV (Tobacco Etch Virus) cleavage site instead, same procedure was followed with TEV protease added instead of PP.

After digestion, the protein samples were then passed again through GSTrap (for GST-tagged proteins), GSTrap followed by His Trap (for His tagged proteins), or GSTrap followed by MBP column (for MBP tagged proteins) in order for the column to capture the digested tag and the PP while tag-free protein pass through. The samples were also run on acrylamide gel of 10 or 15% to follow the success of the digestion process.

3.4.3. Size exclusion chromatography (gel filtration)

After affinity purification, the fractions with obtained protein of interest were concentrated, using Amicon Ultra (Millipore) with a size cutoff depending on the molecular weight of each protein, to a final volume of 1-2 ml. This concentrated protein was then injected alone or mixed with another concentrated protein into either Hiload Superdex 200 16/60 column (GE Healthcare) or Hiload Superdex 75 16/60 column for smaller proteins using Äkta Prime (GE Healthcare). Before injecting, the columns were washed with 180 ml of MQ water then equilibrated with 180 ml of SE buffer. The gel filtration columns were previously calibrated by well-defined protein standards with known molecular weight before use in order to properly estimate the results. Elution of the protein was monitored by absorbance at 280 nm, and samples from the elution peak (s) of the protein were analyzed on SDS-PAGE (10 or 15%). The fractions containing pure protein of interest were then concentrated using Amicon Ultracel concentrators until a volume containing desired concentration of protein was obtained. The protein aliquots were then flash-frozen in liquid nitrogen and stored at -80 °C.

3.5. SDS-PAGE (Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis)

During all steps of different purifications, aliquots of proteins in each step were prepared by mixing with SDS-PAGE loading buffer 6x (0.3 M Tris pH 6.8, 26.1% glycerol, 2% SDS, 7 mg Bromophenol Blue, 675 µl ß-mercaptoethanol, 30% distilled water) and boiled at 95 °C for 10 minutes then centrifuged for one minute to spin down. The samples were then run on 10% polyacrylamide gels (or 15% for smaller protein molecular weight) using a voltage of 200 V and using NuPAGE MOPS SDS at 1X (Life Technologies) as a running buffer.

For visualization, the protein gels were stained with Coomassie Blue solution (1g Coomassie Blue, 40% methanol, 10% acetic acid, 50% distilled water) for 5-10 minutes followed by de-staining in a solution of 10% Methanol, 10% acetic acid, 80% distilled water.

3.6. Protein quantification

The final concentration of the protein was assessed by using the Bradford method (Bio-Rad) against Bovine Serum Albumin (BSA) standard curve by mixing 2 µl of the protein with 998 µl of the Bradford solution in a cuvette and measuring the absorbance at 600 nm using the spectrophotometer Ultrospec 10 Cell density meter (Amersham Biosciences).

Another more accurate approach to measure the concentration of the protein was by mixing the protein in equal amounts with 6M guanidinium chloride solution, and measuring the absorbance at A280 by applying Beer-Lambert law using NanoDrop (Thermo Scientific) taking into account the extinction coefficient of each protein calculated using ProtParam tool (ExPASy).

3.7. Native gel electrophoresis

For the native gel band shift assay, a total amount of 20 µg of protein was incubated with increasing concentrations of total cell extract RNA done in our laboratory by Dr. Sara Zamora Caballero and incubated on ice for around 15-30 minutes. The samples were then mixed with 5x sample buffer (0.3M Tris-HCL pH 8.5, 0.05% (w/v) bromophenol blue, 40% (v/v) glycerol) and loaded on a 10% native polyacrylamide gel using a Tris-Glycine running buffer at pH 8.3. The gel was previously pre-run alone with the buffer for 40 minutes at 100 V before running the samples. The voltage used to run the gel was 100 V and it was run for 4 hours at 4 °C. The later staining and de-staining of the gel were done as described in section 3.5.

3.8. *In vitro* binding experiments

3.8.1. Poly (U) agarose beads binding experiment

10 mg of lyophilized Polyuridylic acid-Agarose powder (Sigma Aldrich; P8563) were weighed and washed with DEPC (diethyl pyrocarbonate) treated, RNAse free, water and left to swell for few minutes. The beads were then equilibrated two times with reaction buffer containing: 50 mM Hepes pH 7.5, 100 mM NaCl, 0.1 mg/ml BSA, and 2.5 mM $MgCl_2$). After that, 200 µg of purified protein (or protein complex) diluted in the reaction buffer were loaded on the beads and left mixing at 4 °C for 30-60 minutes with constant rotation. The beads were then spinned down and the un-bound protein fraction discarded. The beads were after washed 3-4 times with 0.5-1 ml wash buffer not containing BSA, and then the Poly(U) was separated from the beads by the addition of 30 µl of 6x SDS-PAGE loading buffer and boiled for 5 minutes at 95 °C. The samples were then centrifuged spanned down for 1 minute and supernatants loaded on 10% SDS-PAGE or analysis.

3.8.2. Pull down assay

On a small scale, 100-200 µg of tagged bait protein lysate obtained after small-scale expression and small-scale purification as described earlier (section 3.3.1, 3.4.1) was incubated with 70-100 µl of beads previously washed and equilibrated for around 30 minutes at 4 °C with constant rotation. The type of beads used depended on the type of tag fused to the bait protein. Afterwards, the beads were washed with 1 ml SE buffer to remove un-bound protein fraction by centrifugation and discarding supernatants. The prey protein was then added to the beads in equimolar concentrations and left on a rotator for additional 30 minutes. It has to be noted that the prey protein would not have the same tag as the bait protein. Supernatants were then discarded and unbound protein fraction removed by washing the beads 3-4 times with 1ml fresh SE buffer each time. Finally, the beads – bound protein fraction was eluted by adding 30 µl of 6x SDS-PAGE sample buffer and boiled for 10 minutes at 95 °C.

On a large scale however, after sonication and centrifugation the soluble fraction of the bait protein in the supernatants was loaded on a 5ml column depending on the tag of the protein. The column was then washed with Buffer A to remove any unbound protein and the prey protein supernatant was then loaded onto the column. The prey protein would have no tag, or a tag different than that of the bait protein. The column was additionally thoroughly washed with Buffer A and then eluted depending on the type of the protein tag used to catch the bait protein as explained earlier.

The samples from the pull down, in the case of small scale or large scale were then prepared and analyzed on 10% SDS-PAGE gel as explained in section 3.7. The gel was then visualized to check what protein bands are seen in the elution fractions to assess the binding of the protein(s).

3.8.3. Bio-layer Interferometry

The Bio-layer interferometry system (Pall ForteBio) was used to evaluate molecular interactions and calculate binding affinities. To calculate the binding affinity of different protein constructs to a 15 residues long Poly(U) RNA (Sigma Aldrich), Streptavidin sensors (ForteBio) were used to capture the 5' Biotinylated Poly(U) RNA of 15 residues. The streptavidin biosensors were hydrated for 10 minutes before use in a buffer containing 50 mM Hepes pH 7.5, 150 mM NaCl, 5% Glycerol, 1 mM BME, 0.5 mg/ml BSA. The advanced kinetics setup of the BLItz software was used, and a concentration of 50 µg/ml of the bait biotinylated Poly(U) RNA used for immobilization to the biosensors. On the other hand, 4 µl of at least 3 or 4 different increasing concentrations (0.25-20 µM) of each prey protein were utilized to estimate the binding affinity. The prey proteins were diluted in the same hydration buffer containing 0.5 mg/ml BSA that was used throughout the experiment in order to eliminate non-specific binding of the prey protein to the Streptavidin sensor if present. This negative control was in turn used as a reference to calculate the binding affinity. Curve fitting, association constant (Ka), dissociation constant (Kd), and K_D calculations were done using BLItz Pro 1.2 software.

On the other hand, for assessing the binding of different GST tagged Mex67 constructs to Mip6, the bait protein was immobilized on an anti-GST bionsensors (ForteBio). The anti-GST biosensors were also hydrated in the same buffer mentioned earlier for 10 minutes before use. Then different concentrations of the prey protein were used to evaluate the association and dissociation steps. 0.5 mg/ml of BSA in the buffer used throughout the experiment was utilized to eliminate non-specific binding. The curve fitting and the binding affinity was calculated using the BLItz Pro 1.2 software.

During the different biolayer interferometry experiments, the duration of the loading, association, and dissociation steps depended on the saturation of the signal obtained. Additionally, the Streptavidin biosensors were regenerated with 0.5 mM NaOH, while the anti-GST biosensors were regenerated by using 10 mM glycine pH 1.9. The biosensors were then dehydrated by incubating them in 15% Saccharose solution prepared in PBS 1x before being dry stored at room temperature.

3.8.4. Isothermal Titration Calorimetry (ITC)

For the ITC experiments, purified proteins were used. The proteins were dialyzed using Slide-A-Lyzer Dialysis Cassettes (Thermo Scientific) in the same buffer before performing the experiment. Nano ITC from TA instruments was then used to perform the measurements. All samples were spinned down and degassed for 10 minutes using the degassing work station before being slowly and carefully loaded in the cell (200-250 µl) with a Hamilton syringe or in the titration syringe (50 µl). All experiments were performed at 15 °C, and consisted of a serious of 20-30 serial injections depending on each case. The concentration of proteins used either in the syringe or in the cell, plus the number of injections and the volume of injections for the experimental results shown in this study is summarized in table 3.7. A blank experiment where the titrant was used against buffer was used as data reference, as the heat peaks produced from it are the results of dilution effect to be excluded when analyzing the data. The data fitting and analysis was performed using NanoAnalyze v.3.7.0. software (TA Instruments).

Experiment	Syringe	Cell	Injections number	Injection volume
Mex67$_{(528-599)}$/ Mip6$_{(111-480)}$	Mex67$_{(528-599)}$: 800 µM	Mip6$_{(111-480)}$: 25 µM	25	2 µl (1.96 µl)
Mex67$_{(528-599)}$/ Mip6 RRM4$_{(390-480)}$	Mex67$_{(528-599)}$: 690 µM	Mip6 RRM4$_{(390-480)}$: 50, 70	25	2 µl (1.96 µl)
Mip6 RRM4$_{(390-480)}$/ Mex67$_{(528-599)}$	Mip6 RRM4$_{(390-480)}$: 700 µM	Mex67$_{(528-599)}$: 50 µM	30	1.5 µl
Mex67$_{(528-599)}$/ Mip6 RRM4 W442A	Mex67$_{(528-599)}$: 690 µM	Mip6 RRM4 W442A: 50 µM	25	2 µl (1.96 µl)
GST-Mex67 UBA/ Mip6 RRM4$_{(390-480)}$	GST-Mex67 UBA: 612-	Mip6 RRM4$_{(390-480)}$: 20-80 µM	20 or 25	2.4 µl (2.38 µl) or 2 µl (1.96 µl)

	1800 µM			
Mex67$_{(528\text{-}599)}$/Pes4 RRM34	Mex67$_{(528\text{-}599)}$: 800 µM	Pes4 RRM3/4: 25 µM	25	2 µl (1.96 µl)

Table 3.7. ITC conditions and protein concentrations used for the experiments mentioned in this study.

3.8.5. Protein Cross-linking

200 µg of the stable complex Mex67$_{(528\text{-}599)}$/Mip6 RRM4$_{(390\text{-}480)}$ obtained from gel filtration was mixed with different concentration of Sulfo-SMPB (sulfosuccinimidyl 4-maleimidophenyl)butyrate) cross-linker (Thermo Scientific) for 30, 60, 120 minutes at room temperature. Both the protein and the cross-linker were dissolved in SE buffer (refer to table 3.6) into a total volume of 25 µl. The reaction was stopped by adding 1 µl of Tris-HCL pH 8 (0.5M), and samples were heated at 95°C for 10 minutes with the 6x sample buffer and run on 15% SDS-polyacrylamide gel. After visualizing the gel, the cross-linked protein band of interest was extracted from the gel and sent to mass spectrometry analysis service at the University of Valencia, Burjassot, Valencia.

3.9. Protein crystallization and structure resolution

3.9.1. Crystals of Mip6 RRM3$_{(313\text{-}389)}$

After size exclusion chromatography, the protein was concentrated to final concentration and different concentrations were used to set up crystallization screening plates. The crystallization plates were set up using sitting drops vapor diffusion crystallization in 96-well plates at 21 °C using commercial crystal screens available. Crystals were obtained in drops containing 8-10 mg/ml of protein in drops containing 0.5 µl of protein and 0.5 µl of reservoir. The initial crystals were cylinder in shape and diffracted up to 8 Å or needle-like crystals that were not unique. In order to improve the needle crystals, some conditions were chosen to do additive screening with Hampton additives kit. The drop from which the first structure was solved contained 0.5 µl of protein sample, 0.5 µl of reservoir, and 0.1 µl of each added additive. On the other hand, some crystallization plates were set by mixing Mip6 RRM3$_{(313\text{-}389)}$ protein with Poly(U) RNA 7 residues long purchased from Sigma Aldrich in a 1:1.2 protein to RNA ratio.

The obtained crystals were flash-cooled in liquid nitrogen and diffracted at XALOC beam line of ALBA Synchrotron (Barcelona, Spain) at a wavelength of 0.97934 or 0.9795. Data were processed using *XDS* (X-Ray Detector Software) (Kabsch, 2010), then merged and scaled in *Scala* from *CCP4* (Winn *et al.*, 2011). Molecular replacement was then carried out using Phaser-MR module in *Phenix* (Adams *et al.*, 2010). The PDB model used for the molecular replacement of the first structure of Mip6 RRM3 was the RNA recognition motif of the Cap binding complex (CBC) subunit 2 (PDB ID: 3FEY). The initial phases were then used for additional model building using *AutoBuild* module from *Phenix*. Further structural refinement was achieved using a combination of *Phenix refine* suite and manual refinements in *COOT* (Emsley, Lohkamp, Scott, & Cowtan, 2010). The first structure was deposited in the Protein Data Bank (PDB) with PDB ID: 5D77. It was then used as a model for the molecular replacement of the later structure deposited with PDB ID: 5D78.

3.9.2. Crystals of Mex67$_{(528\text{-}599)}$

The crystallization screening plates were set either with purified Mex67$_{(528\text{-}599)}$ alone or by mixing 1:1 ratio with different Mip6 constructs. The crystallization plates were set at 21 °C using the sitting drop vapour diffusion method and crystals were obtained by vapour diffusion using sitting drop method after few months in different conditions (table 4.3). The crystals were cryo-cooled in nitrogen and diffracted at Diamond light source (Harwell, UK) I03 and I04 beam lines. Best crystals diffracted up to 1.3 Å and data were automatically processed at Diamond by xia2 pipeline. Data were then scaled using *Aimless* from *CCP4*. *BALBES* automatic molecular replacement pipeline available online was used to search for a model for molecular replacement (Long, Vagin, Young, & Murshudov, 2007). The search model identified by *BALBES* was that of the UBA domain from the human homolog of Mex67 (PDB ID: 1OAI) and it was used for molecular replacement utilizing Phaser-MR module from *Phenix*. The MR was successful and the obtained initial phases were used to build the rest of the structure using *AutoBuild* from *Phenix* and additional manual building. Cycles of *Phenix refine* and *COOT* were used further structural improvements and refinements. Also, the first structure of Mex67$_{(528\text{-}599)}$ obtained was used as a MR model for the subsequent structure.

3.9.3. Crystals from Pes4 RRM3/4

3.9.3.1. Crystal structure of Pes4 RRM3

The crystals from which the structure was solved were obtained from a pH screen of an initial condition containing 0.2 M Ammonium sulphate and 20% Polyethylene glycol (PEG) 8000 where initial bundle of needle-like crystals were obtained. The crystals were flash-cooled with liquid nitrogen directly or cryo-protected with paratone oil and were diffracted at Diamond light source (Harwell, UK) I03 beam line where the best diffracted to 1.7 Å. Data were automatically processed with *auto-PROC* and further scaled using Aimless from *CCP4* where the resolution was cut to 1.9 Å. Initial phases were obtained by using Mip6 RRM3 (section 3.9.1) as a model for molecular replacement using Phaser-MR from *Phenix*. Next, *AutoBuild* from *Phenix* was used to build the structure. Cycles of *Phenix refine* and *COOT* was utilized in order to obtain the final structure. TLS parameters refinement was utilized in the refinement strategy.

3.9.3.2. Crystal structure of Pes4 RRM4

One of the crystals obtained during the additive screening of the initial needle clusters in a condition containing 0.2 M Ammonium Nitrate, 20% PEG 3350, and 0.1 M Betaine Monohydrate as an additive, belonged to $P2_12_12_1$ space group with cell dimensions: a= 27.501, b= 36.162, c=73.583 Å; α=β=γ= 90°. The crystal diffracted up to 1.1 Å at XALOC beam line of ALBA synchrotron (Barcelona, Spain). The Data were processed by *XDS* and further merged and scaled using *Aimless* from *CCP4*. *Mathews coefficient* in CCP4 was also used to assess the size of the molecule in the crystal and how many molecules to be expected in the asymmetric unit. Since the obtained resolution was very high, *ARCIMBOLDO* suite (D. Rodríguez *et al.*, 2012) in *CCP4* was used to obtain the initial phases. *Buccaneer* (Cowtan, 2006) from CCP4 was further used to build the structure. The model manual refinement was done using *COOT* and further refined using *Phenix refine* for anisotropic refinement until final good R factors were obtained.

3.9.3.3. Crystal structure of RNA-free Pes4 RRM3/4

Crystals of Pes4 RM3/4 grew in 10% PEG 4000, 10% 2-Propanol, 100 mM Sodium citrate pH 5.6, or in 0.2M Ammonium sulfate, 0.1 M Sodium Cacodylate trihydrate pH 6.5, 30% PEG 8000 at 21 °C at a purified protein concentration of 11 mg/ml with 1:1 equal volumes of protein and precipitant (0.5 µl). The crystals were flash-cooled in liquid nitrogen and diffracted at XALOC beam line of ALBA synchrotron (Barcelona, Spain). The crystals diffracted poorly to 2.6-2.8 Å and data were processed by *XDS* and further merged and scaled using *Aimless* from *CCP4*. *Mathews coefficient* from *CCP4* was used to assess the expected number of molecules in the asymmetric unit. Molecular replacement was successful using Pes4 RRM4 as a model. However, the solution was only partial and only RRM4 domain was placed by the program in the structure. The linker and RRM3 had to be placed and fitted manually with consecutive cycles of fitting corrections and refinement with *COOT* followed by *Phenix Refine*. TLS parameters where used in the refinement strategy until the R factors mentioned in this study were reached.

3.9.3.4. Crystal structure of Pes4 RRM3/4 with RNA

The purified protein of Pes4 RRM3/4 obtained from gel filtration was mixed with HPLC purified Poly(U) RNA of 11 residues purchased from Sigma Aldrich at 1:1 or 1.2:1 molar ratio RNA:protein before setting the crystallization screening plates or mixed directly in drop. The crystals obtained by vapor diffusion using sitting drops were small plates that were flash-frozen with nitrogen and diffracted at XALOC beam line of ALBA synchrotron (Barcelona, Spain). The crystals diffracted poorly and the data obtained was processed with both *XDS* and *iMosflm* (Battye, Kontogiannis, Johnson, Powell, & Leslie, 2011). The data was integrated either in the monoclinic $P12_11$ (no. 3) or in I121 (c2, no. 5) space groups and *Molrep* self rotation function from *CCP4* was checked. The MR module from Phenix was used for molecular replacement trials with Pes4 RRM3/4, Pes4 RRM3, or Pes4 RRM4 as a model. Also *BALBES* was tried to search for a model for molecular replacement.

3.10. Structure surface electrostatic potential

The surface electrostatic potential of the structures was calculated using PDB2PQR & APBS (Adaptive Poisson-Boltzmann Solver) online server version 2.1.1 (Dolinsky *et al.*, 2007) and visualized using *Chimera* software (Pettersen *et al.*, 2004). Structure figures as well as figures of superposed structures were generated either by *COOT* or using *Chimera*.

3.11. NMR spectroscopy titration

The NMR (Nuclear Magnetic Resonance) experiments were done and analyzed in Dr. Jose Manuel Pérez Cañadillas lab (Institute of physical chemistry Rocasolano, CSIC, Madrid, Spain). NMR data were acquired in a Bruker AV-600 or AV-800 spectrometers equipped with a z-gradient triple resonance cryoprobe at 20 °C. Protein samples were prepared at concentrations of 50-150 µM in NMR buffer containing 25 mM potassium phosphate pH 7.0, 150 mM NaCl, 2.5 mM

DTT. Backbone triple ^{1}H, ^{13}C and ^{15}N resonance assignments of Mex67$_{(528-599)}$ and Mip6 RRM4 were recorded on a double-labelled protein samples to assign their 2D ^{1}H,^{15}N-heteronuclear single quantum coherence (HSQC). NMR spectra were processed and analyzed with NMRPipe (Delaglio *et al.*, 1995) and CCPN-Analysis software (Vranken *et al.*, 2005).

For the NMR titration experiment, ^{15}N HSQC spectra were recorded at 20 °C for ^{15}N, ^{13}C-labeled Mip6 RRM4 protein (250 µM) titrated with unlabeled Mex67$_{(528-599)}$ protein in the same buffer containing 25 mM potassium phosphate pH 7.0, 150 mM NaCl, 2.5 mM DTT. Three titration points were recorded at 125 µM, 250 µM, and 375 µM of Mex67 allowing the follow of the signals in fast exchange from the free to bound state. The chemical shift perturbation (CSP) was calculated for each amide signal and normalized using $\Sigma\Delta\delta = |\Delta\delta(1H)| + |(0.2)\Delta\delta(15N)|$, where $\Delta\delta(1H)$ and $\Delta\delta(15N)$ represent the chemical shift differences between the free and bound states. The titration of unlabeled Mip6 RRM4 into ^{15}N,^{13}C-labeled Mex67$_{(580-599)}$ could not be analyzed by the CSP as most of the signals exchange in the intermediated regime (broad signals) during the titration experiment. Instead the effects of Mip6 RRM4 on the Mex67$_{(528-599)}$ were analyzed by monitoring the change of intensity of the signals in series of ^{15}N HSQC spectra taken at low saturation values (< 15% of Mip6 RRM4).

4. Results

4.1. Mip6 protein expression and purification

To better understand the structure and function of Mip6, the initial step was to try to express and purify the full length protein of Mip6. For this purpose, the full length recombinant Mip6 $FL_{(1\text{-}659)}$ protein was cloned in a variety of plasmids with different range of N-terminally fused soluble protein tags (6XHis, GST, or MBP tags) as shown in table 3.5. The Mip6 $FL_{(1\text{-}659)}$ protein was then expressed in *E. coli* in Bl21 (DE3) and Bl21 codon plus (DE3) RIPL competent cells using different culture media (LB or ZY media), and expression systems (IPTG induction and Auto-induction). The baculovirus-insect cells expression system was also tried in order to obtain soluble protein. Unfortunately, the protein could not be expressed in a soluble form, even when co-transformed with chaperones (GroEL/GroES system and DnaK-DnaJ-GrpE system) in an attempt to aid the proper folding of the protein.

Only a small fraction of soluble Mip6 $FL_{(1\text{-}659)}$ was obtained from a fused N-terminal MBP tag, expressed with Auto-induction in Bl21 (DE3) codon plus *E. coli* cells at 16°C during a 48-72 hours duration, and purified with affinity chromatography on an amylose resin packed gravity flow column. The eluted protein fractions were visualized on a 10% polyacrylamide SDS-PAGE gel, and the small soluble fraction obtained was further concentrated and loaded in a gel filtration (Superdex 200 16/60) column. From size exclusion chromatography, the protein eluted in the void volume of the column in a single peak, indicating an aggregated form of the protein (figure 4.1). We suspected that a co-purification of nucleic acids with the protein might be the cause, but the addition of RNase A or DNase to the lysis buffer of the protein did not resolve the aggregation. Also several solution additives used frequently for protein stabilization against aggregation such as 200mM L-arginine, 5-10% glycerol, or different salts (KCl instead of NaCl) were used but didn’t improve the result.

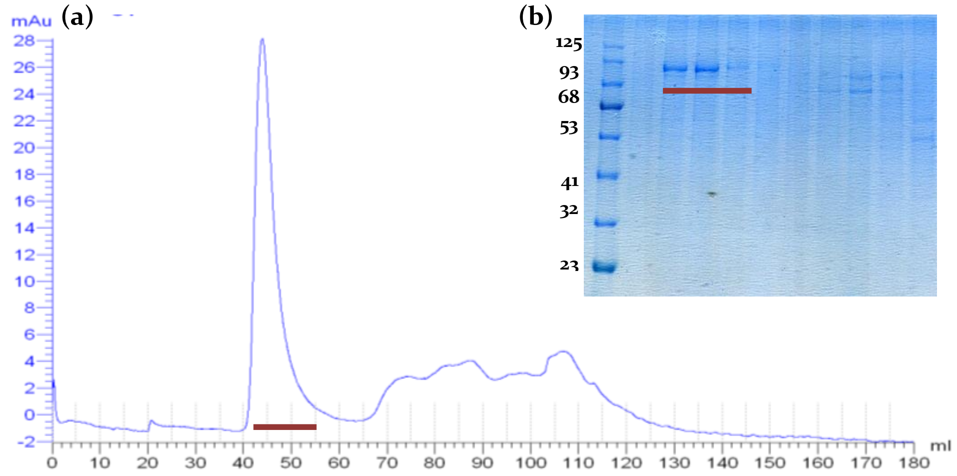


Figure 4.1. **(a)** The gel filtration chromatogram corresponding to the elution profile of MBP tagged Mip6 FL$_{(1-659)}$ as a single peak aggregate eluting in the void volume of the Superdex 200 16/60 column. **(b)** The fractions corresponding to the elution peak containing Mip6 FL$_{(1-659)}$ protein were visualized on a 10% SDS-PAGE and are highlighted with the red bar.

4.2. MIP6 is an RNA binding protein that binds RNA with high affinity

Due to the challenges we faced getting a soluble, non-aggregating Mip6 FL protein, constructing a shorter more stable form of the protein was therefore considered. Running the protein sequence on the RONN disorder prediction server (Yang, Thomson, McNeil, & Esnouf, 2005), predicted the existence of some disordered regions in the protein sequence mainly on its N-terminal region (figure 4.2).

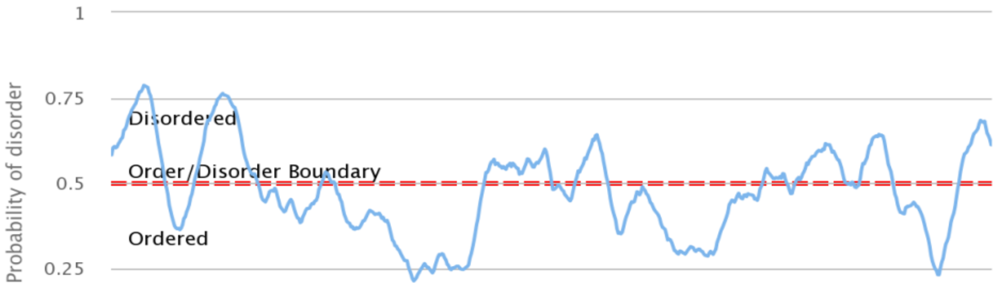


Figure 4.2. Prediction of disordered regions in Mip6 FL sequence. RONN server (Yang *et al.*, 2005) indicated high disordered regions in the N-term of the protein reaching residue 110, and also on the C-term of the protein residues (488-585) and (635-659) which might highlight regions more prone to proteolysis and consequently less stable.

Additionally, as Mip6 was previously suggested to be an RNA binding protein, using Mip6 Fasta sequence to check for conserved domain utilizing the NCBI conserved domain database (CDD) predicted the presence of three RNA recognition motifs (RRMs) and an additional possible fourth RNA binding domain (RBD) in Mip6 protein sequence (figure 4.3) which also goes in agreement with the secondary structure prediction. However though, it was previously reported in literature that Mip6 contained only three RNA binding domains (Segref *et al.*, 1997).

Taking these information into consideration, we generated a construct Mip6$_{(111\text{-}480)}$ that contains all four of the predicted RRMs. This construct was then used to characterize the RNA binding of MIP6 protein.

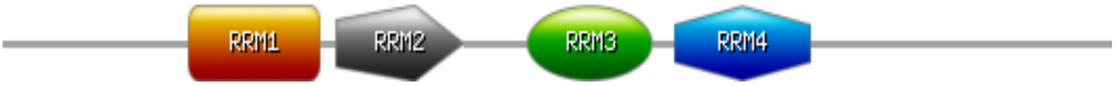


Figure 4.3. Domain organization scheme of Mip6. NCBI domain prediction server predicted the presence of four RRMs where a short linker connects RRM1and2 and RRM3and4 respectively, while a longer linker seems to exist between RRM2 and RRM3.

4.2.1. Expression and purification of Mip6 (111-480)

The GST tagged construct that combines all the predicted RRMs was successfully expressed and purified by affinity chromatography on a GSTrap. The fused GST tag on the N-terminus of the eluted protein was then digested by PreScission protease at 4°C overnight, and the protein was further purified using size exclusion chromatography on a Superdex 75 (16/60) size exclusion column. All the purification steps after the affinity chromatography were carefully completed to prevent the degradation of the obtained protein as the protein was prone to proteolysis (as shown in figure 4.4) possibly due to the presence of relatively long, unstructured linkers between the predicted RRMs. The protein eluted from the gel filtration column at 54.1 ml as a single peak suggesting that Mip6$_{(111\text{-}480)}$ behaves as a monomer in solution (Figure 4.5). The purified protein was then concentrated and flash frozen in liquid nitrogen before being stored at -80°C for further use.

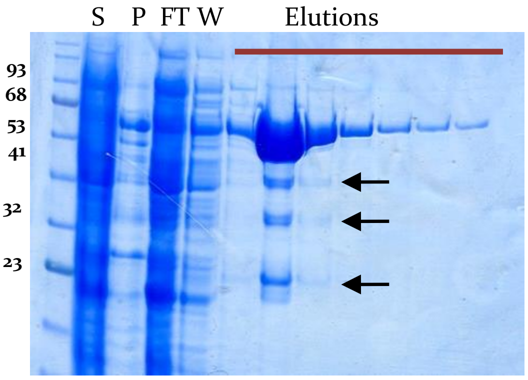


Figure 4.4. Purification of Mip6$_{(111-480)}$. Samples from the different steps: Supernatants (S), non-soluble pellet fraction (P), fraction non-bound to the affinity column (Ft), wash of the affinity column (W), and elutions of affinity chromatography purification of GST-MIP6$_{(111-480)}$ as visualized on a coomassie-stained 10% SDS-PAGE gel. The protein was mostly soluble; some degradation and possible proteolysis is noticed in the most concentrated elution fractions as indicated with the arrows.

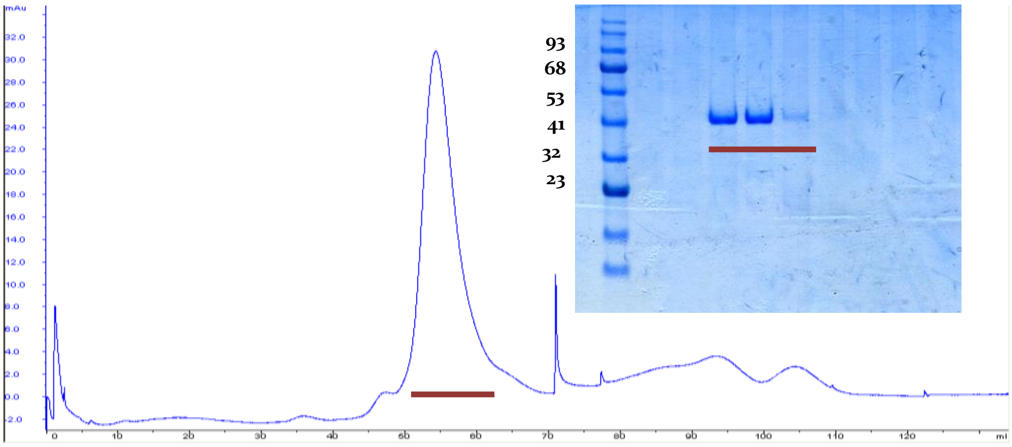


Figure 4.5. Mip6$_{(111-480)}$ elution profile from size exclusion chromatography and its corresponding 10% SDS-PAGE gel. Mip6$_{(111-480)}$ eluted as a single peak corresponding to a monomer that is marked by a red bar on both the chromatogram and the corresponding SDS-PAGE gel where the protein band at 43.52 KDa is visualized.

4.2.2. Mip6$_{(111-480)}$ binds RNA *in vitro*

To investigate whether Mip6 is able to bind RNA *in vitro*, the RNA binding affinity of Mip6$_{(111-480)}$ was assessed using the Bio-layer interferometry system, where an N-terminal 5′ biotinylated general 15-nucleotide long poly(U) RNA was immobilized on streptavidin biosensors. Different concentrations of the purified Mip6$_{(111-480)}$ protein (0.5, 0.75, 1, and 1.25μM) were used as prey protein to calculate the binding affinity to the poly(U) *in vitro* (figure 4.6). Non-specific binding of the prey protein to the streptavidin sensors was minimized to almost no detection by adding 0.5 mg/ml of BSA to the buffer solution used in the experiment. The association and dissociation phases were well fitted with the interaction having a binding affinity in nano-molar range K_D= 0.345 µM proving a high binding affinity to a rather non-specific RNA which supports a role of Mip6 as a universal RNA binding protein. The experiment was repeated confirming the result.

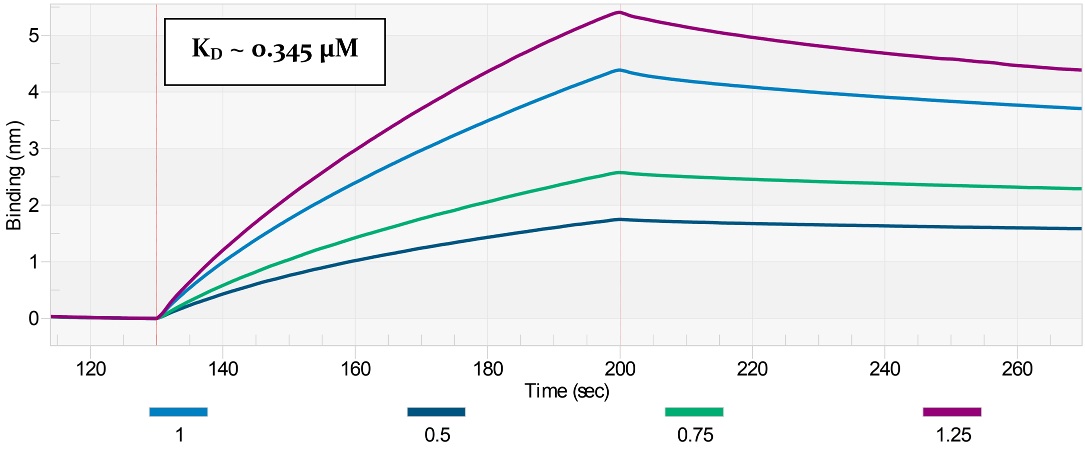


Figure 4.6. Mip6 binds RNA with high affinity. Association and dissociation steps from the Bio-layer interferometry experiment using different molar concentrations (in µM) of Mip6$_{(111\text{-}480)}$ depicted in the figure with different colors. The binding affinity to general Poly(U) RNA immobilized on streptavidin sensors was calculated in the nanomolar range.

4.3. MIP6 has a fourth RNA recognition motif

In literature, it was suggested that Mip6 harbours three RNA recognition motifs. To investigate whether this hypothesis is true, we decided to verify the RNA binding abilities of the separate predicted RRMs and check whether an additional fourth RRM exist. In order to evaluate this hypothesis, different constructs harboring separate predicted RRMs were generated: Mip6 RRM1/2$_{(111\text{-}279)}$, Mip6 RRM3/4$_{(313—480)}$, Mip6 RRM3$_{(313\text{-}480)}$, and several truncations of the predicted Mip6 RRM4 as shown in figure 4.7 below. These truncations were then tested for their ability to recognize and bind RNA independently.

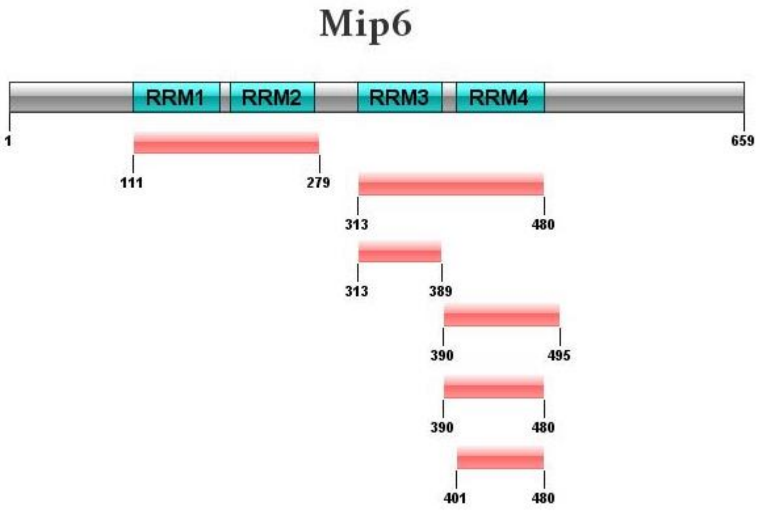


Figure 4.7. The different Mip6 constructs generated to include individual or combination of the predicted RNA binding motifs.

4.3.1. Expression and purification of Mip6 RRM1/2 and MIP6 RRM3/4

Mip6 RRM1/2 and Mip6 RRM3/4 containing residues 111-279 and 313-480 respectively were cloned in a plasmid with an N-terminal GST tag. The plasmids transformed into the bacterial Bl21 codonPlus RIPL strain were then successfully expressed using the Auto-induction expression system at 18°C overnight. The expressed proteins were then purified using affinity chromatography on a GSTrap. A good amount of soluble protein was obtained from both construct. The fused GST tag of the purified proteins was digested afterwards using PreScision protease and the obtained proteins were further purified using size exclusion chromatography. Both proteins eluted from gel filtration as a single, monomeric peak in which its corresponding protein fractions were then concentrated to final concentration and used for further experiments or stored at -80 (figure 4.8).

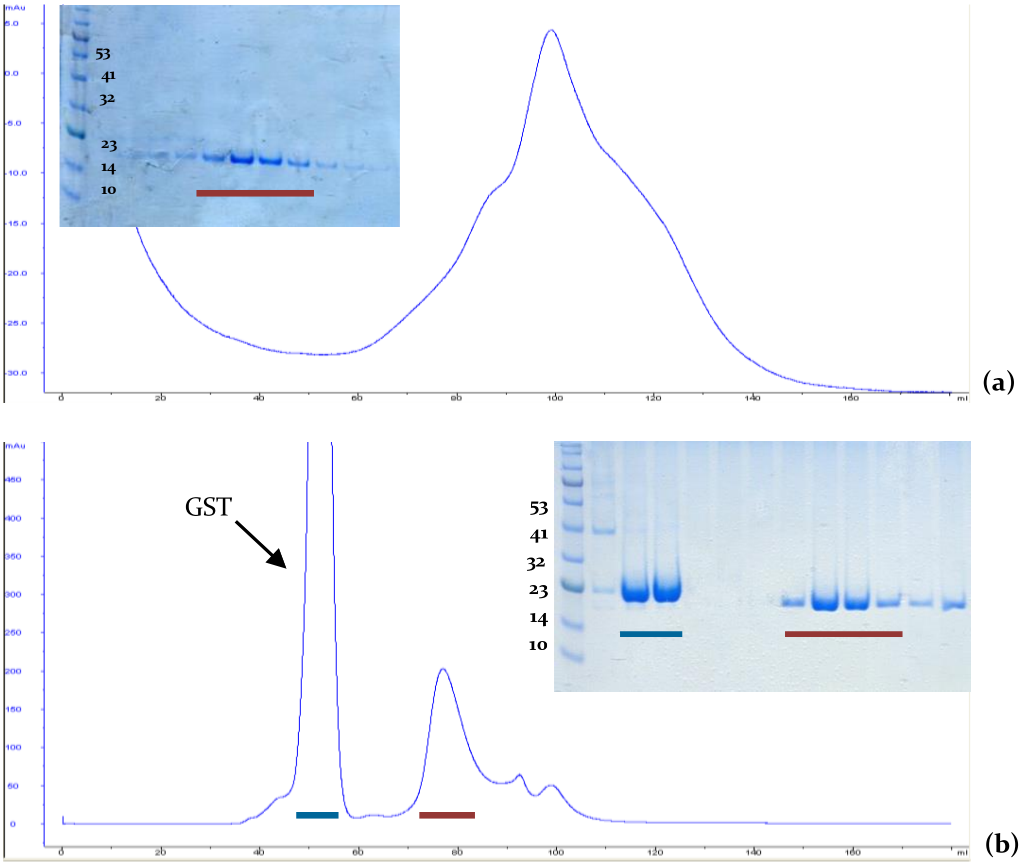


Figure 4.8.

(a). Elution profile from the size exclusion chromatography purification of Mip6 RRM1/2$_{(111-279)}$. The protein eluted as a monomer at 99.1 ml from Superdex 200 16/60 gel filtration column. The fractions corresponding to the peak were visualized on a 10% SDS-PAGE gel and marked with a red bar.

(b). the chromatogram showing the elution profile of Mip6 RRM3/4$_{(313-480)}$ protein from the gel filtration Superdex 75 16/60 column. The protein of 19.92 KDa eluted as at 77 ml of volume as a single, monomeric peak. The corresponding fractions were visualized on a 10% SDS-PAGE gel where the purified protein is marked with a red bar, while the fractions with a slightly higher band corresponded to the excess free GST (marked in dark blue) after the GST tag digestion.

4.3.2. Expression and purification of Mip6 RRM3$_{(313-389)}$

Mip6 RRM3$_{(313-389)}$ was expressed in BL21 codon plus (DE3) RIPL bacterial cells using the Auto-induction expression system in ZY medium at 20°C overnight. The cells were collected, lysed, and purified using affinity chromatography on a 5 ml GSTrap. A good soluble amount of the protein was obtained. The proteins' N-terminal GST tag was then removed by incubating it with PreScision protease at

4°C overnight. The protein was further purified with size exclusion chromatography where it eluted in a single peak. Analysis of the purified Mip6 RRM3$_{(313\text{-}389)}$ showed more than 95% purity and the protein was then concentrated and flash frozen in liquid nitrogen and stored at -80 °C (figure 4.9).

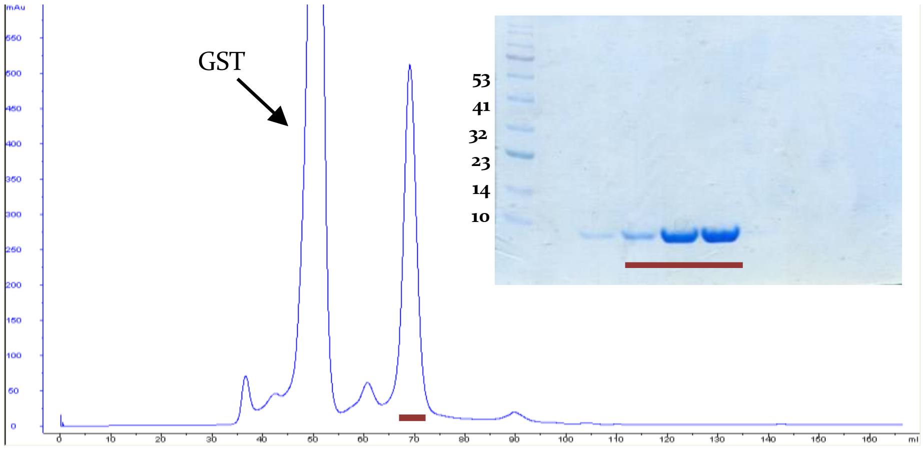
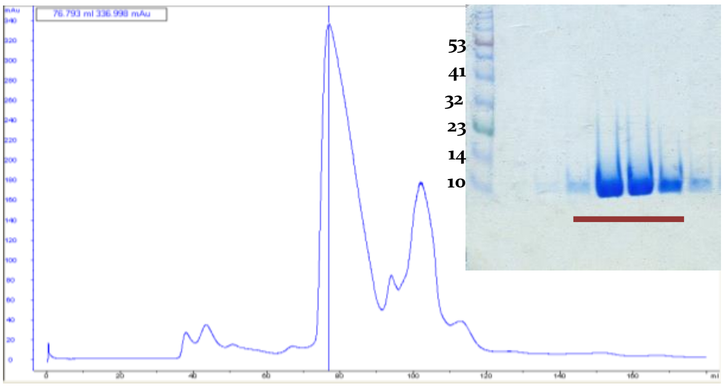


Figure 4.9. Gel filtration elution profile of Mip6 RRM3$_{(313\text{-}480)}$. The protein eluted in a single peak in a volume corresponding to the monomeric size of the protein. A 10% SDS-PAGE showed a good amount of the protein as clean and pure. The protein peak and its correspondence on the gel are represented by the red bar. The additional peak shown with an arrow was corresponding to excess free GST after tag removal.

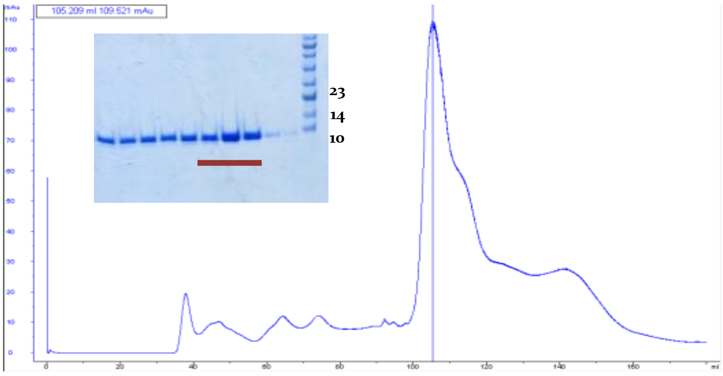
4.3.3. Expression and purification of Mip6 RRM4

On the other hand, due to the lack of previous characterization of a fourth RRM on Mip6 and the subsequent lack of information on the possible boundaries, three constructs containing RRM4 were expressed and purified: 6xHis tagged Mip6 RRM4$_{(390\text{-}495)}$, 6xHis-Mip6 RRM4$_{(390\text{-}480)}$, and GST- Mip6 RRM4$_{(401\text{-}480)}$. The proteins were purified, depending on its fused tag, using affinity chromatography on either HisTrap column or GSTrap followed by exclusion chromatography on Superdex 75 (16/60). The proteins then were either digested with PreScession protease or left non-digested depending on the requirements of the subsequent experiments.

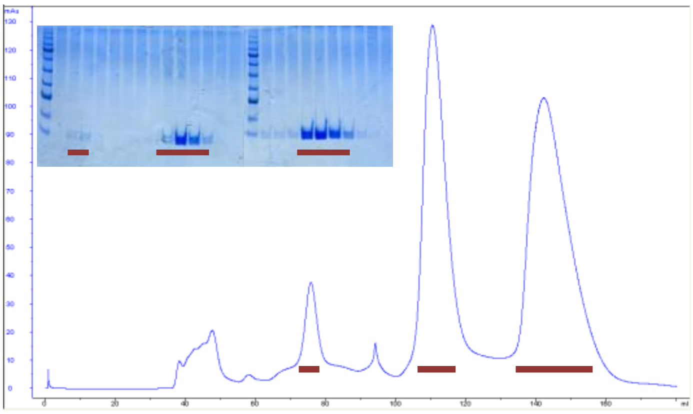
It has to be noted that in the case of the 6xHis tagged Mip6 RRM4$_{(390\text{-}495)}$, the major part of the protein eluted from gel filtration in the volume (~76.79 ml) corresponding to a monomer, while in the case of the other two constructs (Mip6 RRM4$_{(390\text{-}480)}$ and Mip6 RRM4$_{(401\text{-}480)}$) the protein did not elute from gel filtration in the volume expected for its size, instead it eluted in the retention volume of the column after the expected size possibly due to non-specific interaction with the column matrix or maybe due to a possible compact conformation making the protein behave as if it was of a lower molecular weight. Only a very small portion eluted in a smaller peak corresponding to a dimer. This was noticed especially when the protein was in higher concentrations. Increasing ionic strength by using higher concentrations of NaCl salt in the solution buffer did not help resolve the retention problem. On some occasions, when the protein co-purified with nucleic acids, an additional step of HiTrap Heparin purification was used to remove any co-purifically bound nucleic acid and to get cleaner protein.



(a)



(b)



(c)

Figure 4.10. Size exclusion chromatography elution profile of Mip6 RRM4.

(a). Gel filtration chromatogram corresponding to the elution profile of 6xHis tagged Mip6 RRM4$_{(390-495)}$ and the 10% SDS-PAGE gel corresponding to the obtained protein. The protein eluted mostly in the volume corresponding to its monomeric size.

(b). The gel filtration elution profile of Mip6 RRM4$_{(390-480)}$ with 6xHis tag eluted from gel filtration after the elution volume for its expected size. The fractions corresponding to the highest peak are marked in red.

(c). The elution profile of Mip6 RRM4$_{(401-480)}$ after 6xHis tag removal consisted of three peaks: a smaller peak where the elution volume corresponds to a dimer but the amount of protein in this peak was very small, and another two peaks after the protein expected elution volume. The corresponding fractions were visualized on a 10% SDS-PAGE gel and marked with a red bar that corresponds to each peak.

4.3.4. Mip6 RNA recognition motifs independently bind poly-uridylic acid *in vitro*

After expressing and purifying the constructs of Mip6 containing the different RRMs, their separate ability to recognize and bind RNA *in vitro* was evaluated. For this purpose, polyuridylic acid (PolyU) agarose beads were used. The proteins in the purified form were incubated at 200 µg with the beads for one hour with constant rotation and then washed 3-4 times with buffer containing (50mM Hepes pH 7.5, 100mM NaCl, 2.5mM $MgCl_2$) prior to elution. As shown in figure 4.11, binding to the Poly(U) of the predicted RRMs: Mip6 RRM1/2$_{(111-279)}$, RRM3/4$_{(313-480)}$, RRM3$_{(313-389)}$, and RRM4$_{(390-480)}$ was verified as the proteins were visualized bound to the Poly(U) RNA on polyacrylamide gel indicating a clear RNA binding. However, the binding of Mip6 RRM3 to Poly(U)-agarose was notably lower than the other RRMs or RRM4 as the intensity of the band corresponding to Mip6 RRM3$_{(313-389)}$ as seen on the SDS-PAGE was much lower.

Although, unlike RRM3 and RRM4, Mip6 RRM1 and RRM2 were not tested for RNA binding separately, but rather joined in the Mip6 RRM1/2$_{(111\text{-}279)}$ construct and thus our results were not indicative if each of them on its own would be able to bind RNA, it is most probably the case. Thus, our results confirm Mip6 as an RNA binding protein containing multiple RRM domains, most likely four, capable of binding RNA independently. For this experiment, BSA was used as a negative control.

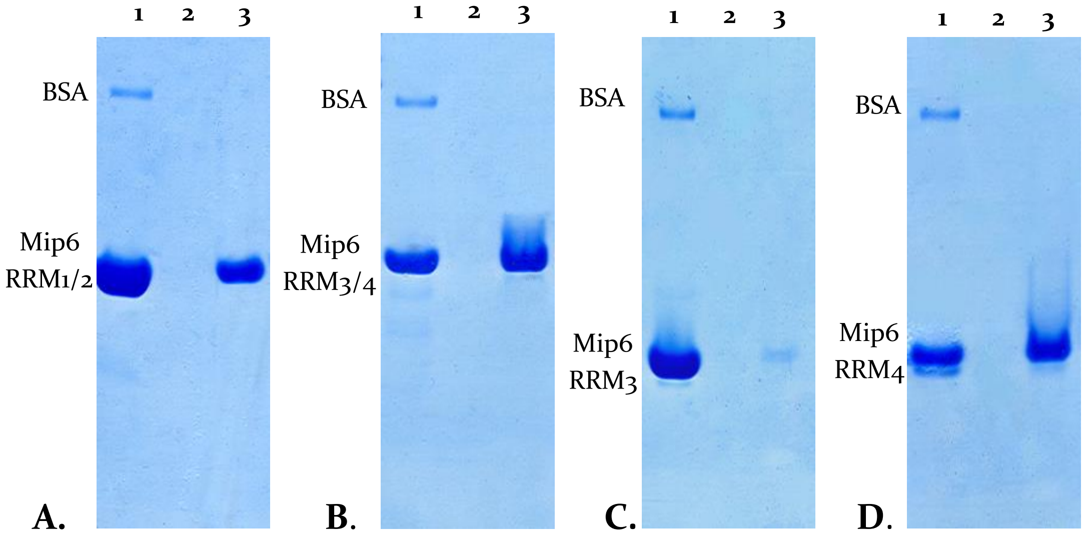


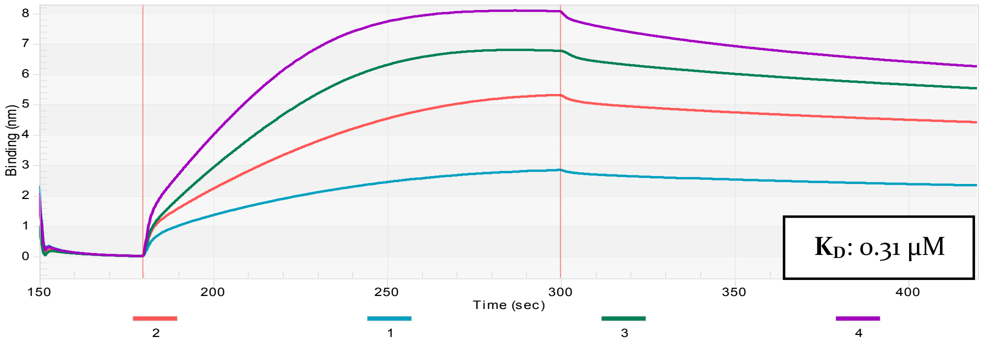
Figure 4.11. Mip6 has four RNA binding domains.
Truncations containing RRM domains of Mip6 were incubated with poly(U) beads (1: protein input), beads were then washed thoroughly (2: wash), and then eluted (3: proteins that were attached to the beads after washing) and samples were visualized on 10% SDS-PAGE gel stained with Coomassie. Proteins co-eluted from resin together with polyU indicating binding.
A. Mip6 RRM1/2$_{(111\text{-}279)}$ **B.** Mip6 RRM3/4$_{(313\text{-}480)}$ **C.** Mip6 RRM3$_{(313\text{-}389)}$ **D.** 6xHis-Mip6 RRM4$_{(390\text{-}480)}$

4.3.5. Mip6 RRM4 binds RNA with high affinity

After characterizing a fourth RNA binding domain in Mip6, we decided to go further and precisely measure the binding affinity of the protein truncation containing the boundaries of the expected Mip6 RRM4$_{(390\text{-}480)}$ to a general 15-nucleotide long Poly(U) RNA immobilized on streptavidin sensors using Bio-layer interferometry. Mip6 RRM4 was able to bind Poly(U) RNA with a good micro molar affinity of K_D = 5.7 µM assuring its position as an RNA binding motif (figure 4.12).

Additionally, From the intensities of the bands visualized on the polyacrylamide gel in the Poly(U) beads binding experiment, we could not assess whether the binding to RNA is better when two RRMs are joined rather than independent, therefore the binding affinity of Mip6 RRM3/4$_{(313-480)}$ to the biotinylated Poly(U) RNA was also measured. A binding affinity of K_D 0.31 µM was obtained indicating a stronger interaction with the two RRMs combined (figure 4.12). This possibly suggests the presence of an additional RRM may provide a larger binding platform for a better cooperative recognition and interaction with RNA.

(a) Mip6 RRM3/4$_{(313-480)}$



(b) Mip6 RRM4$_{(390-480)}$

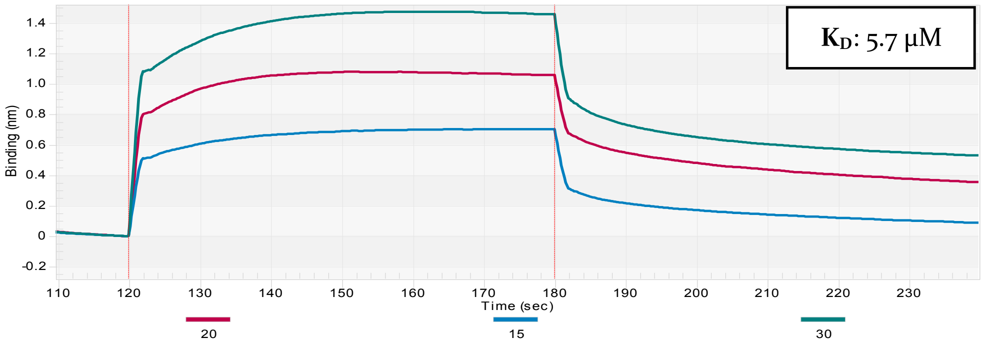


Figure 4.12. RNA binding affinity calculations for Mip6 RRM3/4 and Mip6 RRM4 using Bio-layer interferometry. Association and dissociation curves obtained for Mip6 RRM3/4$_{(313-480)}$ **(a)** and Mip6 RRM4$_{(390-480)}$ **(b)** showed a high affinity binding to the general biotinylated Poly(U) RNA of 15 residues immobilized on streptavidin sensors. The curve corresponding to each protein concentration used (in µM) is depicted in different colors.

Crystallization trials of all the above utilized constructs comprising different RRM combinations from Mip6 with or without poly(U) RNA was carried out, but till the day crystals that led to structure solving were only obtained from Mip6 RRM3$_{(313-389)}$.

However, our attempts to crystallize Mip6 RRM1/2 with Poly(A) RNA of 11 residues rapidly produced crystals that diffracted to 1.25-1.35 Å and a high quality data was collected. Yet, the solved structure did not contain any protein, instead the structure turned out to be that of a parallel RNA double-helix formed by Poly(A) with one double-helix in the asymmetric unit (figure 4.13). In the structure, the double-helix was formed by two symmetric adenine nucleotides stabilized by hydrogen bonds formed between the hydrogen atom of the amino group of one adenine and the Nitrogen N7 atom of the symmetric adenine on one hand, and another hydrogen bond between the oxygen atom of the phosphate backbone of one strand and the hydrogen atom of the amino group of the opposite adenine on the other strand (figure 4.13.b). Interestingly, the terminal unpaired nucleotides of the helix were pairing with the symmetry-related unpaired nucleotide thus forming inter-duplex base pairing which resulted in a continuous double helical structure (figure 4.13.a).

Unfortunately, comparing our data with previously solved structures on the protein data base (PDB) led us to the knowledge that the structure of parallel double helix poly(A) was published shortly before we obtained our result (Safaee *et al.*, 2013) (PDB ID: 4JRD), after a high order helical Poly(A) RNA structure was just a prediction for more than 50 years.

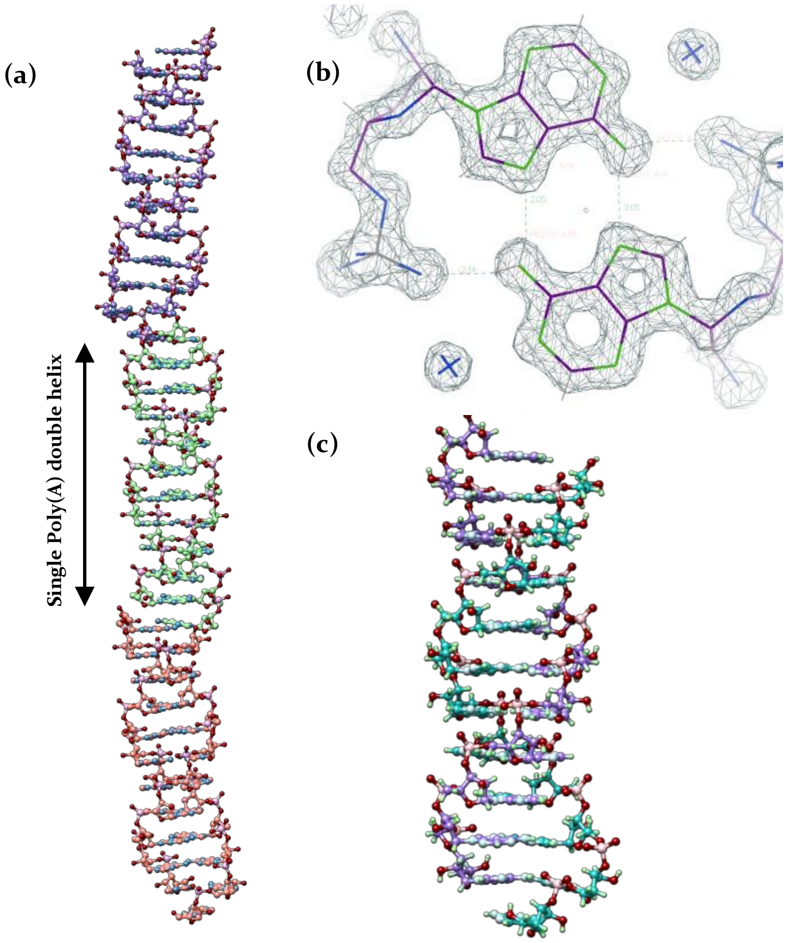


Figure 4.13. Poly(A) double helix crystal structure. (a) The continuous Poly(A) double helix formed by three individual helices shown in different colours. **(b)** 2Fo-Fc map countered at 1.20 δ showing the electron density surrounding the A-A base pairing responsible for the double helix formation with their formed hydrogen bonds formed between the amino groups and Nitrogen atome of symmetrical adenines on one hand and between the oxygen group of the phosphate backbone and the amino group of the opposite adenine on the other hand. The environmental distances are also shown. **(c)** An overview of the symmetrical double-stranded Poly(A) helix as present in the asymmetric unit of the crystal where one nucleotide is overhanging on each end. Each strand in the double-helix is represented in distinct colour.

Resolution (Å)	1.25
Space group	$P4_12_12$
Unit-cell parameters	
a,b,c (Å)	22.89, 22.89, 163.673
α, β, γ (°)	90, 90, 90
Wavelength (Å)	0.97949
Total reflections	129468 (3230)
Unique reflections	12847 (1059)
Completeness (%)	98.04 (84.45)

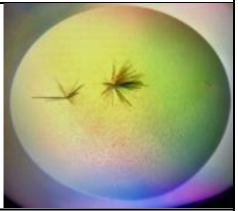
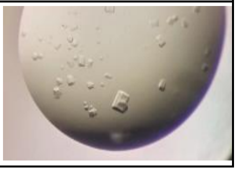
Multiplicity	10.1 (3.1)
Mean I/Sigma(I)	14.82 (2.29)
R_{merge}	0.1031 (0.6267)
R_{meas}	0.1083
R_{work}/R_{free}	0.1448/0.1957
CC1/2	0.999(0.72)
Water molecules	65
Clashscore	0
Average B factors	11.20

Table 4.1. Data collection and refinement statistics from a crystal of Poly(A) double helix structure. The refinement statistics shown are obtained after two rounds of refinement using *Phenix.refine* module. Statistics from the highest resolution shell are indicated in parenthesis.

4.4. Crystal structure of Mip6 $RRM3_{(313-389)}$

4.4.1. Crystallization

Crystals of purified Mip6 $RRM3_{(313-389)}$ started to appear few days after setting the crystallization screening plates at 21°C. Different forms of crystals were obtained in different growth conditions. Some crystals diffracted poorly to 8 Å, but other crystals in forms of needles diffracted to a better resolution. The conditions where clusters of needles grew were used for additive screening, where single crystals were consequently obtained. Additionally, co-crystallizations with Poly(U) RNA was also tried. Table 4.2 summarizes the crystallization conditions where the crystals were obtained with the corresponding resolution obtained when X-ray diffracted.

Crystallization condition	Screen	Additive	Resolution	Crystals form
1.6M Sodium citrate tribasic dihydrate pH 6.5	Crystal screen I and II (CSIYII) (Hampton)		2-2.5 Å	
3 M Sodium formate	JBSII (Jena Bioscience)		8 Å	

1.6M Sodium citrate tribasic dihydrate pH 6.5	CSIYII (Hampton)	2M Ammonium Nitrate	1.3 Å	
0.2M Lithium sulfate, 0.1M Tris Ph 8.5, 40% PEG 400	JCSG (molecular dimensions)		1.25 Å	

Table 4.2. Mip6 RRM3$_{(313-389)}$ crystals growth conditions where different crystal forms with varying x-ray diffraction resolution were obtained.

4.4.2. Data collection and processing

Subsequently, complete data sets were collected from the highest diffracting crystals without RNA at 1.3 Å and with added Poly(U) RNA at 1.25Å were used for scaling and further processing in the aim to solve the structure of Mip6 RRM3$_{(313-489)}$. The crystals belonged to the tetragonal $P4_32_12$ and monoclinic $P12_11$ space groups respectively. Due to the prior characterization of Mip6 RRM3$_{(313-389)}$ as a RNA recognition motif (RRM), an RRM model from the PDB (PDB ID: 3FEY) was initially used for molecular replacement using MR *Phaser* from *Phenix*. Unit cell dimensions from both datasets belonging to different space groups indicated the presence of one molecule of Mip6 RRM3$_{(313-389)}$ in the asymmetric unit with no RNA in the structure in the case where RNA was added. The structures from both crystals were further refined using *Phenix.refine*. The final models were 97 and 99% Ramachandran favoured with no Ramachandran outliers.

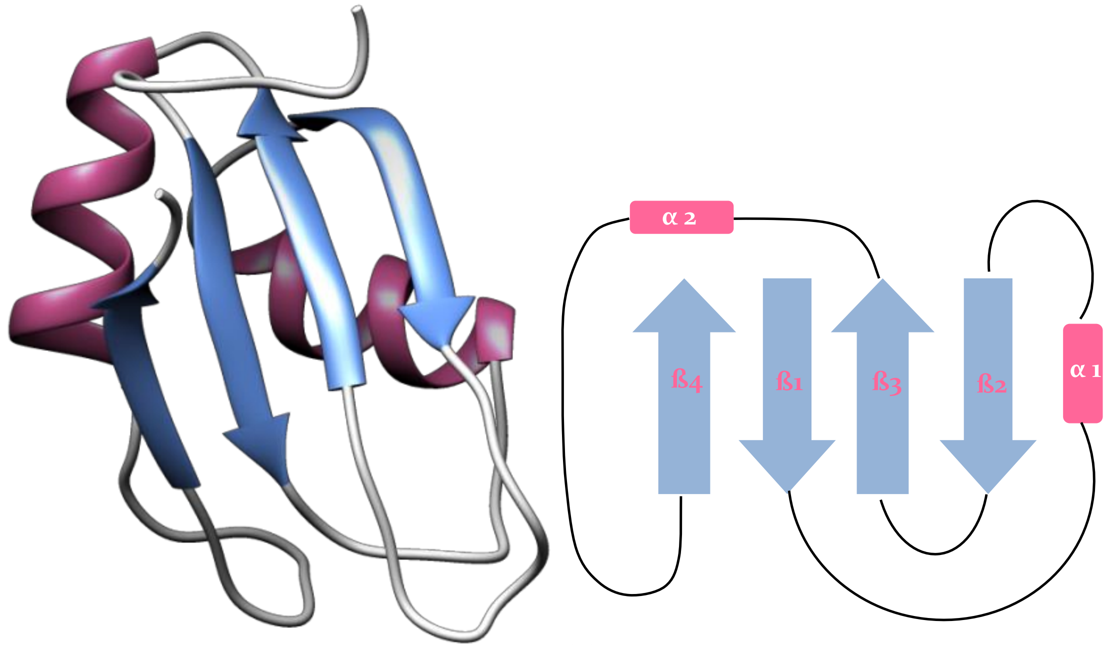
	Crystal 1	Crystal 2 (RNA added)
Resolution (Å)	1.3	1.25
Space group	$P4_32_12$	$P12_11$
Unit-cell parameters		
a,b,c (Å)	36.310, 36.310, 91.816	25.842, 47.023, 30.563
α, β, γ (°)	90, 90, 90	90, 109.9, 90
Wavelength (Å)	0.97934	0.9795
Total reflections	178972 (6700)	56990 (350)

Unique reflections	15749 (1444)	14119 (206)
Completeness (%)	99.29	74.01
Multiplicity	11.4 (4.6)	4.0 (1.7)
Mean I/Sigma(I)	26.38 (2.74)	25.88 (3.16)
R_{merge}	0.04538 (0.4646)	0.03091 (0.2076)
R_{meas}	0.04746	0.03519
R_{work}/R_{free}	0.1514/0.1818	0.1371/0.1771
CC1/2	0.999(0.919)	0.999(0.924)
Ligands	22	13
Water molecules	51	99
Protein residues	86	81
Ramachandran plot		
Favoured (%)	99	97
Outliers (%)	0	0
Clashscore	0	0
Average B factors	24.60	14.80
PDB code	5D77	5D78

Table 4.3. Data collection and refinement statistics from two crystals of Mip6 RRM3$_{(313\text{-}389)}$. Statistics from the highest resolution shell are indicated in parenthesis.

4.4.3. Mip6 RRM3$_{(313\text{-}389)}$ structures reveal different loop conformations

The high resolution structures of Mip6 RRM3$_{(313\text{-}389)}$ showed a typical canonical RRM fold of βαββαβ with four stranded antiparallel β sheets, and two alpha helices packed against the β-sheet (figure 4.14). The β sheets formed a hydrophobic core especially β1 and β3 that were entirely hydrophobic. These two central β sheets, β3 and β1 are known to be involved in the hydrophobic interaction with nucleic acids in other RRMs. Indeed, analyzing the electrostatic potential of the domain, the surface of one side of Mip6 RRM3 structure showed a positive charged surface. This indicates a possible site of RNA interaction as we already have shown to exist in vitro (figure 4.15).



(a) **(b)**

Figure 4.14.Crystal structure of Mip6 RRM3$_{(313-389)}$. (a) An overview of Mip6 RRM3$_{(313-389)}$ structure we crystallized. It consists of two alpha helices (shown in pink), and four anti-parallel beta sheets (shown in blue). **(b)** A schematic representation of the secondary structure arrangement in the Mip6 RRM3 structure.

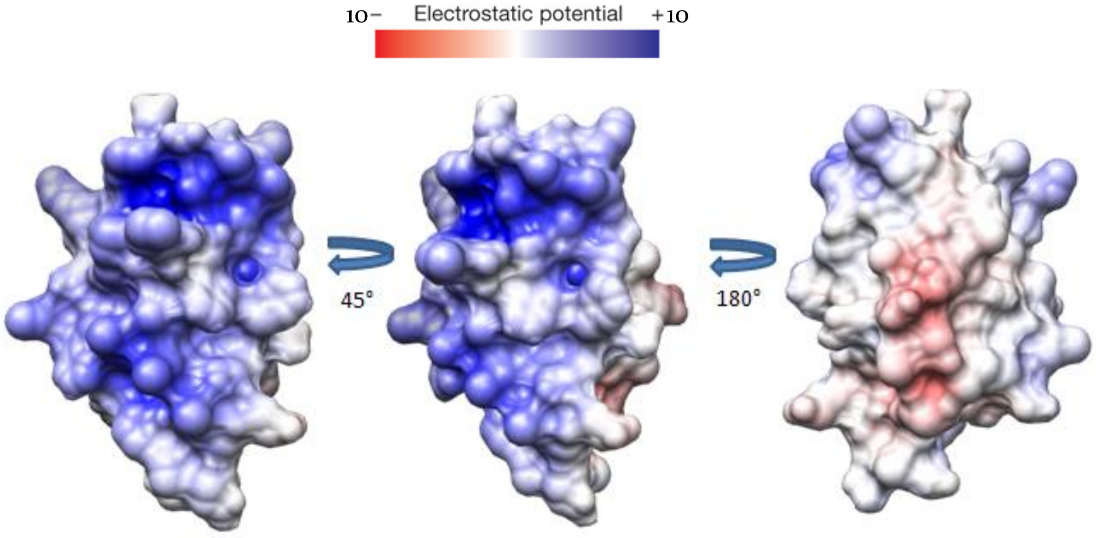


Figure 4.15. Positively charged surface of Mip6 RRM3. Electrostatic potential calculated with APBS software and visualized using *Chimera* software shows the positively charged surface of Mip6 RRM3 domain in blue, which indicates a possible site for binding nucleic acids. The negatively charged areas are depicted in red.

Superposing the two structures of Mip6 RRM3 obtained from two different datasets of crystals obtained from different conditions showed two different loop conformations. The two structures were essentially similar with R.M.S.D of 1.42 Å with 76 aligned residues, but the obvious difference was the flexibility and movement of loop 3 (β2-β3 loop) primarily, and to a lower extent in loop 5 (α2-β4 loop) as shown in (figure 4.16). It is worth mentioning here that loop3 definitely had higher B-factors than the rest of the structure which is probably due to the flexibility and movement of that loop. The flexibility of the loop might be due to the change of the position of this loop in the free form or bound to different nucleic acid sequences.

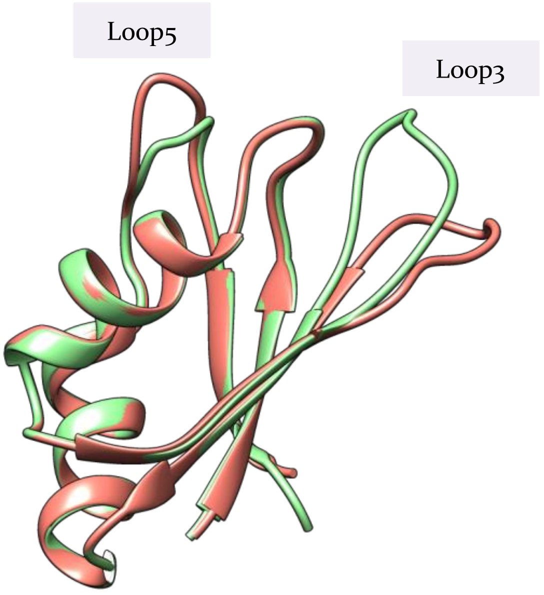


Figure 4.16. Mip6 RRM3 crystal structure showed two different structural conformations. Superposing the two structures obtained from Mip6 RRM3 domain showed different conformation in loop3 (linking ß2 and ß3) and in loop 5 (linking α2 to ß4) to a lower degree.

By comparing the Mip6 RRM3 structures we obtained to that of RRM1 from Sex lethal protein (Sxl) of *Drosophila melanogaster* bound to RNA (PDB code: 1B7F), which the advanced sequence search on the PDB identified as having high sequence similarity with Mip6 RRM3, we could see an overall conservation on the protein folding of the RRM excluding loop 3 (β2-β3 loop) that showed high degree of movement and also in loop 5 (α2-β4 loop) in lower degree. These loops are in some RRMs involved in the recognition of RNA, in addition to the major role of ß1 and ß3 that typically forms the platform for RNA binding like seen in the Sxl RRM1/RNA crystal structure as shown in figure 4.17.

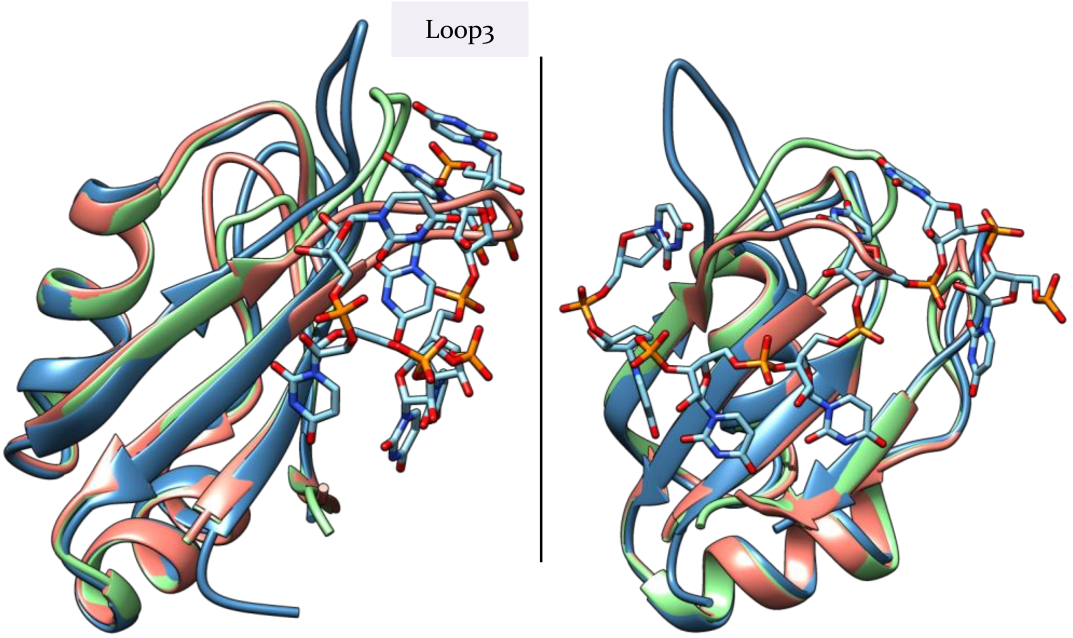


Figure 4.17. Conformational movement in Mip6 RRM3 loops. Superposition of Mip6 RRM3 two structures (represented in salmon and green colours) with RRM1 of Sxl protein (shown in blue) bound to RNA (PDB code: 1B7F). The comparison revealed general domain homology, but major flexibility in loop 3 and to a lesser extend in loop5. As visible in the figure, ß2 and ß3 of Sxl RRM1 are majorly involved in the RNA binding as well as loop 3.

4.5. Characterizing the interaction between Mex67 and Mip6

A previous study (Segref *et al.*, 1997), where a yeast two hybrid experiment was performed, suggested the possible involvement of the C-terminal domain of Mex67 in the interaction with Mip6. To verify these claims, we designed and primarily worked with a construct that includes the C-terminal domain of Mex67 due to the fact that the full length Mex67 protein was big in size and tricky to get large amounts of soluble protein from in bacterial cells.

4.5.1. Expression and purification of Mex67 C-term

GST-tagged Mex67 C-term $_{(528\text{-}599)}$ of interest was successfully expressed in *E. coli* using Auto-induction as an expression system. A good amount of soluble protein was obtained after the purification of the expressed protein on GSTrap. The fused GST tag was then digested using PreScission protease, and the obtained protein was further purified using size exclusion chromatography (Figure 4.18). In gel filtration, the protein eluted as a single peak where its corresponding protein fractions were pure as visualized in acrylamide gel. Good amounts of the protein were obtained and were concentrated to a final concentration of 30 mg/ml then flash frozen in liquid nitrogen and stored at -80°C for further use.

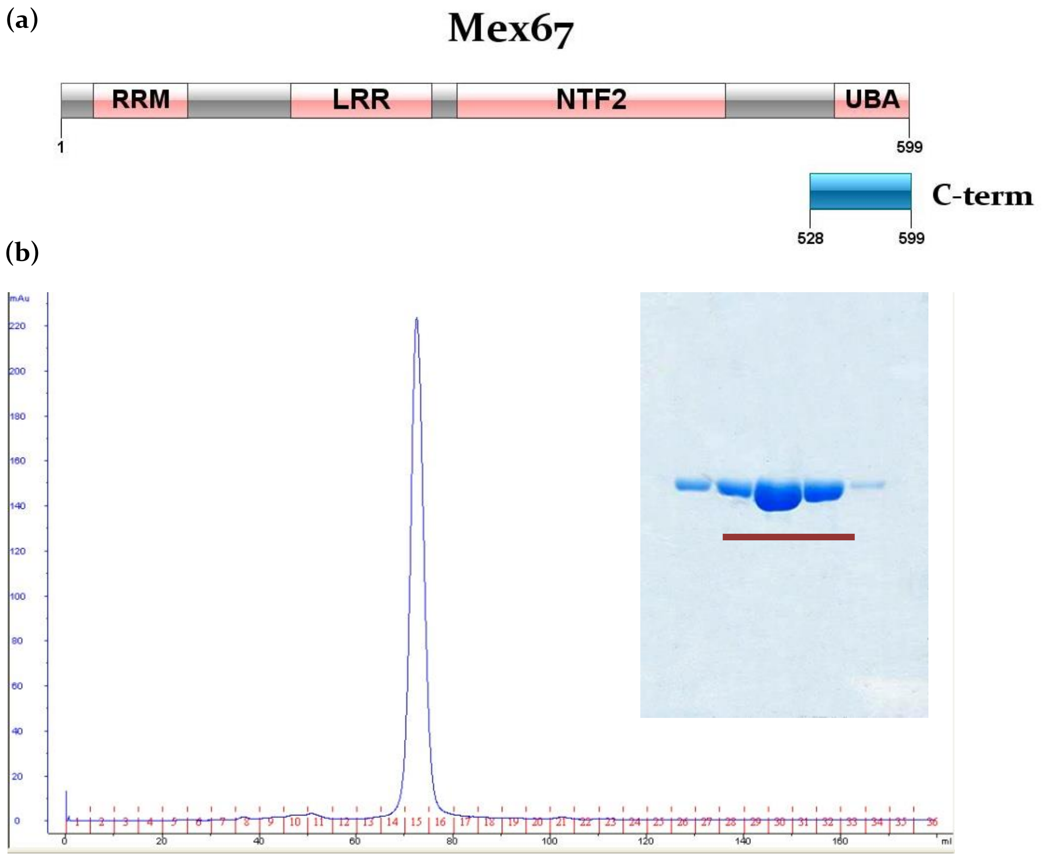


Figure 4.18. Purification of Mex67 C-term $_{(528-599)}$ using size exclusion chromatography. (a) A scheme depicting the full length Mex67 protein and its C-terminal domain boundaries used in this study. **(b)** Chromatogram obtained from Superdex 75 16/60 shows the elution profile of the Mex67 C-term$_{(528-599)}$ as a single peak at ~72.4 ml. the corresponding fractions were visualized on a 10% SDS-PAGE gel highlights the purely obtained protein marked with the red bar.

4.5.2. Mex67 binds Mip6 *in vitro* with high affinity

To assess the interaction between Mex67 and Mip6 *in vitro*, Mip6$_{(111-480)}$ was tested for the binding of the C-terminal domain of

Mex67$_{(528-599)}$ where both protein were expressed and purified as mentioned earlier and Isothermal Titration Calorimetry (ITC) was used to assess the interaction. 56 µM of purified Mip6$_{(111-480)}$ (section 4.2.1) was used in sample cell and 800µM of purified Mex67 C-term$_{(528-599)}$ (section 4.5.1) as a ligand. As shown in figure4.19, our results showed an exothermic reaction between the two proteins and a good binding profile with a calculated affinity of K_D= 1.067 µM when Mex67$_{(528-599)}$ was titrated into Mip6$_{(111-480)}$ thus confirming the protein-protein interaction and complex formation between both proteins. Other interaction parameters were also determined with ΔH (Kj/mol) of -29.85 and ΔS (J/mol-k) of 10.74 were given.

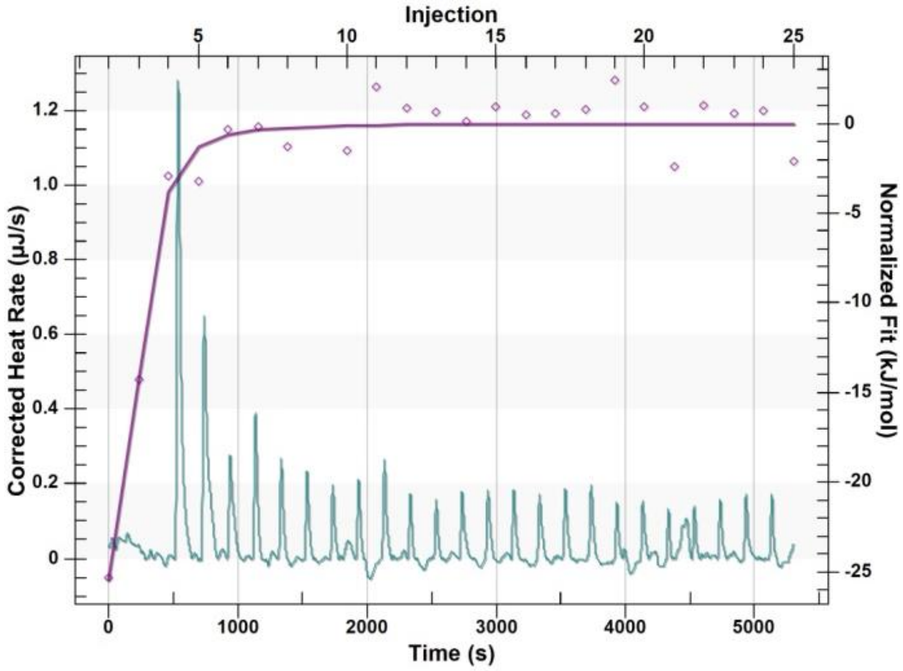


Figure 4.19. . Mex67$_{(528-599)}$ interacts with Mip6. ITC experiment with 800 µM of Mex67 C-term$_{(528-599)}$ titrated into 56 µM Mip6$_{(111-480)}$ showed micro molar binding affinity. Raw heat peaks from injections are shown consecutively in pink, in addition to the curve obtained after integration and fitting of data obtained over the time course of the experiment. A binding affinity of 1.067 µM was calculated. Nanoanalyze software was used to fit the data and generate the figure.

4.5.3. Mex67 interacts with Mip6 through its RRM4

Going a step further to narrow down the possible sites of Mex67 binding on Mip6 and define the interaction boundaries, a series of subsequent deletions of Mip6, as shown in the figure 4.20 below, were generated and tested for Mex67 binding.

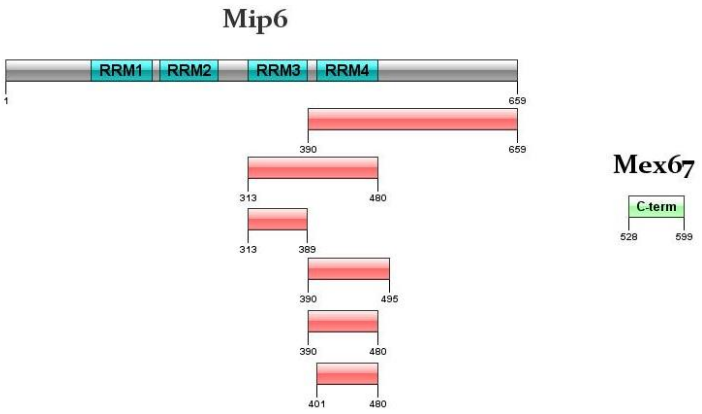


Figure 4.20. A scheme depicting the full length Mip6 protein and the generated protein truncations used to map the minimal boundaries for the site of binding of Mex67 C-terminal domain$_{(528\text{-}599)}$ on Mip6.

We commenced with small scale purification of MBP tagged Mip6 construct comprising the whole C-term of Mip6 (MBP-Mip6$_{390\text{-}659}$) and GST tagged Mex67 C-term$_{(528\text{-}599)}$ co-tranformed in Bl21 codon plus bacterial cells and expressed at 20°C overnight. The small scale purification was carried out as indicatd in section 3.4.1 on either glutathione sepharose beads to capture the GST-Mex67$_{(528\text{-}599)}$ or amylose resin to capture the MBP-Mip6$_{(390\text{-}659)}$. The bound proteins were seperated from the beads by the addition of 6x loading buffer and boiling for 10 minutes at 95°C. The samples were then visualized on a 10% SDS-PAGE gel (figure 4.21). The two proteins eluted together from either beads indicating binding, although noticibly the Mip6$_{(390\text{-}659)}$ showed signs of degradation and also a lower band appeared of the protein appeared on the SDS-PAGE probably due to cleavage (proteolysis).

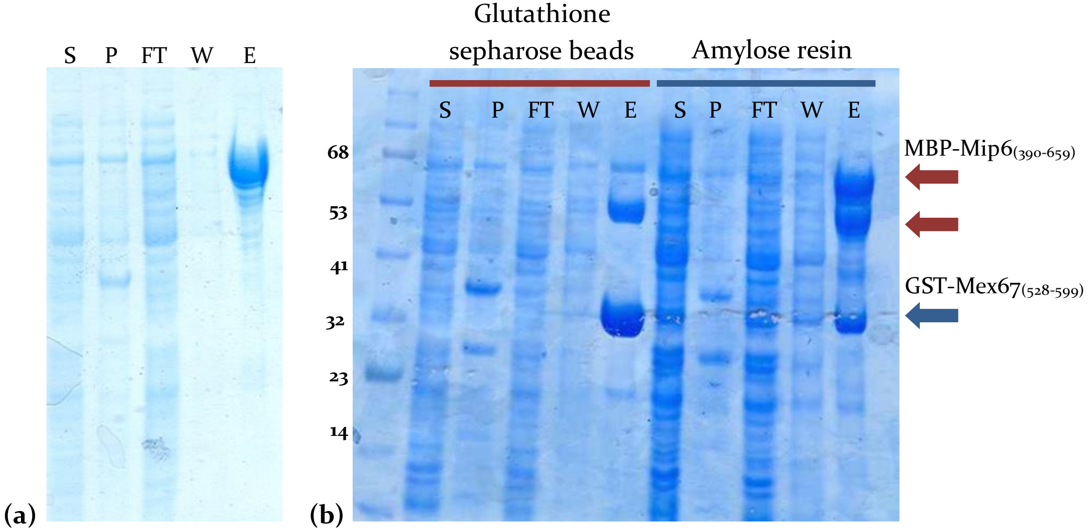


Figure 4.21. Small scale purification of co-expressed Mip6$_{(390-659)}$ and Mex67$_{(528-599)}$. (a) Control small scale purification of MBP tagged Mip6$_{(390-659)}$ on amylose resin shows the band corresponding to the protein at the expected size. **(b)** Small scale purification of the co-expressed protein on glutathione sepharose beads and amylose resin respectively showing the co-elution of both proteins which indicates interaction. A possible cleavage in the Mip6$_{(390-659)}$ protein is visible (red arrows). Aliquots from the different step of the pull down assay of Supernatants (S), Pellets (P), non-bound fraction (FT), wash of the beads (W), and finally elution of the protein bound to the beads (E) were visualized on 10% SDS-PAGE.

On the other hand, a purified Mip6 RRM3/4$_{(313-480)}$ that comprised both RRM3 and RRM4 (section 4.3.1) was used to determine whether it interacts with Mex67$_{(528-599)}$ or not. For this purpose, the Bio-layer interferometry system where 80 µg/ml of GST-tagged Mex67 C-term$_{(528-599)}$ immobilized on anti-GST biosensors was used as a bait, while an increasing concentrations of Mip6 RRM3/4$_{(313-480)}$ were used to establish the association and dissociation steps. A good binding profile was observed between the two proteins enabled us to calculate the binding affinity between both proteins (figure 4.22). A calculated micro molar binding affinity of K_D= 1.42 µM was obtained, which indicates high affinity between both proteins. An addition of 0.5 mg/ml of BSA was used in the buffer used in this experiment to eliminate any non-specific binding of the Mip6 RRM3/4$_{(313-480)}$ protein to the anti-GST sensors.

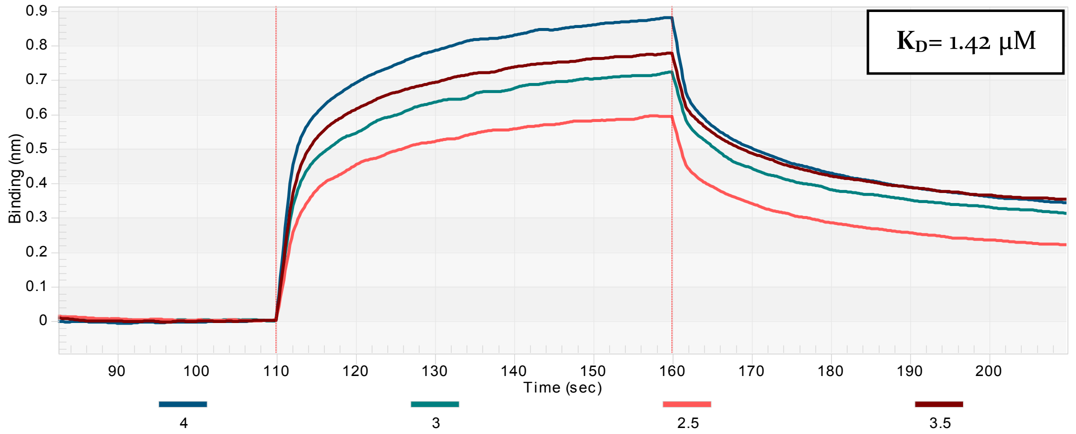


Figure 4.22. Mip6 RRM3/4$_{(313\text{-}480)}$ binds Mex67 C-term$_{(528\text{-}599)}$ with high affinity. Association and dissociation steps of Bio-layer interferometry using increasing concentrations of Mip6 RRM3/4$_{(313\text{-}480)}$ depicted in distinctive colours shows the interaction profile between the two proteins. The calculated global binding affinity calculated is shown.

Taking these attained results into account, it started to seem relevant to think that RRM4 of Mip6 is the domain with the minimal boundaries required for the interaction with Mex67, especially that Mip6 RRM3$_{(313\text{-}389)}$ did not co-elute with Mex67$_{(528\text{-}599)}$ from gel filtration after both proteins were mixed in equimolar concentration and injected in size exclusion chromatography (figure 4.23) which indicates that the two proteins were not holding together. To prove this hypothesis we decided to go a step further to investigate whether truly Mip6 RRM4 alone was sufficient for the formation of the complex. For this purpose, different constructs of MiP6 RRM4 with different boundaries were investigated: Mip6 RRM4$_{(390\text{-}495)}$; Mip6 RRM4$_{(390\text{-}480)}$; Mip6 RRM4$_{(401\text{-}480)}$ respectively (section 4.3.3).

A small scale pull down assay was carried out on glutathione sepharose beads, using GST-Mex67 C-term$_{(528\text{-}599)}$ as a bait protein while His tagged Mip6 RRM4$_{(390\text{-}495)}$ and Mip6 RRM4$_{(390\text{-}480)}$ were used as prey proteins. The beads were washed thoroughly four times to remove any non-bound protein, before retaining the protein in the beads by adding 6x sample buffer and heating the samples for 10 minutes at 95°C. Both Mip6 RRM4$_{(390\text{-}495)}$ and Mip6 RRM4$_{(390\text{-}480)}$ eluted together from the beads with the GST-Mex67 C-term$_{(528\text{-}599)}$ showing interaction (figure 4.23). Subsequent large scale purifications confirmed the results on a larger scale.

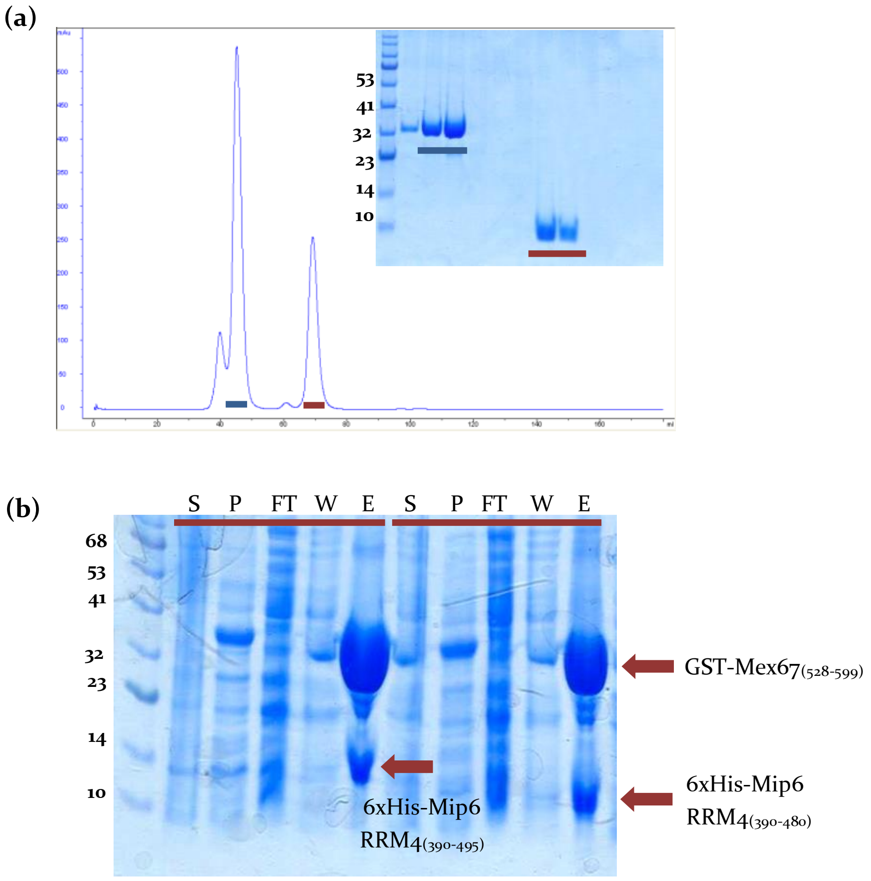
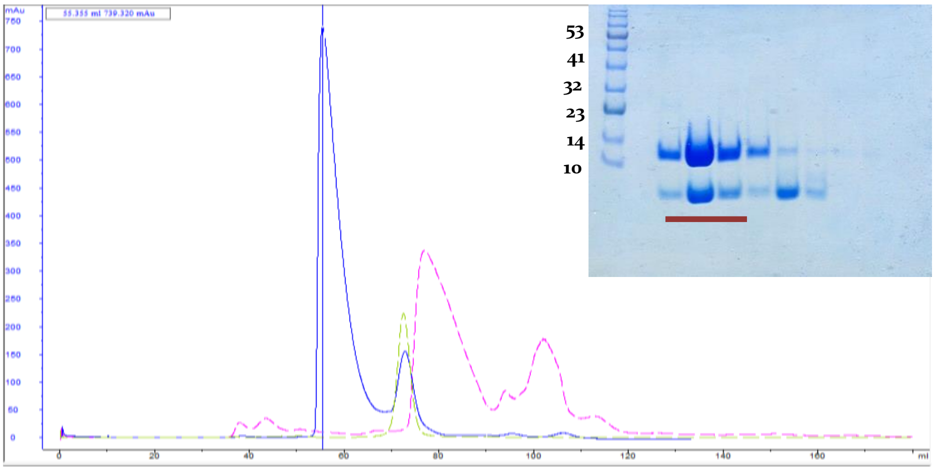
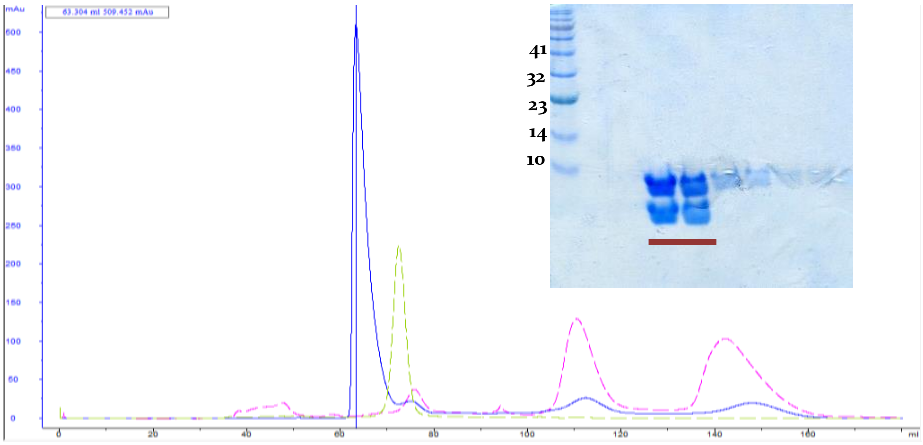


Figure 4.23. (a) No interaction between Mex67$_{(528-599)}$ and Mip6 RRM3$_{(313-389)}$. Performing size exclusion chromatography with pre-incubated equimolar concentration of Mex67$_{(528-599)}$ and Mip6 RRM3$_{(313-389)}$ using Superdex 75 16/60 column showed that the proteins did not hold together, instead each protein eluted in a separate peak, GST-Mex67$_{(528-599)}$ (blue) and Mip6 RRM3$_{(313-389)}$ (red)as visualized on both the chromatogram and the 10% SDS-PAGE.
(b) Pull down assay shows co-elution of Mex67$_{(528-599)}$ with Mip6 RRM4. 10% SDS-PAGE gel representing the different pull down assay steps: supernatant (S), pellet (P), non-bound(FT), wash (W), and Elution (E), done one glutathione sepharose beads using GST tagged Mex67$_{(528-599)}$ and His tagged Mip6 RRM4$_{(390-495)}$ and Mip6 RRM4$_{(390-480)}$ respectively showed the co-elution of both proteins together in the elution fraction.

Subsequently, each of the Mip6 RRM4 constructs: Mip6 RRM4$_{(390-495)}$, Mip6 RRM4$_{(390-480)}$, and Mip6 RRM4$_{(401-480)}$ was over-expressed and purified as described earlier (section 4.3.3), then mixed with equimolar concentration with a purified Mex67 C-term$_{(528-599)}$ (section 4.5.1) and injected in a gel filtration column Superdex 75 16/60. The formation of a stable complexes even with the shortest deletion construct Mip6 RRM4$_{(401\text{-}480)}$ was confirmed as the two proteins co-eluted from the gel filtration column in a single peak corresponding to a higher molecular weight than that of the elution profile of each protein alone. A shift in the peak when alone or in complex was discrete. This was confirmed by visualizing the gel filtration fractions on a 10% SDS-PAGE gel (figure 4.24) and consistent with a stable complex formation with the minimal Mip6 RRM4 boundaries utilized.



(a)



(b)

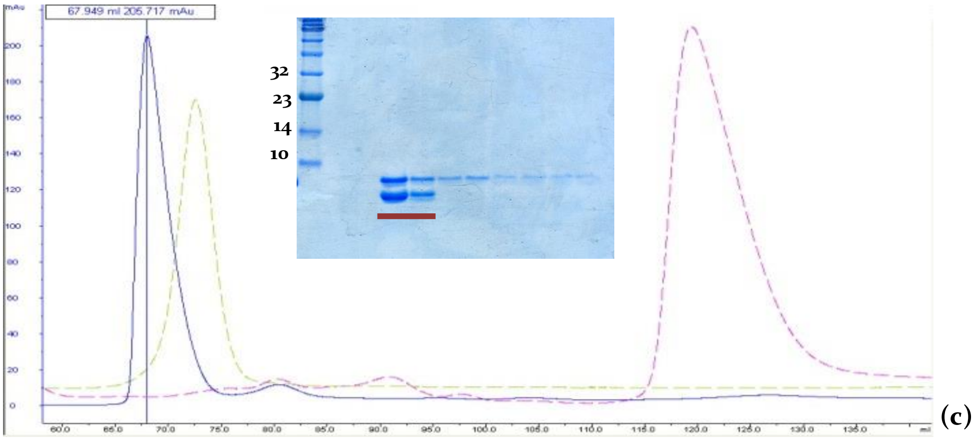


Figure 4.24. Stable complex formation between Mex67$_{(528\text{-}599)}$ and Mip6 RRM4 in gel filtration. Superposition of chromatograms representing the elution profile of the complex formed between Mex67$_{(528\text{-}599)}$ and several truncations of Mip6 RRM4: 390-495 (**a**), 390-480 (**b**), and 401-480 (**c**) from 16/60 Superdex 75 gel filtration column with the elution profile previously obtained of each of the proteins separately .The elution peak of the complex is demonstrated in blue and the corresponding fractions visualized on 10% SDS-PAGE gel are marked with a red bar. On the other hand, the light green peak corresponds to the separate elution peak of Mex67 C-term$_{(528\text{-}599)}$, while the pink peak corresponds to the elution profile of the truncated Mip6 RRM4 construct.

Furthermore, ITC biophysical technique was used in order to quantify the interaction between Mex67$_{(528\text{-}599)}$ and Mip6 RRM4$_{(390\text{-}480)}$ and determine the kinetics of the interaction. Both proteins tags were removed and both proteins were dialyzed overnight in the same buffer. 690 µM of Mex67$_{(528\text{-}599)}$ was used as a titrant in the syringe, while 50 µM or 70 µM of Mip6 RRM4$_{(390\text{-}480)}$ in the sample cell. The experiment was carried out at 15°C. The analysis and fitting of the resulting data from each injection, calculated a good binding affinity in the low micro molar range with K_D =1.717 µM with a stoichiometric ratio n= 0.965 indicating a probable 1:1 binding ratio (figure 4.25). However, the calculated stoichiometric ratio was slightly lower when higher concentration of Mip6$_{(390\text{-}480)}$ was used (n=0.72 or n=0.8 at 70µM of concentration) . Interestingly, the binding affinity obtained is almost identical to those obtained between Mex67$_{(528\text{-}599)}$ and Mip6$_{(111\text{-}480)}$ and Mip6$_{(313\text{-}480)}$ which clearly suggests that Mip6 RRM4 boundaries are sufficient for the interaction and complex formation.

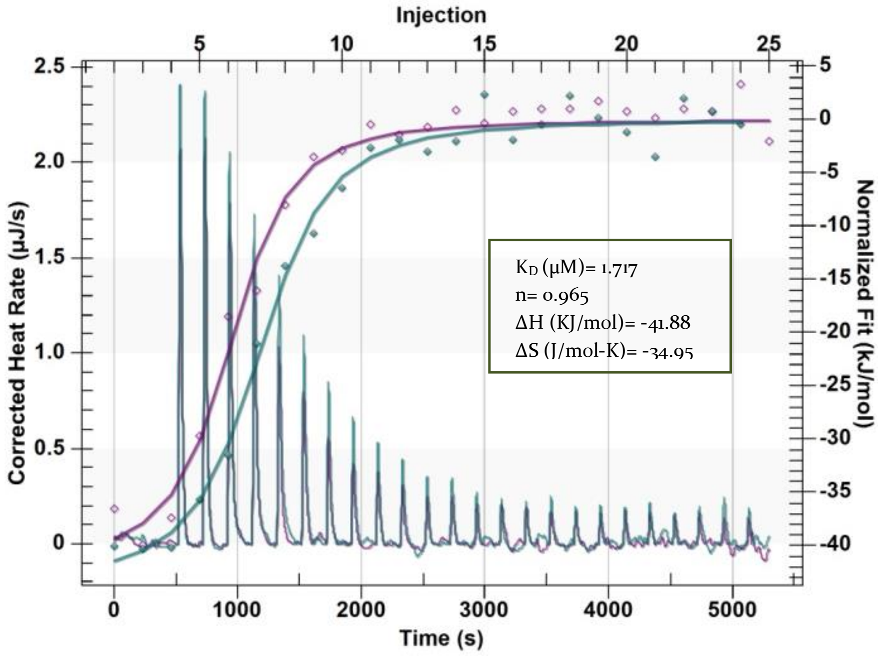


Figure 4.25. Mex67$_{(528-599)}$) binds Mip6 RRM4$_{(390-480)}$ with high affinity. Overlapping results from two ITC experiments with two different concentrations of Mip6 RRM4$_{(390-480)}$ in the sample cell (70 and 50 µM consecutively). Raw heat peaks from the injections are shown consecutively in pink (50 µM) and cyan (70 µM), in addition to the fitting curves obtained that allowed the calculation of a binding affinity of 1.717 µM and 2.21 µM respectively. The thermodynamic parameters for the experiment with 50 µM Mip6 RRM4$_{(390-480)}$ are shown on the graph. The Nanoanalyze software was used to fit the data and generate the figure.

4.6. Crystal structure of Mex67 C-term$_{(528-599)}$

4.6.1. Crystallization and data processing

The stably formed complexes between Mex67 C-term$_{(528-599)}$ and Mip6 RRM4, obtained from the previously indicated size exclusion chromatography (figure 4.24) were used to set crystallization screening plates in an attempt to get a crystal structure of the complex. Other attempts were carried out through forming the complexes with different Mip6 constructs by mixing directly before setting the screening plates. Also, limited in situ proteolysis of the complex using different enzymes trypsin, chymotrypsin, or subtilisin was carried out in the attempt to obtain a more stable complex by digesting any flexible regions. During our attempts to crystallize the Mex67$_{(528-599)}$/Mip6RRM4 complex, few crystals containing only Mex67 C-terminal domain were obtained, but none comprising the complex.

Crystals containing Mex67$_{(528-599)}$ appeared at 21°C in a range of different growth conditions where some them are summarized in table 4.4 below.

Condition	Proteins in drop	Resolution (Å)
0.1M Bicine pH 9; 20% PEG 6000	Mex67 C-term$_{(528-599)}$	8
0.01M Nickel(II) chloride hexahydrate; 0.1M Tris pH 8.5; 1M Lithium sulfate monohydrate	Mex67 C-term$_{(528-599)}$/Mip6 RRM3/4	1.42
35% Tacsimate pH 7	Mex67 C-term$_{(528-599)}$/Mip6 RRM3/4	1.3
3M Sodium formate	Mex67 C-term$_{(528-599)}$/Mip6 RRM4$_{(390-495)}$	1.5
1.5M Ammonium sulfate; 0.1M Tris pH 8.5; 12% Glycerol	Mex67 C-term$_{(528-599)}$/Mip6 RRM4$_{(401-480)}$	1.5
0.1M MES monohydrate pH 6.5, 1.6M Magnesium sulfate heptahydrate	Mex67 C-term$_{(528-599)}$/Mip6 RRM4$_{(401-480)}$	-
1.5 M Ammonium sulfate, 0.1 M Tris pH 8.5, 12% glycerol	Mex67 C-term$_{(528-599)}$/Mip6 RRM4$_{(401-480)}$	-

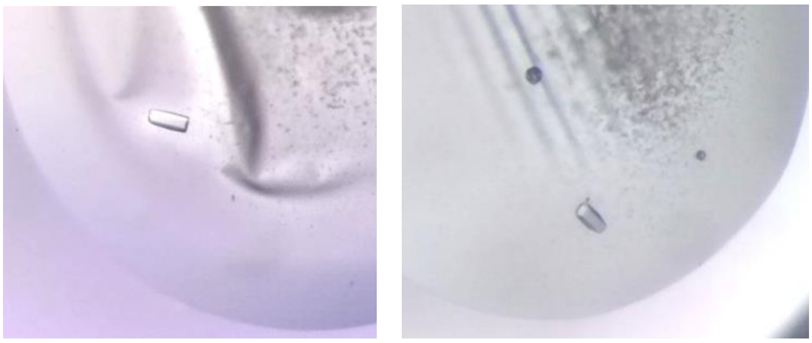


Table 4.4. Crystallization conditions in which crystals containing Mex67 C-term$_{(528-599)}$ were obtained and the resolution to which each crystal diffracted. Additionally, images of two obtained crystals of Mex67 C-term$_{(528-599)}$ are shown.

Complete data sets were collected from crystals that diffracted up to 1.42 and 1.3 Å (mentioned in table 4.4) and were further scaled and processed in order to solve the structure of Mex67$_{(528-599)}$. The crystals belonged to the hexagonal space groups P6$_5$ and P6$_1$ respectively. The molecular replacement for the first 1.42Å structure was performed with the MR *Phaser* module from *Phenix* initially using the PDB structure of the C-terminal UBA domain from the human TAP (Nxf1) protein, the human homologue of Mex67, as a model for molecular replacement (PDB code: 1OAI), the attained structure in turn served as a molecular replacement model for the later crystal. Good initial phases were obtained, and the structures obtained indicated one molecule of Mex67$_{(528\text{-}599)}$ in the asymmetric unit in both crystals. In the 1.3 Å, hydrogen atoms were added since resolution was high enough and map showing hydrogen densities was clear. Superposing both structures did not show any major differences or different conformations. The two structures were overlapping with an R.M.S.D. of 1.054 Å from the 67 identical residues with only a slight movement on the N-terminal part of the Mex67 UBA domain noted. Table 4.5 summarizes the data collection and final refinement statistics from both crystals of Mex67$_{(528\text{-}599)}$.

	Crystal 1	Crystal 2
Resolution (Å)	1.42	1.3
Space group	$P6_5$	$P6_1$
Unit-cell parameters		
a,b,c (Å)	37.65, 37.65, 74.745	71.42, 71.42, 20.34
α, β, γ (°)	90, 90, 120	90, 90, 120
Wavelength (Å)	0.97624	0.9795
Total reflections	296733(23933)	284363 (28535)
Unique reflections	11287 (1117)	14960 (1491)
Completeness (%)	99.47	100
Multiplicity	26.2 (20.1)	19 (19.1)
Mean I/Sigma(I)	17.56 (2.62)	25.5 (1.7)
R_{merge}	0.09804 (1.362)	0.04496(1.83)
R_{meas}	0.09997	0.04623
R_{work}/R_{free}	0.1856/0.2098	0.2098/0.2465
CC1/2	1(0.823)	1(0.616)
Ligands	6	3
Water molecules	29	31
Protein residues	68	68
Ramachandran plot		
Favoured (%)	99	98
Outliers (%)	1.5	0
Clashscore	4.51	2.74
Average B factors	26.40	24.50

Table 4.5. Data collection and refinement statistics from two crystals of Mex67 C-term$_{(528-599)}$. Statistics from the highest resolution shell are indicated in parenthesis.

4.6.2. Mex67 C-term structure made of alpha helices

The crystal structure of Mex67 C-term$_{(528-599)}$ included the Mex67 UBA domain plus an additional 17 residues on its N-terminus. The structure is made of six alpha helices connected with short loops, in which four of these helices formed the Mex67 UBA domain. The N-terminal part is formed of one two turn helix and a shorter one turn helix, while the UBA domain is formed by three complete helices in addition to a C-terminal small one turn fourth helix that is said to undergo conformational changes when UBA binds its partners stabilizing the interaction (Hobeika *et al.*, 2009) (figure 4.26).

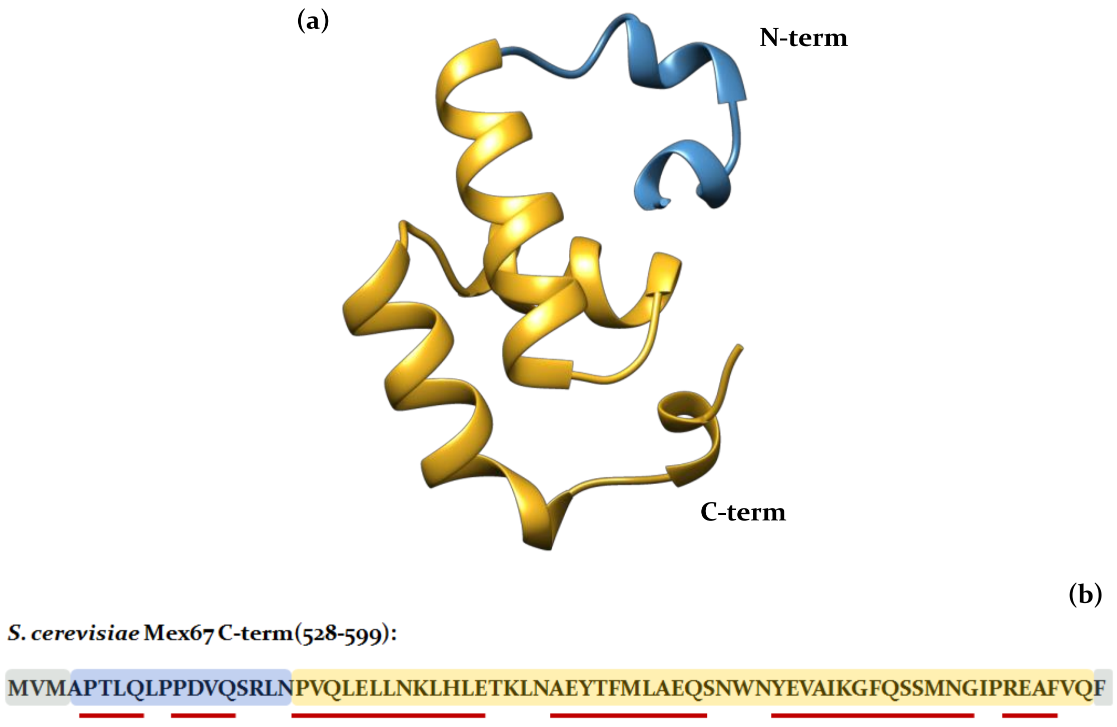
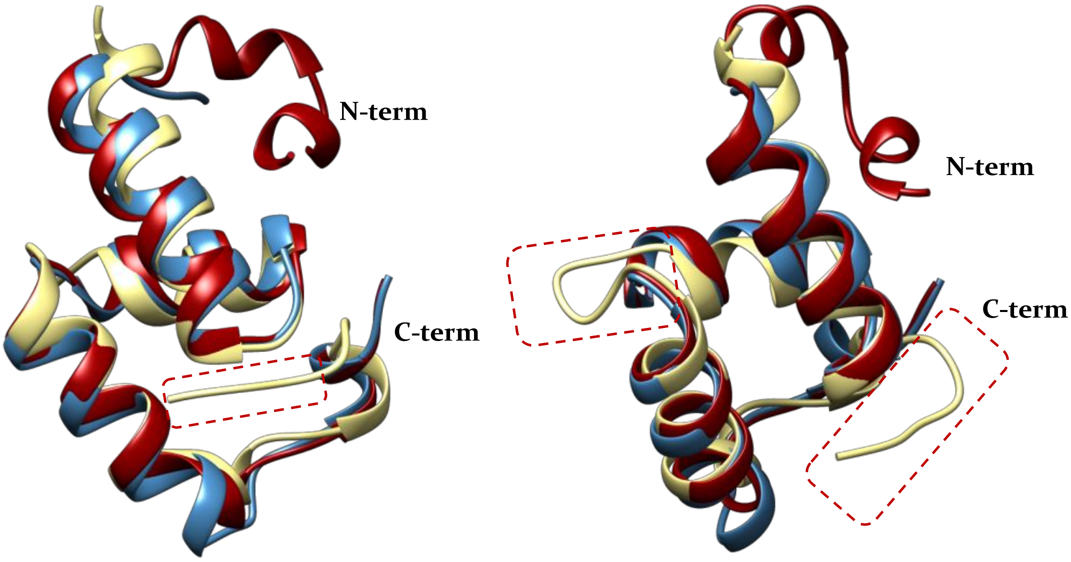

(a)
N-term
C-term
(b)
S. cerevisiae Mex67 C-term(528-599):
MVMAPTLQLPPDVQSRLNPVQLELLNKLHLETKLNAEYTFMLAEQSNWNYEVAIKGFQSSMNGIPREAFVQF

Figure 4.26. Crystal structure of Mex67 C-term$_{(528-599)}$. (a) A general overview of the Mex67 C-terminal domain structure that crystallized. It is made of six alpha helices linked by short linkers. The alpha helices that form the Mex67 UBA domain are shown in yellow while the additional N-term in the structure formed by two alpha helices are seen in blue. **(b)** A scheme of the sequence of Mex67 C-terminal domain$_{(528-599)}$ we crystallized representing the different characteristics of the protein mentioned in (a), the first three residues and the last one shown in grey are not visible in the structure.

4.6.3. Evolutionary conservation of the Mex67 C-term UBA domain

The C-terminal domain of yeast Mex67 structure was then compared to the PDB X-ray structures of the UBA domain alone from the homologues of Mex67 in *Homo sapiens* (PDB code: 1OAI) and *Chaetomium thermophilum* (PDB code: 4WP2). Regardless of the weak conservation in the sequence of the UBA domain between the three species, the structural alignment showed a high structural conservation (figure 4.27). Using TM-align (Yang Zhang & Skolnick, 2005) as a calculation tool for structural comparison, Mex67 UBA domain structure from *S. cerevisiae* presented the highest conservation to that of *H. sapiens* with R.M.S.D of 0.89 Å from 57 aligned residues, while it revealed a higher variation and lower conservation with that of *C. thermophilum* with R.M.S.D of 2.08 with 55 aligned resides. The UBA domain from the *C. thermophilum* homologue presented a C-terminal extension that doesn´t exist both in yeast and human Mex67 UBA, and a slightly longer first helix with a longer protruding loop connecting helix 2 and helix 3 of the domain (known to play an important role in the interaction with the FG repeats of nucleoporins) (figure 4.27). The reason to these differences in the *C. thermophilum* structure is still not clear although it might has a role in the increased thermo-stability known for proteins from *C. thermophilum* .

(a)



(b)

```
H.sapiens       -PEQQEMLQAFSTQSGMNLEWSQKCLQD-NNWDYTRSAQAFTHLKAKGEIPEVAFMK----
S.cerevisiae    -PVQLELLNKLHLETKLNAEY-TFMLAEQSNWNYEVAIKGFQSS--MNGIPREAFVQF---
C.thermophilum  NATKAQLIAEVSRRTGMNVEXSQMCLTGAANWNLELALQSFEQQ--KANVPPEAFISQPQV
                   : :::  .  .: :* *     *    **:   : :.*        :*  **:.
```

Figure 4.27. Evolutionary structural conservation of Mex67 UBA domain. (a) Superposition between Mex67 C-term$_{(528\text{-}599)}$ structure from *S. cerevisiae* (shown in red), The UBA domain structure of its homologue from *H. sapiens* (shown in blue), and the UBA domain structure of the *C. thermophilum* homologue (shown in yellow). The C-terminal extension in the *C. thermophilum* UBA structure and the longer protruding loop between helix 2 and helix 3 are highlighted in the dotted red box. **(b)** The sequence alignment of Mex67 UBA domain across the three species highlighted with the colour corresponding to each structure.

4.6.4. Mex67 C-term$_{(528\text{-}599)}$ electrostatic surface is not positively charged

The electrostatic surface potential of Mex67 C-term$_{(528\text{-}599)}$ calculated using PDB2PQR & APBS software did not show any significant positively charged patches on the surface of the structure. On the contrary, the overall structure was neutral with one side of the structure showing a clear negatively charged pocket just above the FXFG and Hpr1 binding pocket (figure 4.28).

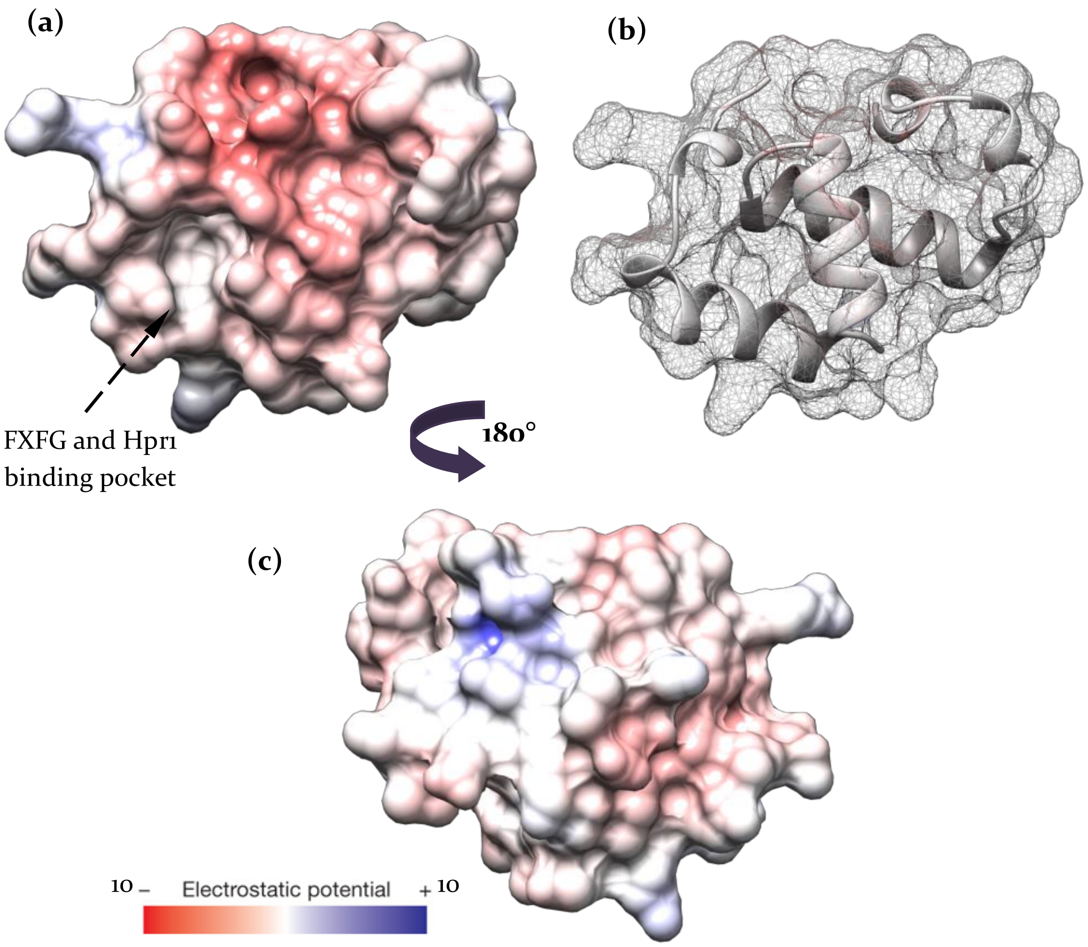


Figure 4.28. Mex67 C-term$_{(528\text{-}599)}$ surface is not positively charged. Surface Electrostatic Potential calculations visualized using *Chimera* shows the negatively electrostatic patches (in red) and the positively charged areas (in blue) on the Mex67 C-terminal domain surface **(a)** with the corresponding structural fold **(b)** and its other side surface when rotated 180°. Neutral patches are shown in white.

This led us to predict that the C-terminal domain of Mex67 did not have the capacity to bind nucleic acids on its own *in vitro*. To elaborate on this, 200 µg of purified Mex67 C-term$_{(528\text{-}599)}$ (section 4.5.1) were incubated with polyuridylic acid (PolyU) agaorse beads for 1 hour at 4°C with constant mixing. Later, the non bound fraction was removed, and the beads were washed three times before adding 6x sample buffer and boiling for 10 minutes at 95°C to separate from the beads any Poly(U) bound protein. The result showed that indeed Mex67 C-term$_{(528\text{-}599)}$ was not binding RNA *in vitro* since it did not co-purify with Poly (U)from the beads, instead it was seen in the non bound fraction (figure 4.29). Furthermore, Mex67$_{(528\text{-}599)}$ showed no binding to a biotinylated Poly (U) of 15 residues immobilized on streptavidin sensor using the Bio-layer interferometry system thus confirming the inability of Mex67 C-terminal region to bind RNA independently *in vitro* (figure 4.39).

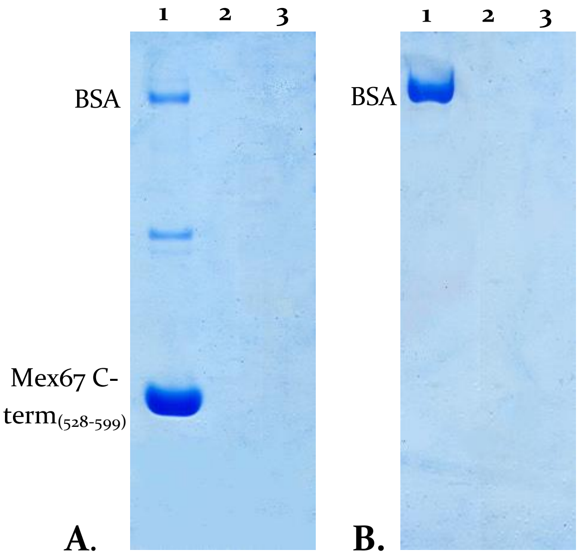


Figure 4.29. Mex67 c-term$_{(528-599)}$ does not bind RNA *in vitro*. A. 200 µg o Mex67$_{(528-599)}$ (1: input) were incubated with polyuridylic acid agarose beads for 1 h at 4°C, then washed three times (2: wash), and finally eluted (3: separation from beads). The samples were then visualized on a 10% SDS-PAGE gel and no Mex67$_{(528-599)}$ was not seen in the bound to the Poly(U) beads. **B.** BSA was used as a negative control.

4.7. Mex67 UBA domain is the binding site of Mip6

Additionally, to investigate whether the UBA domain of Mex67 required for proper mRNA export and previously shown to be the site of binding of the FG nucleoporins and Hpr1 is also the site of binding of Mip6, two additional constructs of Mex67 C-term were generated: Mex67 UBA$_{(545-599)}$ comprising exclusively the C-terminal UBA domain of Mex67, and Mex67 ΔUBA$_{(481-544)}$ composed of the N-terminal region of the UBA domain starting just after the NTF2 domain.

4.7.1. Expression and purification

The same protocol used for expressing and purifying Mex67 C-term$_{(528-599)}$ (section 4.5.1) was followed to express and purify Mex67 UBA$_{(545-599)}$ and Mex67 ΔUBA$_{(481-544)}$. Both construct showed good solubility, although Mex67 ΔUBA$_{(481-544)}$ appeared to be a bit less soluble and stable especially when trials to remove its GST tag were performed possibly affected by its location in a disordered region. The

GST tag was kept or was further removed by using PreScission protease depending on the requirement of latter experiments. Additionally, Size exclusion chromatography was used as a final purification step for both protein constructs. In both cases the protein eluted in a single peak in a volume reflecting a possible dimerization of the proteins.

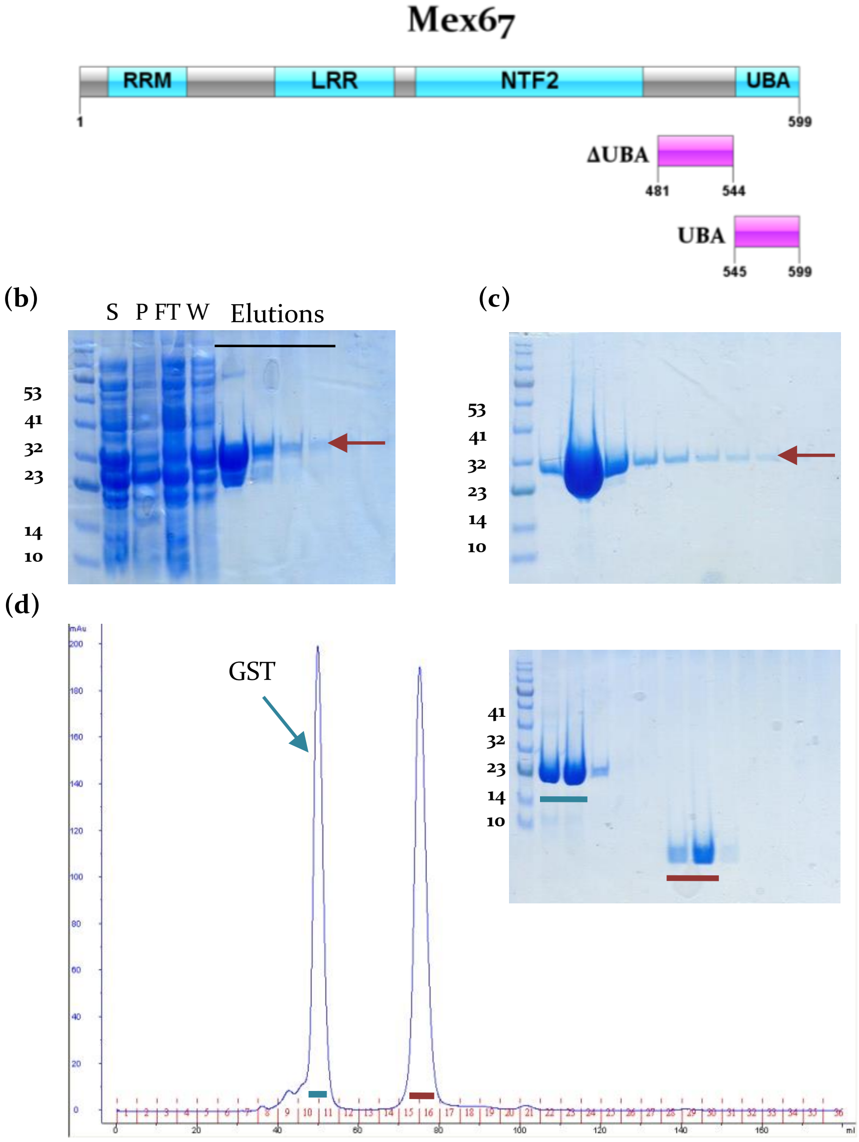
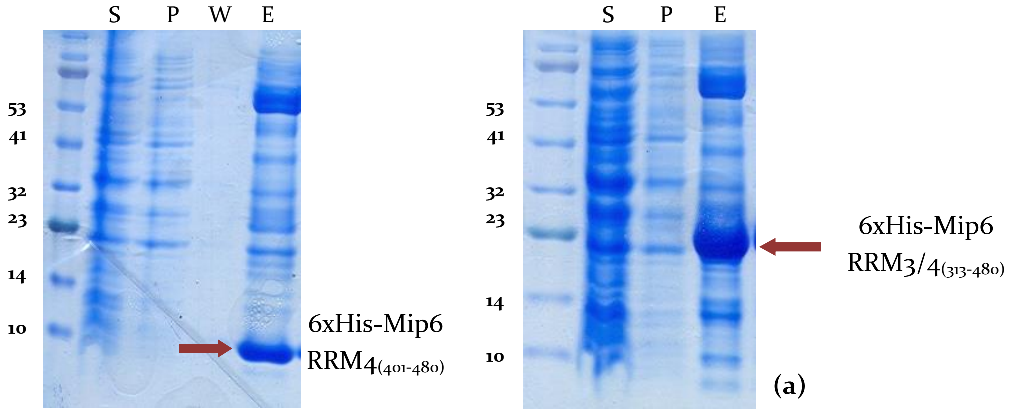


Figure 4.30. Purification of Mex67 ΔUBA$_{(481\text{-}544)}$ and Mex67 UBA$_{(545\text{-}599)}$. **(a)** An illustration shown the boundaries of Mex67 ΔUBA$_{(481\text{-}544)}$ and Mex67 UBA$_{(545\text{-}599)}$ constructs used. **(b)** 10% SDS-PAGE gel with GST-Mex67 ΔUBA$_{(481\text{-}544)}$ samples from the affinity chromatography purification on a 5ml GSTrap. **(c)** 10% SDS-PAGE gel with samples of GST-Mex67 UBA$_{(545\text{-}599)}$ purified on GSTrap affinity chromatography column. **(d)** Chromatogram showing Mex67 UBA$_{(545\text{-}599)}$ elution profile from size exclusion Superdex 75 16/60 column after the GST tag digestion and

the corresponding 10% SDS-PAGE gel with the Mex67 $\text{UBA}_{(545\text{-}599)}$ protein fractions highlighted in red.

4.7.2. Removing Mex67 UBA domain disrupts the interaction with Mip6

A small scale pull down assay using GST-Mex67 $\Delta\text{UBA}_{(481\text{-}544)}$ on one hand and 6xHis tagged Mip6 $\text{RRM4}_{(401\text{-}480)}$ or 6xHis tagged Mip6 $\text{RRM3/4}_{(313\text{-}480)}$ on the other hand was performed on nickel agarose beads to capture the His tag. Proteins were expressed on a small scale using autoinduction at 20°C overnight, and their cells were then lysed separately. The soluble fraction obtained from 6xHis- $\text{RRM4}_{(401\text{-}480)}$ or Mip6 $\text{RRM3/4}_{(313\text{-}480)}$ were incubated solely with the nickel charged agarose beads respectively for 15 minutes at 4°C prior to the addition of the second GST-Mex67 $\Delta\text{UBA}_{(481\text{-}544)}$ protein. After an additional incubation of 1 hour, the beads were washed extensively four times with buffer prior to elution. The two proteins did not elute together from the beads, instead 6xHis tagged Mip6 $\text{RRM4}_{(401\text{-}480)}$ and 6xHis tagged Mip6 $\text{RRM3/4}_{(313\text{-}480)}$ each eluted alone from the Nickel agarose beads with no detection of Mex67 $\Delta\text{UBA}_{(481\text{-}544)}$. As a control, the same procedure was carried out using GST-Mex67 $\text{C-term}_{(528\text{-}599)}$ shown earlier to form a complex with Mip6 (section 4.5). Our result indicates that Mex67 $\Delta\text{UBA}_{(481\text{-}544)}$ was not sufficient on its own to achieve interaction with Mip6 RRM4 and the removal of the UBA domain disrupted the binding (figure 4.31). The experiment pointed out the fact that indeed Mex67 UBA is required for the complex formation.



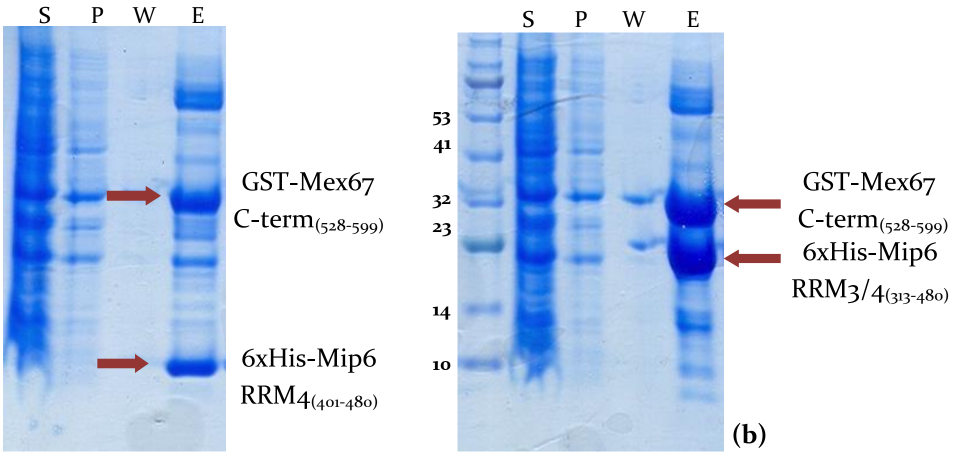
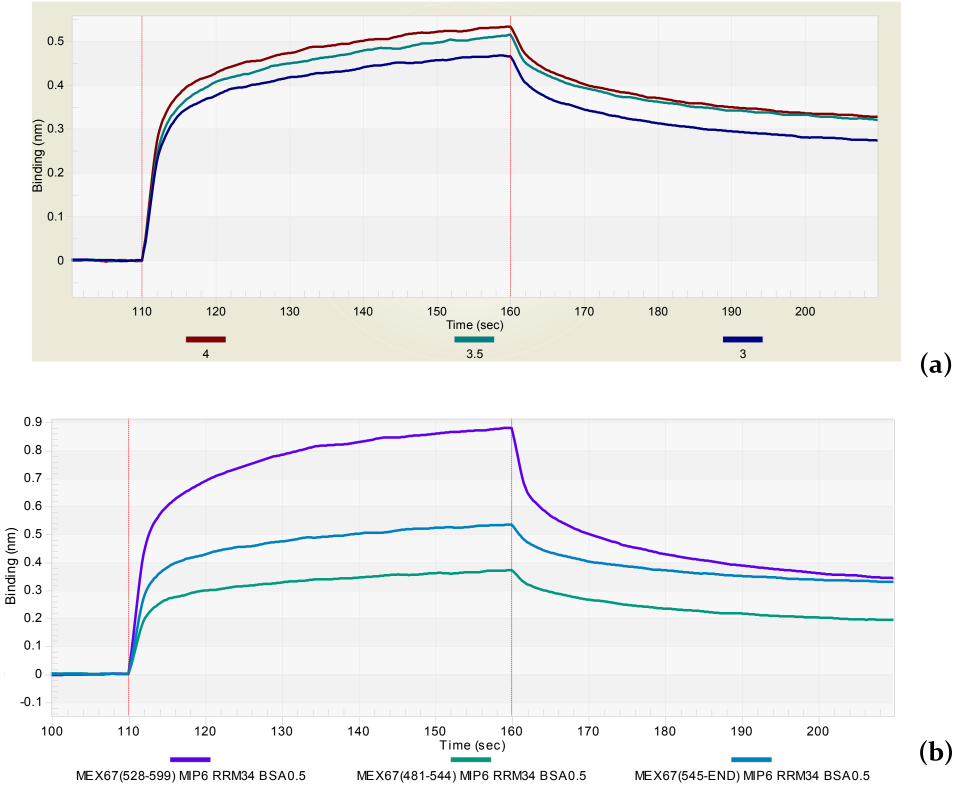


Figure 4.31. Mex67 UBA domain deletion disrupts the interaction with Mip6. (a) 10% SDS-PAGE gel with samples from the pull down assay on Ni-NTA beads using His tagged Mip6 RRM4$_{(401-480)}$ or Mip6 RRM3/4$_{(313-480)}$ constructs as a bait while GST-Mex67 ΔUBA$_{(481-544)}$ as a prey. S: the supernatants fraction, P: the insoluble pellet fraction, W: wash, and E: the elution of the beads. Interaction lost and only the His tagged Mip6 RRM4 constructs eluted from the beads. **(b)** As a control, pull down assay on Nickel beads using GST-Mex67 C-term$_{(528-599)}$ as a prey and His-Mip6 RRM4$_{(401-480)}$ or Mip6 RRM3/4$_{(313-480)}$ as bait proteins. Both proteins co-eluted from the beads indicating interaction.

4.7.3. Mex67 UBA domain is the site of binding of Mip6

To verify that Mex67 UBA domain was necessarily required for Mip6 interaction, purified protein of either GST-Mex67 ΔUBA$_{(481-544)}$ or GST-Mex67 UBA$_{(545-599)}$ (section 4.7.1) were tested using Bio-layer interferometry for binding Mip6 RRM3/4$_{(313-480)}$. However, the Bio-layer interferometry system required tagged proteins to carry on the experiment, so the GST tag fused N-terminally to Mex67 constructs was not removed and the GST-tagged proteins were immobilized on anti-GST sensors. For this experiment, increasing concentrations of purified Mip6 RRM3/4$_{(313-480)}$ (section 4.3.1) were used as a prey protein since Mip6 RRM4$_{(401-480)}$ size was on the lower range of the recommended detectable sizes for this technique. While we did not detect interaction of GST-Mex67 ΔUBA$_{(481-544)}$ to Mip6 RRM3/4$_{(313-480)}$, interestingly GST-Mex67 UBA$_{(545-599)}$ was binding with an affinity Mex67 UBA$_{(545-599)}$ comparable to that obtained for Mex67$_{(528-599)}$ (figure 4.32), and almost identical to that calculated for the binding GST-Mex67 C-term$_{(528-599)}$ to Mip6 RRM3/4$_{(313-480)}$ (figure 4.22) with no significant difference in the association and dissociation constants.



Construct	**Binding affinity (K_D)**
GST-MEX67 C-term$_{(528\text{-}599)}$	1.42 μM
GST-MEX67 UBA$_{(545\text{-}599)}$	1.94 μM
GST-MEX67 ΔUBA$_{(481\text{-}544)}$	No binding detected

(c)

Figure 4.32. Mex67 UBA domain is sufficient for the interaction with Mip6 RRM4. (a) The association/dissociation profile of Mip6 RMM3/4$_{(313\text{-}480)}$ to GST-Mex67 UBA$_{(545\text{-}599)}$ immobilized on an anti-GST sensor. **(b)** The association/dissociation profile of Mip6 RRM$_{(313\text{-}480)}$ at the same concentration, binding to the three different GST tagged Mex67 c-terminal constructs immobilized on anti-GST sensor using Bio-layer interferometry. The binding affinity in molar calculated for each experiment is shown in the table **(c)** showing clearly that the binding affinity obtained with GST-MEX67 UBA$_{(545\text{-}599)}$ is almost identical to that calculated with GST-MEX67 C-term$_{(528\text{-}599)}$.

4.7.4. Mex67 UBA domain needs to be in a specific conformation to be able to bind Mip6

Afterwards, we decided to reconstruct the complex Mex67 UBA$_{(545\text{-}599)}$/Mip6 RRM4 *in vitro* to check whether it can stably hold together in gel filtration which consequently indicates strong binding. For this purpose, equimolar concentration of purified non-tagged protein from both Mex67$_{(545-599)}$ and Mip6 RRM4$_{(401-480)}$ were mixed and left on ice half an hour before injecting in gel filtration superdex75 16/60 column. To our surprise, the two proteins eluted from gel filtration independently in two separate peaks (figure 4.33) which indicated that the removal of the N-terminal GST tag of Mex67$_{(545-599)}$ was possibly disrupting the binding. The elution of Mex67 UBA$_{(545-599)}$ as shown in figure 4.33 preceded the elution of Mip6 RRM4$_{(401-480)}$ however its lower molecular weight, due to the fact that Mex67 UBA$_{(545-599)}$ was eluting from gel filtration in a volume corresponding to a possible dimer formation while Mip6 RRM4$_{(401-480)}$ was eluting after its expected elution volume as discussed earlier in section 4.3.3.

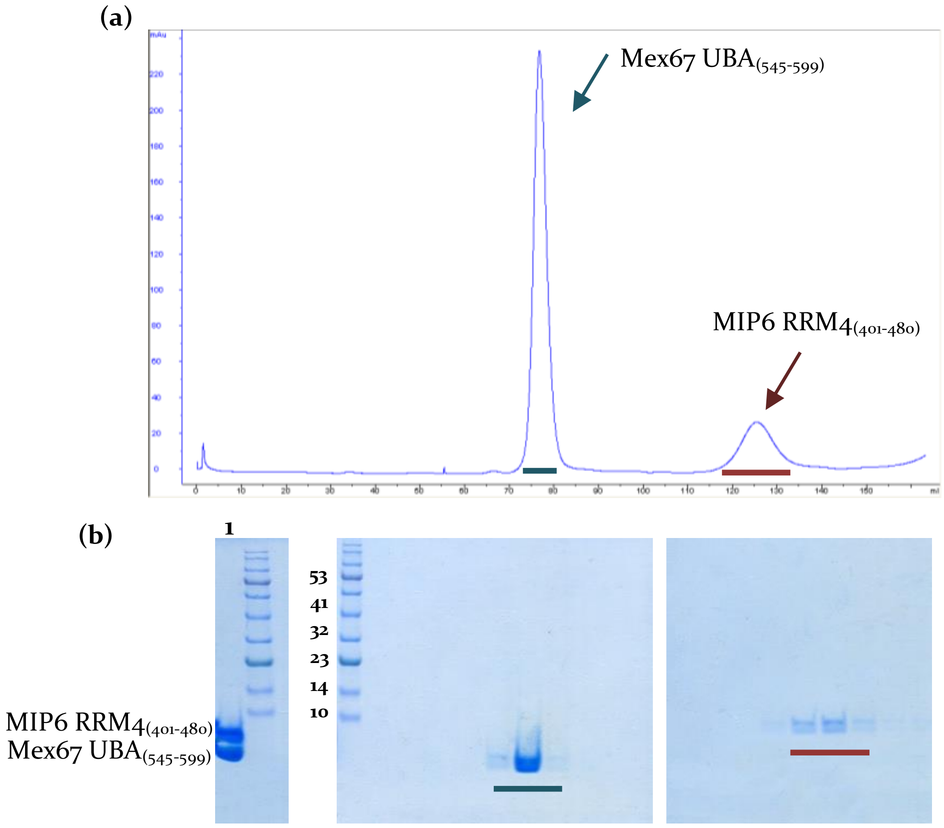


Figure. 4.33. (a) Gel filtration chromatogram showing the elution profile after the injection of the equimolar pre-mixing of Mex67 UBA$_{(545-599)}$ with Mip6 RRM4$_{(401-480)}$. The protein eluted in two separate peaks. **(b)** The input proteins when injected (1) and the samples corresponding to the gel filtration elution peaks were visualized on a 10% SDS-PAGE gel. The bands corresponding to Mex67 UBA$_{(545-599)}$ elution peak is shown in dark blue, while that of RRM4$_{(401-480)}$ elution is marked in red.

To elaborate on these results and better understand the effect on the N-terminal tag fused to the Mex67 UBA$_{(545\text{-}599)}$, ITC experiments were carried out to study the binding kinetics of a GST- Mex67 UBA$_{(545\text{-}599)}$ or a non-tagged Mex67 UBA$_{(545\text{-}599)}$ to Mip6 RRM4$_{(390\text{-}480)}$. To our surprise, increasing concentrations up to 1 mM of no tag Mex67 UBA$_{(545\text{-}599)}$ (GST tag was removed using PreScission protease (section 4.7.1) titrated into 40 µM of Mip6 RRM4$_{(390\text{-}480)}$ in the sample cell at 15°C did not show any binding (figure 4.34). The experiment was repeated several times with different concentrations and the same no binding result was obtained. However, since the previous Bio-layer interfermotry assay necessitated a tagged protein immobilized on corresponding sensors for the experiment to be carried out, and hence a GST tagged Mex67 UBA$_{(545\text{-}599)}$ was subsequently used (section 4.7.3), the ITC experiment was repeated with the same concentration of Mip6 RRM4$_{(390\text{-}480)}$ in the sample cell but this time using the tagged GST-Mex67 UBA$_{(545\text{-}599)}$ as a titrant. The kinetically determined values were consistent with those obtained from Bio-layer interferometry, replacing the endogenous N-terminal residues of the Mex67 UBA domain with another coming from the fused N-terminal GST tag replenished the binding of Mex67 UBA$_{(545\text{-}599)}$ to Mip6$_{(390\text{-}480)}$ with an affinity of K_D= 3.3 µM comparable to that of Mex67 Nt-UBA$_{(528\text{-}599)}$ (figure 4.32 c). The experiment was also repeated using different concentration and the same result was secured. Therefore, our results demonstrate that Mex67 UBA was the site of binding of Mip6 but where a mandatory N-terminal extension on the UBA domain is required, not necessarily sequence specific but more likely for certain conformational purposes putting the Mex67 UBA domain in a conformation required for Mip6 identification or maybe for a role in stabilizing the interaction once Mip6 is actually bound.

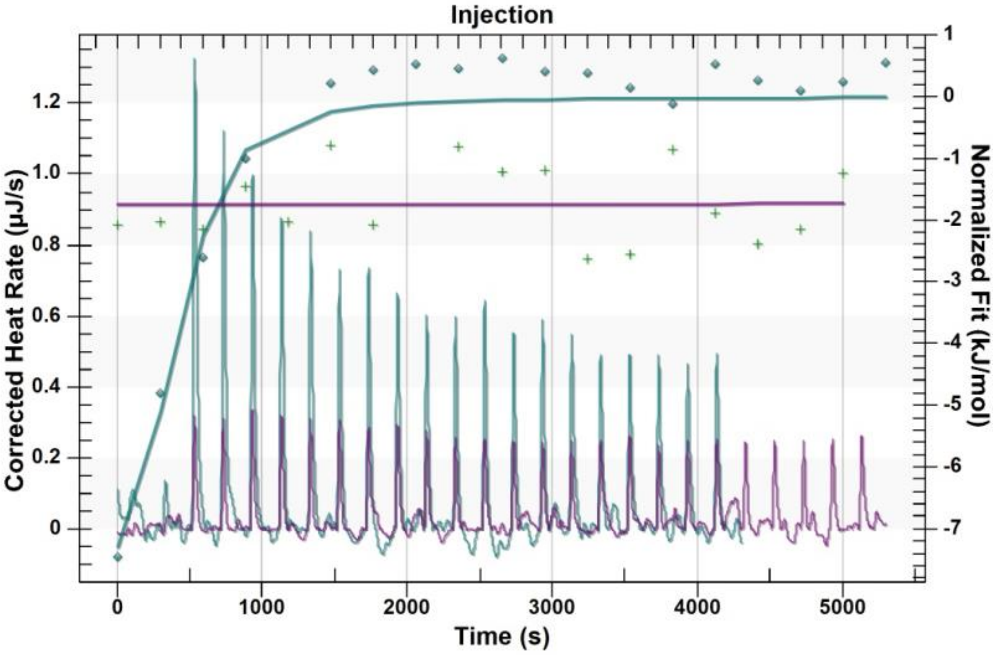


Figure 4.34. Mex67 UBA domain needs an N-terminal extension to be able to bind Mip6. Graph showing the difference in binding profile between the obtained raw heat peaks and the corresponding integrated binding curve over the course of time obtained from ITC experiment between Mip6 RRM4$_{(390-480)}$ and GST-Mex67$_{(545-599)}$ (green), and the no binding obtained when no tag Mex67 UBA$_{(545-599)}$ was used as a titrant instead (blue). The calculated binding affinity of GST-Mex67$_{(545-599)}$ to Mip6 RRM4$_{(390-480)}$ was 3.3 µM comparable for that of non-tagged Mex67$_{(528-599)}$.

4.8. Chemical cross-linking

To complement the results, we proceeded with a cross-linking experiment, a process where two proteins or two regions of a protein are linked by covalent bonds formed by a chemical reaction between the reactive group of a chosen cross-linker and a functional group of the protein. This procedure aimed to stably link the two proteins Mex67$_{(528-599)}$ and Mip6 RRM4$_{(390-480)}$ in the attempt to have a closer look on the exact sites of interaction. Since the sequence of Mex67$_{(528-599)}$ had lysines (-NH2 amine group) and no cysteines, while Mip6 RRM4$_{(390-480)}$ sequence contained two cysteines (-SH thiol group), Sulfo-SMPB (sulfosuccinimidyl 4-maleimidophenyl)butyrate) was our cross-linker of choice as it is a water soluble amine to sulfhydryl crosslinker (figure 4.35).

Different concentrations of the cross-linker were incubated with 200 µg of the stably formed Mex67$_{(528-599)}$/Mip6 RRM4$_{(390-480)}$ complex, obtained after size exclusion chromatography (figure 4.24. b), for 1 hour at 25°C. The reaction was stopped by adding 1 µl of Tris-HCl 0.5M, and samples were heated at 95°C for 10 minutes with the 6x sample buffer and run on 15% SDS-PAGE gel. The cross-linked shifted band visualized on the SDS-PAGE corresponding to the size of the cross-linked complex was then gel extracted and sent for mass spectrometry analysis (University of Valencia, Burjassot, Valencia). The samples were digested with trypsin and analyzed using MALDI TOFTOF. The resulting MS and MS/MS data were then analyzed using MASCOT software (Matrix-Science) for peptide identification.

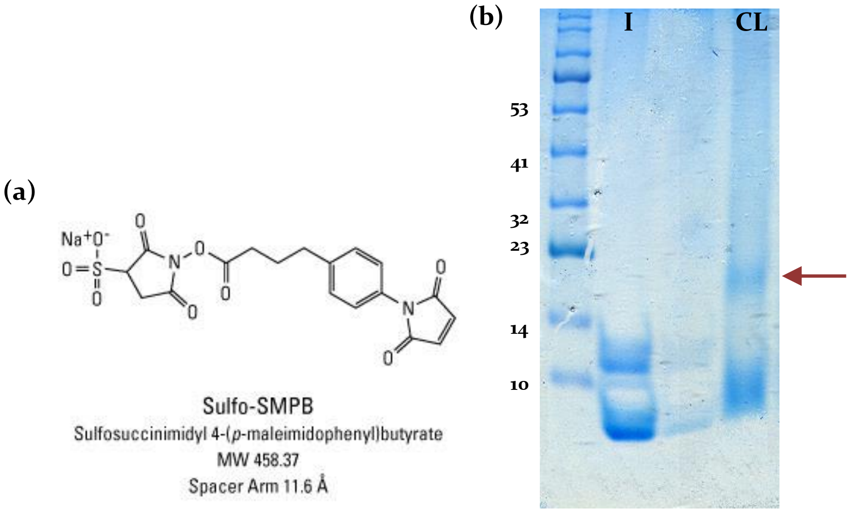


Figure. 4.35. Cross-linking Mex67$_{(528-599)}$ to Mip6 RRM4$_{(390-480)}$.
(a) Chemical structure of the cross-linker Sulfo-SMPB used for the experiment. **(b)** 15% SDS-PAGE gel showing the input Mex67$_{(528-599)}$/ Mip6 RRM4$_{(390-480)}$ complex before adding the cross-linker (I), and the obtained result after adding the cross-linker (CL). The resulting band (~20 KDa) sent for mass spectrometry is shown with a red arrow.

The peptides identified by MASCOT in either Mex67$_{(528-599)}$ or Mip6 RRM4$_{(390-480)}$, representing the regions not involved in the interaction, were residues 528-554 and 583-593 of Mex67; and residues 408-415 and 459-480 of Mip6 RRM4 respectively (figure 4.36). This meant that the site of Mex67 binding to Mip6 RRM4 was most probably somewhere between residues 555-582 of the UBA domain (figure 4.36). Interestingly, when the residues were visualized on the obtained Mex67 C-term$_{(528-599)}$ structure, the central α helix 2 of the UBA domain is seen as the core site involved in interaction. These results complemented our results and ensured that role of Mex67 UBA domain as the platform of physical interaction with Mip6. Although the results obtained from this experiment were not ideal as it couldn't provide us with an exact site of binding but it emphasized our previous results and gave us a better idea on the mechanism of interaction.

(a)
Mip6 RRM4:
GPLGSMKPVHNQIGTNKKTKVYLKNLSFNCNKEFISQLCLQEKIRFSEIKITNYNS
LNWTFCGHVECFSRSDAERLFNILDRRLIGSSLVEASWSKN
Mex67:
GPLGSMVMAPTLQLPPDVQSRLNPVQLELLNKLHLETKLNAEYTFMLAEQSNWNYE
VAIKGFQSSMNGIPREAFVQF

(b)

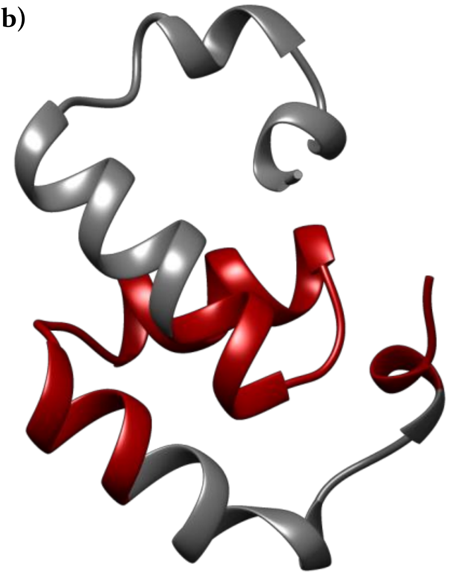
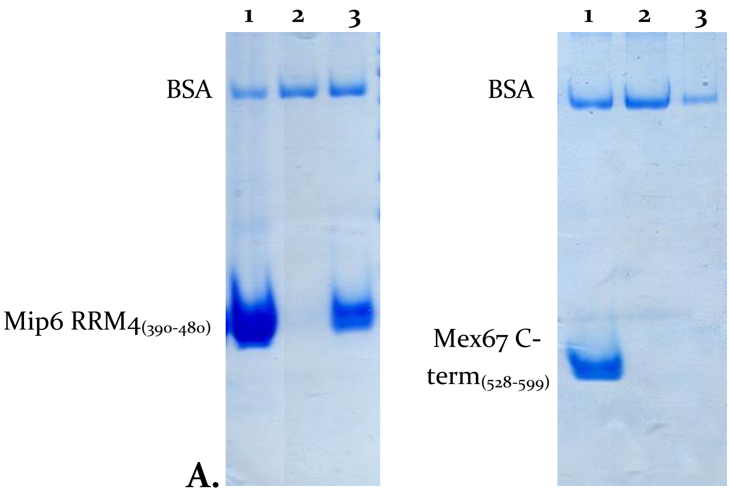


Figure 4.36. MASCOT search results. **(a)** The matched peptides, after trypsin treatment, identified by MASCOT for each protein are highlighted in red. Residues in grey are additional due to the cloning technique and not part of the protein sequence. **(b)** Structural representation of the results on Mex67$_{(528-599)}$ structure, where the regions that represent the probable site of binding as deducted from MASCOT results are highlighted in red. As shown on figure, these include majorly whole alpha helix 2 of the UBA, the last part of the first α helix of the UBA domain, and the beginning of α3.

4.9. Mip6 RRM4 RNA binding interfere with its interaction with Mex67

We then asked whether the binding of Mip6 RRM4 to RNA (section 4.3.5) interferes with its ability to interact with Mex67 or vice versa. To answer this question, the pre-formed stable complexes of Mex67$_{(528-599)}$/Mip6 RRM4$_{(390-480)}$ or Mex67$_{(528-599)}$/Mip6 RRM4$_{(401-480)}$ obtained directly from gel filtration (section 4.5.3; figure 4.24) were checked for the ability to bind RNA *in vitro*. This was determined by incubating 100 or 200 µg of each pre-formed complex with polyuridylic acid (PolyU) agarose beads for one hour on a rotator at 4°C. The non-bound fraction was then removed, and the beads washed extensively three times. Notably, The incubation of a preformed Mip6 RRM4/Mex67$_{(528-599)}$ complex with the Poly(U) agarose beads hindered the binding of Mip6 RRM4 to the Poly (U) (figure 4.37). On the other hand, incubating Mip6 RRM4 with the Poly(U) agarose beads first for 20 minutes before the addition of equimolar concentration of Mex67$_{(528-599)}$ and further incubation for an additional 45 minutes resulted in the elution of Mip6 RRM4 alone bound to the Poly(U). A control fraction showed the co-elution of Mip6 RRM4$_{(390-480)}$ with poly(U) when incubated separately with the Poly(U) agarose beads. The same results were obtained when the experiments were repeated.



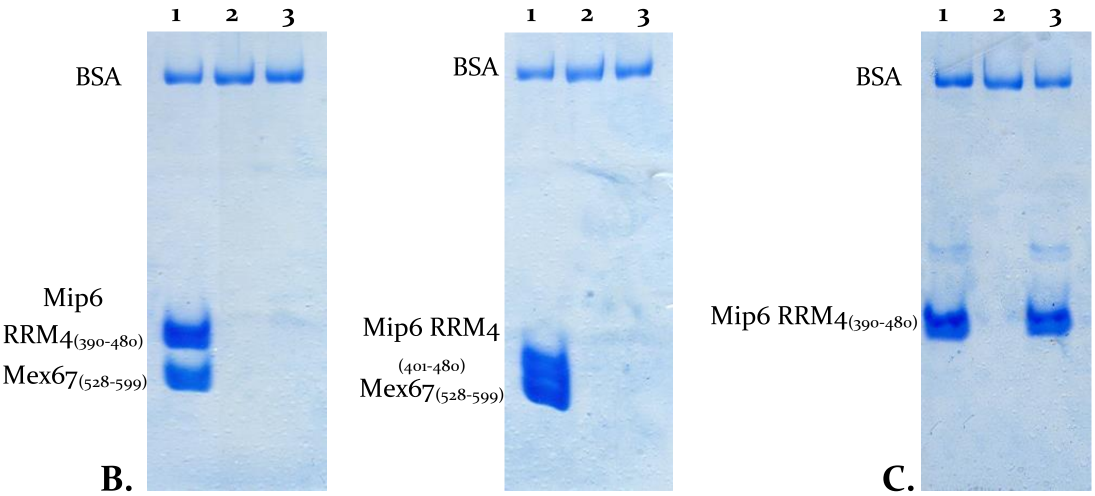


Figure 4.37. The complex formation with Mex67 affects binding of Mip6 RRM4 to RNA *in vitro*. **A.** As a control for the experiment incubation of Mip6 RRM4$_{(390-480)}$ or Mex67 C-term$_{(528-599)}$ separately with the Poly (U) agarose beads showed binding of Mip6 RRM4$_{(390-480)}$ but no binding of Mex67 C-term$_{(528-599)}$ as expected. **B.** The incubation of preformed complexes Mex67$_{(528-599)}$/Mip6 RRM4$_{(390-480)}$ or Mex67$_{(528-599)}$/Mip6 RRM4$_{(401-480)}$ with Poly (U) beads showed no co-elution with Poly (U) from the agarose beads in both cases as visualized on the coomassie stained 10% SDS-PAGE, while the primarily incubation with Mip6 RRM4$_{(390-480)}$ for 20 minutes before the addition of Mex67$_{(528-599)}$ for further incubation before the beads being washed and eluted lead to the elution of Mip6 RRM4$_{(390-480)}$ solely with the Poly (U) **(C)** . All samples were visualized on 10% SDS-PAGE gels.

1: Input; 2: wash; 3: protein eluting with Poly(U) from beads

Additionally, the native gel electrophoresis gel was employed using a pre-formed complex Mex67$_{(528-599)}$/Mip6 RRM4$_{(390-480)}$ and adding increasing concentration of total cell RNA extract. On the native PAGE Mex67$_{(528-599)}$ appeared as clear band while the complex Mex67$_{(528-599)}$/Mip6 RRM4$_{(390-480)}$ appeared as a higher smeared band and a Mip6 RRM4$_{(390-480)}$ band was not visible probably due to the close pI of the protein to the pH of running native. The isoelectric point (pI) of Mip6 RRM4$_{(390-480)}$ was on the basic side pI= 9.39, the isoelectric point of Mex67$_{(528-599)}$ was pI=5.13, while the theoretical calculation of pI for the complex was around 9. Due to this difference in isoelectric points between the proteins, obtaining a good native gel was tricky and after several trials the best result were from a 10% native gel electrophoresis of pH 8.3 is shown in figure 4.38. However, the presence of RNA extract in increasing concentrations did not affect the smeared band of the complex and no shift was seen. Although the native gel result was not brilliant, but it was compatible with the results obtained from the Poly(U) pull down experiment indicating that Mex67$_{(528-599)}$ and RNA were not binding simultaneously to Mip6 RRM4, nor competing for the binding although the binding affinity of both partners to Mip6 RRM4 was comparable, instead the binding of one partner first was inhibiting the binding of the other.

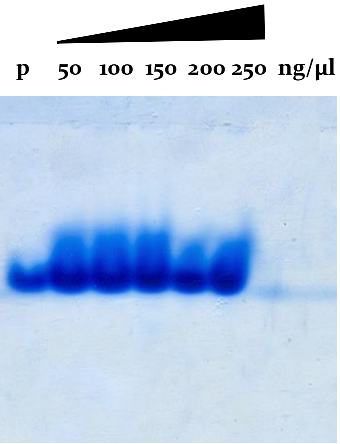
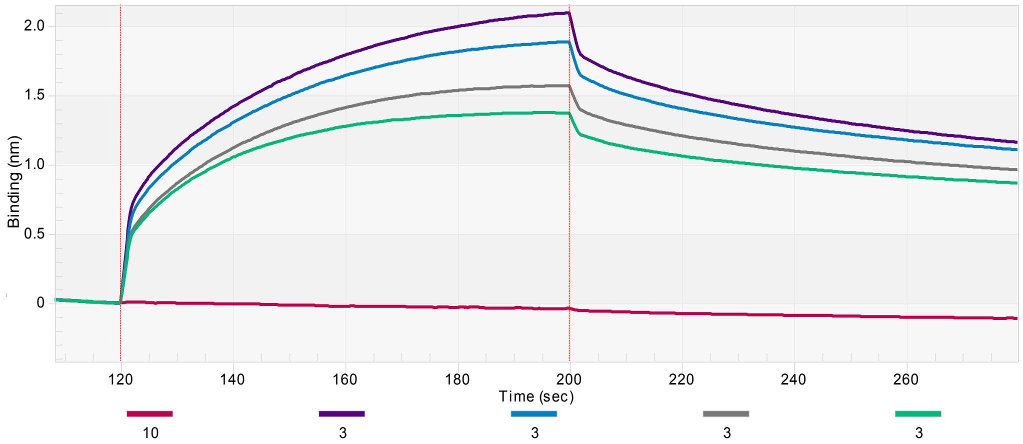


Figure 4.38. Native gel electrophoresis showed no change nor shift in the smeared band of the complex Mex67$_{(528-599)}$/Mip6 RRM4$_{(390-480)}$ upon increasing concentrations of RNA. P represents the protein band of Mex67$_{(528-599)}$ when run independently on the gel.

Nevertheless, a Mip6 construct (Mip6 RRM3/4$_{(313-480)}$) containing an additional RRM other than RRM4 was still able to interact with RNA when Mex67 was bound when tested using Bio-layer interferometry. Indeed, forming an Mip6 RRM3/4$_{(313-480)}$/ Mex67$_{(528-599)}$ of 1:1, 1:2, or 1:4 molar ratio respectively did not affect its binding the biotinylated Poly(U) RNA immobilized on the streptavidin biosensors (figure 4.39). Instead, the calculated binding affinity of all was a slightly lower but similar to the binding of just RRM3/4$_{(313-480)}$ to the Poly(U) RNA. A negative control showed no binding to the Poly(U) by Mex67$_{(528-599)}$. This suggests that although Mex67 binding affected the RNA binding of Mip6 RRM4, the protein was still able to bind RNA probably through its other RRM domain.

(a)



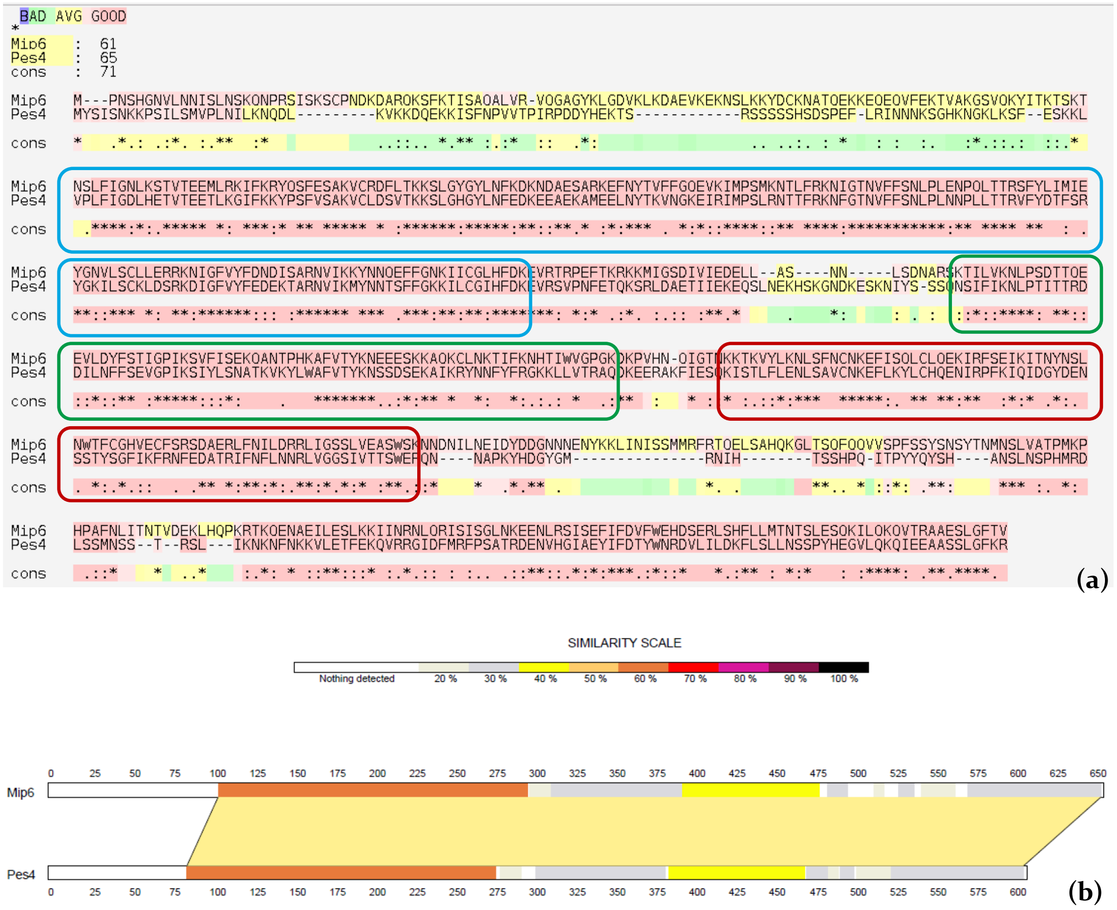
(b)

Sample ID	Concentration (μM)	K_D (μM)
Btn-PolyU15/MIP6RRM3/4$_{(313-480)}$ (Positive control)	3	1.456
Btn-PolyU15:MIP6RRM3/4$_{(313-480)}$/MEX67$_{(528-599)}$ 1:1 ratio	3	1.337
Btn-PolyU15:MIP6RRM3/4$_{(313-480)}$/MEX67$_{(528-599)}$ 1:2 ratio	3	1.305
Btn-PolyU15:MIP6RRM3/4$_{(313-480)}$/MEX67$_{(528-599)}$ 1:4 ratio	3	1.002
Btn-PolyU15:MEX67$_{(528-599)}$	10	No binding detected

Figure 4.39. Mex67$_{(528-599)}$ did not interfere with Mip6 RRM3/4 interaction with RNA. (a) association and dissociation profiles obtained from the interaction of Mip6 RRM3/4$_{(313-480)}$ to sensor immobilized Btn-Poly(U) RNA of 15 residues (violet), plus when complexed with Mex67$_{(528-599)}$ at 1:1, 1:2, and 1:4 molar ratio (blue, grey, and green respectively). Mex67$_{(528-599)}$ showed no binding on its own to Poly(U) (red). **(b)** Table summarizing the concentrations used for each experiment with the calculated binding affinity in micro molar corresponding to each.

4.10. Pes4, a Mip6 Paralogue, is also an RNA binding protein

Pes4, also known as DNA polymerase epsilon suppressor 4, is a yeast protein from *Saccharomyces cerevisiae* that was said to show high homology to Mip6 (Segref *et al.*, 1997). A sequence alignment indeed showed high sequence homology. The two proteins share ⍰ 52.45% of sequence similarity and ⍰ 37% of percentage identity. The ordered RRMs regions show the highest similarity that gets lower in the not ordered regions in the N and C term (figure 4.40).



```
 BAD AVG GOOD
*
Mip6     :  61
Pes4     :  65
cons     :  71

Mip6   M---PNSHGNVLNNISLNSKQNPRSISKSCPNDKDARQKSFKTISAQALVR-VQGAGYKLGDVKLKDAEVKEKNSLKKYDCKNATQEKKEQEQVFEKTVAKGSVQKYITKTSKT
Pes4   MYSISNKKPSILSMVPLNILKNQDL---------KVKKDQEKKISFNPVVTPIRPDDYHEKTS------------RSSSSSHSDSPEF-LRINNNKSGHKNGKLKSF--ESKKL

cons   *    .*.: .:*. :.**  :*                ..:.. *.** :.:*  ::  .*:                .. ..:. : *   : :  :.    :*.::.:  ::.*

Mip6   NSLFIGNLKSTVTEEMLRKIFKRYQSFESAKVCRDFLTKKSLGYGYLNFKDKNDAESARKEFNYTVFFGQEVKIMPSMKNTLFRKNIGTNVFFSNLPLENPQLTTRSFYLIMIE
Pes4   VPLFIGDLHETVTEETLKGIFKKYPSFVSAKVCLDSVTKKSLGHGYLNFEDKEEAEKAMEELNYTKVNGKEIRIMPSLRNTTFRKNFGTNVFFSNLPLNNPLLTTRVFYDTFSR

cons    .****:*:.***** *: ***:* ** ***** * :******:*****:**::**.* :*:***  . *:*::****::** ****:***********:** **** **   : .

Mip6   YGNVLSCLLERRKNIGFVYFDNDISARNVIKKYNNQEFFGNKIICGLHFDKEVRTRPEFTKRKKMIGSDIVIEDELL--AS----NN-----LSDNARSKTILVKNLPSDTTQE
Pes4   YGKILSCKLDSRKDIGFVYFEDEKTARNVIKMYNNTSFFGKKILCGIHFDKEVRSVPNFETQKSRLDAETIIEKEQSLNEKHSKGNDKESKNIYS-SSQNSIFIKNLPTITTRD

cons   **::*** *: **:******::: :****** ***  .***:**:**:*******: *:*  .:*. :.:: :**.*       .      *:     :  . :   ::*::****: **::

Mip6   EVLDYFSTIGPIKSVFISEKQANTPHKAFVTYKNEEESKKAQKCLNKTIFKNHTIWVGPGKQKPVHN-QIGTNKKTKVYLKNLSFNCNKEFISQLCLQEKIRFSEIKITNYNSL
Pes4   DILNFFSEVGPIKSIYLSNATKVKYLWAFVTYKNSSDSEKAIKRYNNFYFRGKKLLVTRAQQKEERAKFIESQKISTLFLENLSAVCNKEFLKYLCHQENIRPFKIQIDGYDEN

cons   ::*::** :*****:::*:     .    *******..:*:** *  *:  *:.:.: *   .:**  :   * : * :.::*:***   *****:.  ** **:**   :*:*  .*:.

Mip6   NWTFCGHVECFSRSDAERLFNILDRRLIGSSLVEASWSKNNDNILNEIDYDDGNNNENYKKLINISSMMRFRTQELSAHQKGLTSQFQQVVSPFSSYSNSYTNMNSLVATPMKP
Pes4   SSTYSGFIKFRNFEDATRIFNFLNNRLVGGSIVTTSWERQN----NAPKYHDGYGM---------------RNIH--------TSSHPQ-ITPYYQYSH----ANSLNSPHMRD

cons   . *:.*.::  . .** *:**:*:.**:*.*:*  :**.::*    *   .*.**  .         *..        **.. * ::*:  .**:     *** :.  *:

Mip6   HPAFNLITNTVDEKLHQPKRTKQENAEILESLKKIINRNLQRISISGLNKEENLRSISEFIFDVFWEHDSERLSHFLLMTNTSLESQKILQKQVTRAAESLGFTV
Pes4   LSSMNSS--T--RSL---IKNKNFNKKVLETFEKQVRRGIDFMRFPSATRDENVHGIAEYIFDTYWNRDVLILDKFLSLLNSSPYHEGVLQKQIEEAASSLGFKR

cons    .::*    *  ..*     :.*: * ::**:::* :.*.:: : :.. .::**::.*:*:***.:*::*   *.:** : *:*     : :****: .**.****.
```

Figure 4.40. Sequence alignment between Mip6 and Pes4. (a) Alignment was obtained from using T-coffee multiple sequence alignment server. RRM 1 and 2 are shown in the blue box, while RRM3 is shown green box, and RRM4 shown in red. **(b)** A graph representing the similarity scale shows the highest percentage similarity in RRM1 and 2 by about 60%, while RRM3 and RRM4 show 30% and 40% similarity respectively. Computation was done using Expasy SIM alignment tool and visualized graphically using LALNVIEW program.

4.10.1. Expression and purification of Pes4 RRM3/4

Due to the difficulty of producing a soluble full length Pes4 protein and since our interest was more on RRM3 and RRM4 of the protein, a 6xHis tagged Pes4 RRM3/4$_{(303-473)}$ was obtained by over expression in a large scale ZY media using the auto-induction expression system. After cell lysis, the soluble fraction of the protein was purified on a 5 ml HisTrap affinity column. A good amount of soluble protein was obtained from the affinity chromatography. The protein was then incubated with PreScission protease overnight in order to remove the tag and was further purified using size exclusion chromatography on a superdex 75 16/60 gel filtration column. The protein eluted from gel filtration as a single peak at ~58 ml corresponding to the size of a dimer. The corresponding protein fractions visualized on 10% SDS-PAGE gel reflected pure protein (figure 4.41) that was then concentrated and flash -frozen in liquid nitrogen before being stored at -80% for further use.

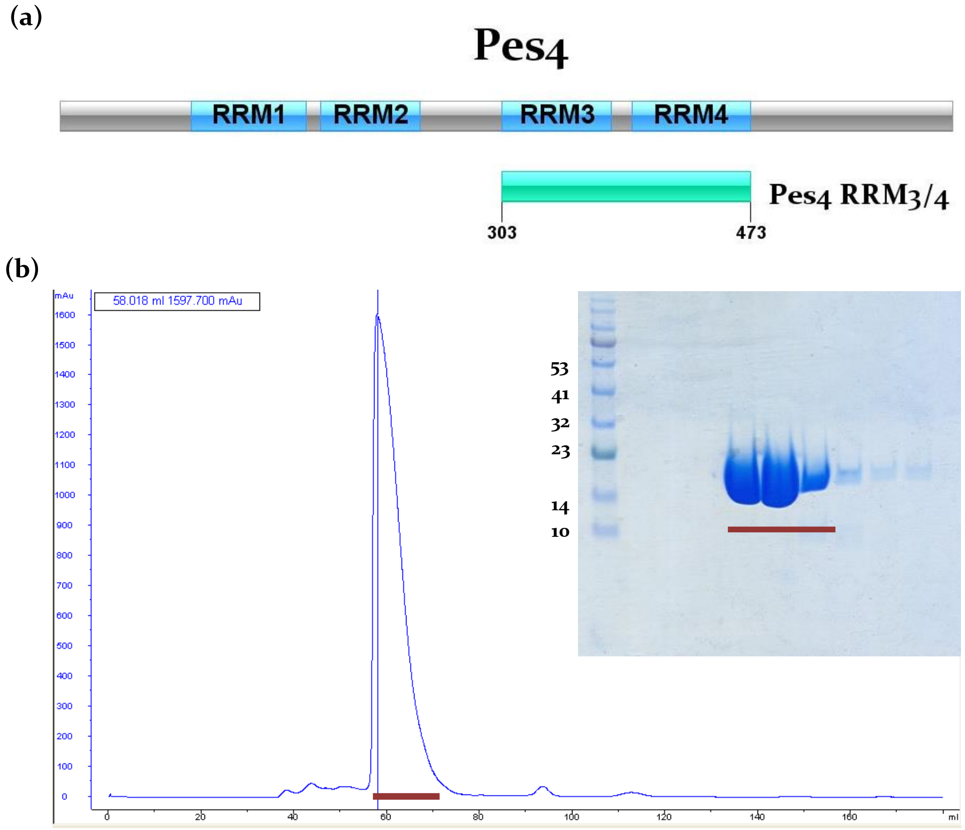


Figure 4.41. Size exclusion chromatography elution profile of Pes4 RRM3/4. (a) Illustration showing the boundaries of the Pes4 construct used in this study. **(b)** The protein eluted from gel filtration as a single peak at~ 58ml of volume. The corresponding fractions were visualized on a 10% SDS-PAGE gel. The protein appeared at its proper size and is highlighted with a red bar.

4.10.2. Pes4 binds RNA *in vitro* with high affinity

The purified Pes4 RRM3/4 containing the predicted RRM3 and RRM4 as aligned with Mip6 was then used to investigate whether Pes4 is also an RNA binding protein. A biotinylated Poly (U) RNA of 15 residues immobilized on streptavidin sensors was used as bait using the biolayer interfermotry system. Increasing concentrations of Pes4 RRM3/4 (0.25, 0.5, 0.75, and 1 µM) were then used as a prey to see the binding profile. The obtained results showed a good binding affinity of K_D= 1.212 μM of the protein to the Poly (U), similar to that obtained from the Mip6 RRM3/4 (figure 4.42). 0.5 mg/ml of BSA was added to the buffer used for this experiment in order to diminish any non specific bindings. Association constant were Ka (1/Ms)= 3.108 e4 M while dissociation constant Kd (1/s)= 3.108 e-2 M respectively.

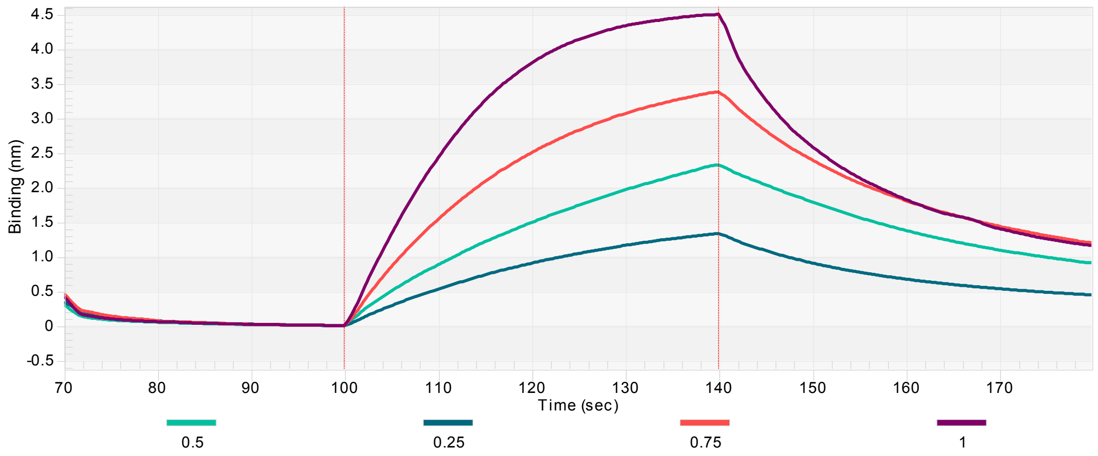


Figure 4.42. Pes4 is binds RNA with micro molar affinity in vitro. The graph representing the association and dissociation profile of each used concentration of Pes4 RRM3/4 to the sensor associated Poly (U) depicted in different colours. A binding affinity of **K_D= 1.212 μM** was calculated.

4.11. Pes4 RRM3/4 does not interact with Mex67 *in vitro*

Due to the fact that Pes4 shared a big sequence similarity with Mip6 as mentioned earlier especially in the ordered regions where the RNA recognition motifs exist (40% sequencing similarity for RRM4), it seemed logical and intriguing for us to check whether Pes4 RRM3/4 region was also able to interact with Mex67.

For this purpose, an equimolar concentrations of Mex67$_{(528\text{-}599)}$ and Pes4 RRM3/4 were incubated for half an hour before being injected in a Superdex 75 16/60 gel filtration column. Each protein eluted from gel filtration in independent peaks as the chromatogram shown in figure 4.43 illustrates. Corresponding fractions run on SDS-PAGE gel assured the result. This pointed to that Pes4 and Mex67 did not have the ability to form any stable complex *in vitro*.

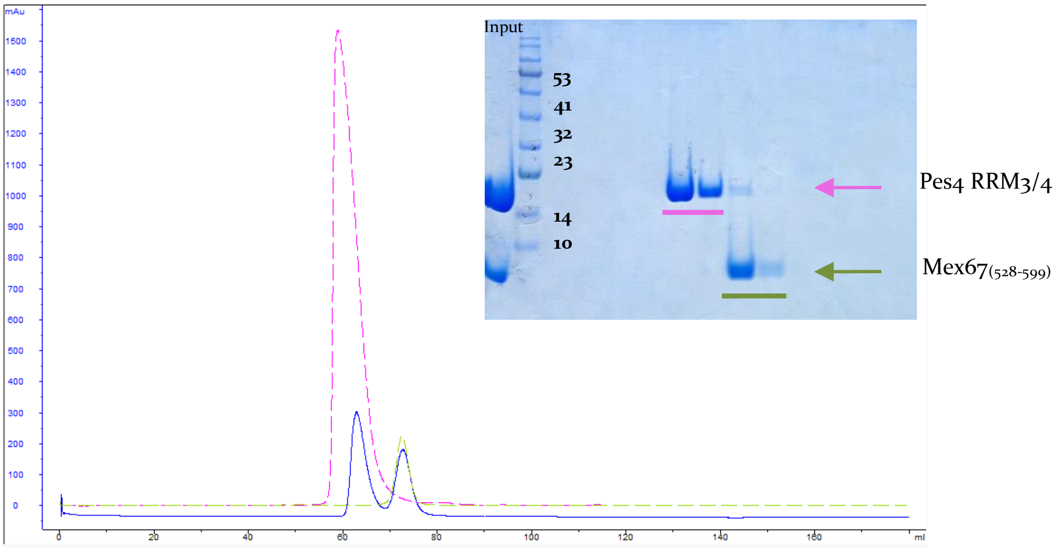


Figure 4.43. Pes4 RRM3/4 does not hold to Mex67$_{(528\text{-}599)}$ in gel filtration. The figure shows a superposition of the elution profile of Pes4 RRM3/4 and Mex67$_{(528\text{-}599)}$ when incubated then injected together in Superdex 75 16/60 gel filtration column in blue with the elution profiles of each protein when injected separately (pink for the elution profile of Pes4 RRM3/4 and light green for that of Mex67$_{(528\text{-}599)}$). The 10% SDS-PAGE gel shows the input containing both proteins as injected in gel filtration, and the fractions where each protein eluted as demonstrated.

To make sure that this was indeed the case, we utilized ITC as a more accurate and sensitive approach for assessing interactions and studying its kinetics. The experiment was carried out at 15°C, where 25 µM of Pes4 RRM3/4 was placed in the sample cell and titrated with 800 µM of Mex67$_{(528\text{-}599)}$. The experiment was repeated using different concentrations of both proteins in order to be certain of the result obtained. As expected, no binding heat was detected between both proteins confirming that no interaction was happening and that indeed Pes4 RRM3/4 did not have the ability to bind Mex67 or at least the binding was very weak to be detected (figure 4.44).

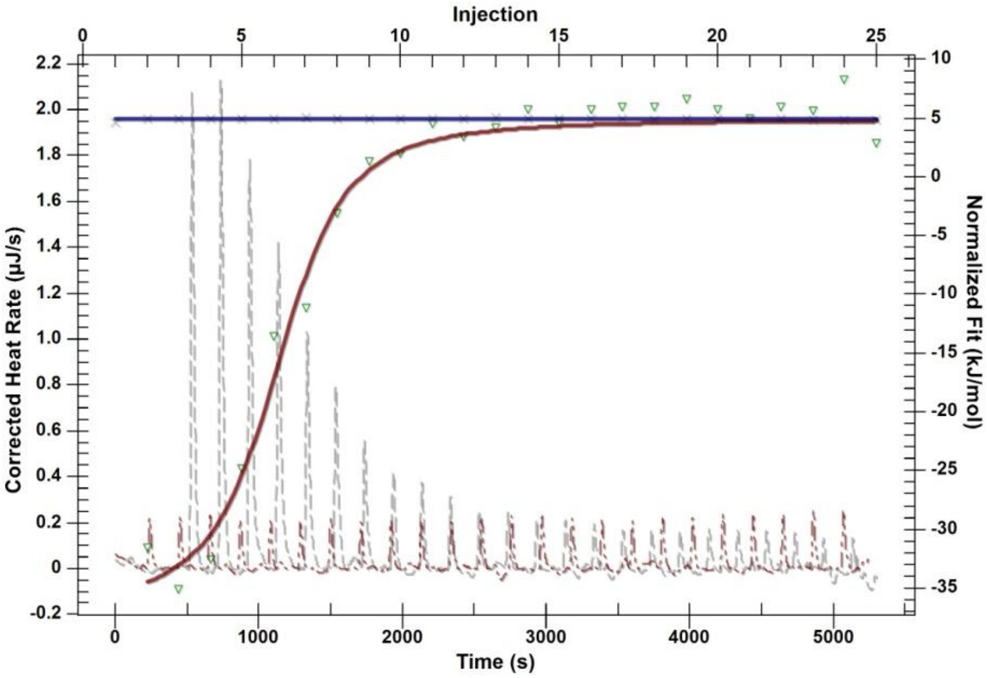


Figure 4.44. No interaction between Pes4 RRM3/4 and Mex67$_{(528-599)}$. Graph showing the binding heat peaks (grey) and the corresponding fitted curve (red) of Mip6 RRM4$_{(390-480)}$ when titrated by Mex67$_{(528-599)}$ compared to the no binding profile (red peaks and the corresponding blue line) obtained when was done using Pes4 RRM3/4 instead.

4.12. Crystal structures of Pes4

Pes4 RRM4 shared 42% percent of identical sequence with Mip6 RRM4, yet it was not able to interact with Mex67 C-term. To have a structural insight on Pes4, which in turn would give us an insight on the conserved regions with Mip6 RRM4, screening plates were set with Pes4 RRM3/4 in an attempt to crystallize it. And indeed crystals were obtained and structures solved.

4.12.1. Crystal structure of Pes4 RRM3

4.12.1.1. Crystallization and data processing

Initial crystals obtained from Pes4 RRM3/4 were clusters of needles that were not unique (figure 4.45).

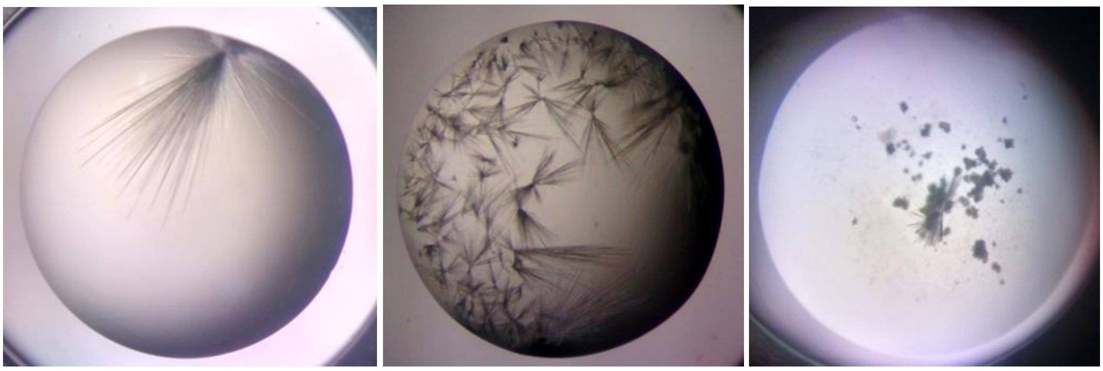


Figure 4.45. Pes4 RRM3/4 crystals. Some crystallization drops showing the form of the initial needle like crystals obtained from Pes4 RRM3/4 in different growth conditions.

Optimizations of the crystals were attempted by varying additive screens of some initial conditions. However, the crystals appearance did not improve much, but one pHs screen resulted in crystals that belonged to the $P3_221$ space group and diffracted up to 2.8 or 1.7 Å. In other growth conditions some crystals with similar unit cell dimensions diffracted to 2.4-2.8 Å:

Growth condition	Resolution
30% PEG 8000, 200 mM Ammonium Sulfate, 1M MES/NaOH pH 7.4	**1.7 Å**
30% PEG 8000, 200 mM Ammonium Sulfate, K2HPO4/NaH2PO4 pH 8.6	**2.8 Å**
1.26 M Sodium phosphate monobasic monohydrate, 0.14 M Potassium phosphate dibasic pH 5.6	**2.4-2.8 Å**

The best diffracting crystals belonged to the trigonal space group $P3_221$. Preliminary data analysis suggested that the unit cell volume wasn't big enough to fit the two RRM domains in the Pes4 RRM3/4 protein used initially to set the crystallization plates, instead it suggested enough space for only one RRM domain in the asymmetric unit. Data from the best diffracting crystal were processed with *XDS* and scaled with *Aimless* in *CCP4*. Resolution was cut at 1.9 Å and phases were obtained by molecular replacement using the coordinates of Mip6 RRM3 structure (section 4.4.3) as a model. The structure turned out to be that of Pes4 RRM3 with one molecule in the asymmetric unit and no sufficient cell volume to accommodate RRM4. This was probably due to some kind of in situ proteolysis of the protein in the most vulnerable and flexible linker region between both RRMs.

	Pes4 RRM3
Resolution (Å)	1.9
Space group	$P3_221$
Unit-cell parameters	
a,b,c (Å)	82.826, 82.826, 31.985
α, β, γ (°)	90, 90, 120
Wavelength (Å)	0.97623
Total reflections	198116 (19105)
Unique reflections	10165 (1003)
Completeness (%)	100
Multiplicity	19.5 (19)
Mean I/Sigma(I)	10.52 (1.74)
R_{merge}	0.3853 (2.211)
R_{meas}	0.3957
R_{work}/R_{free}	0.1899/0.2308
CC1/2	0.974(0.521)
Ligands	14
Water molecules	37
Protein residues	81
Ramachandran plot	
Favoured (%)	100
Outliers (%)	0
Clashscore	4.31
Average B factors	33.90

Table 4.6. Data collection and refinement statistics of Pes4 RRM3 crystal. Statistics from the highest resolution outer shell are indicated in parenthesis.

4.12.1.2. Pes4 RRM3 reveals a canonical RRM structure

Like Mip6 RRM3 (section 4.4.3), the structure of Pes4 RRM3 was that of an RRM with a β1α1β2β3α2β4 topology that forms a four stranded anti-parallel β sheets, packed against two alpha helices. The structure included additional 8 residues from the linker that connects Pes4 RRM3 to RRM4 (residues 377-384) that were probably the site of the proteolysis that led to the crystallization of RRM3 on its own (figure 4.46).

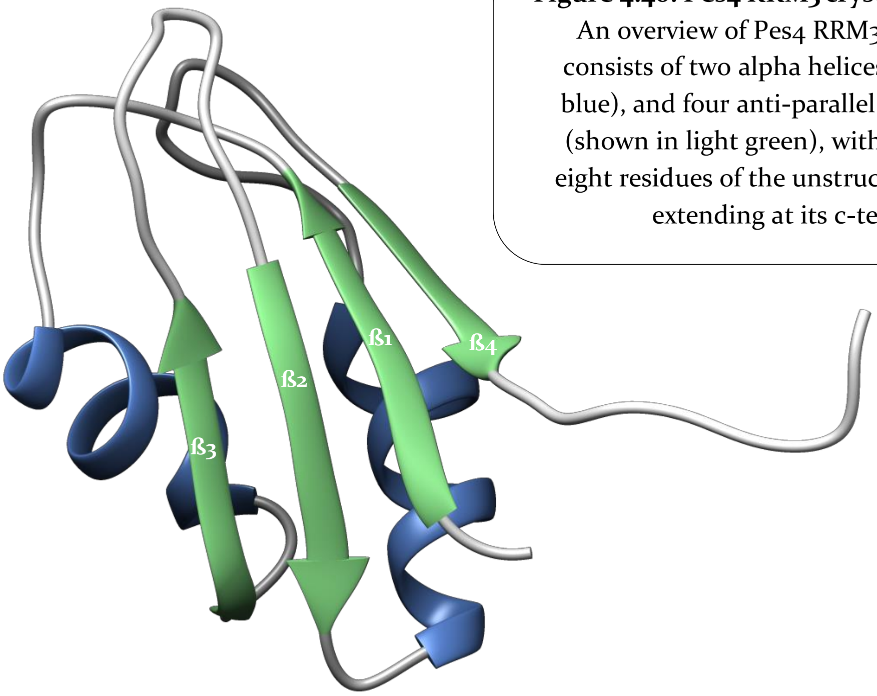


Figure 4.46. Pes4 RRM3 crystal structure. An overview of Pes4 RRM3 structure consists of two alpha helices (shown in blue), and four anti-parallel beta sheets (shown in light green), with additional eight residues of the unstructured linker extending at its c-term.

4.12.2. Crystal structure of Pes4 RRM4

4.12.2.1. Crystallization and data processing

Other types of crystals were obtained from the optimization with additive screens of some initial conditions. One crystal obtained in crystallization conditions containing 0.2 M Ammonium Nitrate, 20% PEG 3350, and 0.1 M Betaine Monohydrate as an additive, diffracted to a resolution of 1.1 Å. The crystal belonged to the orthorhombic space group $P2_12_12_1$. Data were processed using *XDS* and further scaled using *Aimless* from *CCP4* software. *Mathews coefficient* calculations indicated that the unit cell was big enough to fit only of the domains of the initial construct used to set the plates. Since the resolution obtained was high at 1.1 Å, we used *ARCIMBOLDO* program (D. Rodríguez *et al.*, 2012; D. D. Rodríguez *et al.*, 2009) to obtain phases and find a protein solution. The initial phases obtained from *ARCIMBOLDO* were then used for further model building using *Buccaneer* (Cowtan, 2006) . The structure obtained was that of Pes4 RRM4 with residues (391-469), possibly due to a proteolytic cleavage of the protein with only one molecule of Pes4 RRM4 present in the asymmetric unit. The structure model was further built in *COOT* and refined using *Phenix.refine*. Table 4.7 summarizes the final Data collection and refinement statistics of the structure.

It is most likely that the proteolysis event that led to the crystal formation of both Pes4 RRM3 and Pes4 RRM4 independently occurred during the crystallization process considering that the crystallization plates were set at 21°C for several weeks with the probability of the presence of proteases in any of the crystallization.

	Pes4 RRM4
Resolution (Å)	1.1
Space group	$P2_12_12_1$
Unit-cell parameters	
a,b,c (Å)	27.501, 36.162, 73.583
α, β, γ (°)	90, 90, 90
Wavelength (Å)	0.97921
Total reflections	376258 (37012)
Unique reflections	30588 (3013)
Completeness (%)	100
Multiplicity	12.3 (12.3)
Mean I/Sigma(I)	21.51 (3.64)
R_{merge}	0.05842(0.6737)
R_{meas}	0.06103
R_{work}/R_{free}	0.1464/0.1725
CC1/2	0.999(0.888)
Ligands	4
Water molecules	94
Protein residues	79
Ramachandran plot	
Favoured (%)	100
Outliers (%)	0
Clashscore	0
Average B factors	12.90

Table 4.7. Data collection and refinement statistics of Pes4 RRM4 crystal structure. Statistics from the outer shell are indicated in parenthesis.

4.12.2.2. Pes4 RRM4 shows a difference!

The structure of Pes4 RRM4 also has an RRM fold with four stranded anti-parallel β sheets, and two alpha helices packed against the β-sheet (figure 4.47). However, the very obvious and striking difference is the extended ß2 and ß3 with the shortened loop3 that links them. As mentioned earlier, and as seen in the previous RRM3 domain structures from Mip6, Pes4, and Sxl (figure 4.17, 4.46), this loop is usually long, extended, flexible loop that in some cases play a significant role in the RNA binding of the protein. Instead, in Pes4 RRM4 ß2 and ß3 are longer at the expense of the much shortened loop (loop3), but ß2 and ß3 typically forms the surface for RNA binding so it is possible that this feature might aid in extending the RNA binding surface. Consequently, loop 2 is longer in RRM4 and α1 is pushed to a different conformation (figure 4.48). It is not sure though if this feature also extends to the structure of Mip6 RRM4 since our attempts to solve its structure was not successful yet, or if this feature plays any role in the ability of the protein to bind Mex67. However, the sequence corresponding to this particular region is seen to be the least conserved between Mip6 RRM4 and Pes4 RRM4 sequence.

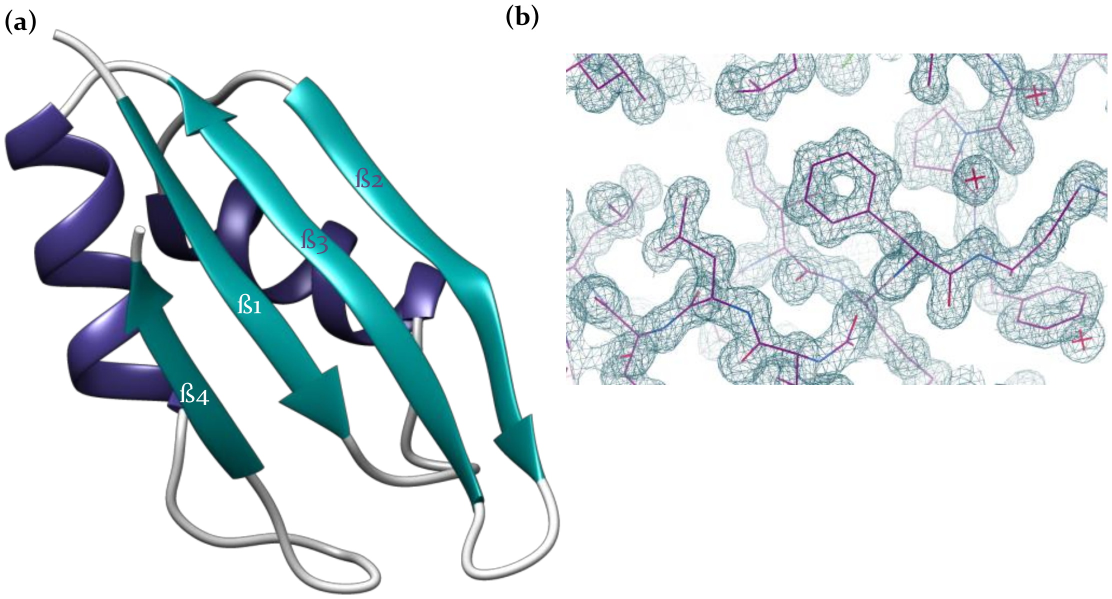


Figure 4.47. Pes4 RRM4 crystal structure. (a) An overview of Pes4 RRM4 structure consists of two alpha helices (shown in pink), and four anti-parallel beta sheets (shown in blue). ß2 and ß3 are visibly longer and more extended with a shortened small loop connecting them. (b) The quality of the obtained *2Fo-Fc* map contoured at 1 δ showing the electron density around amino acids with their hydrogen atoms and water molecules around it.

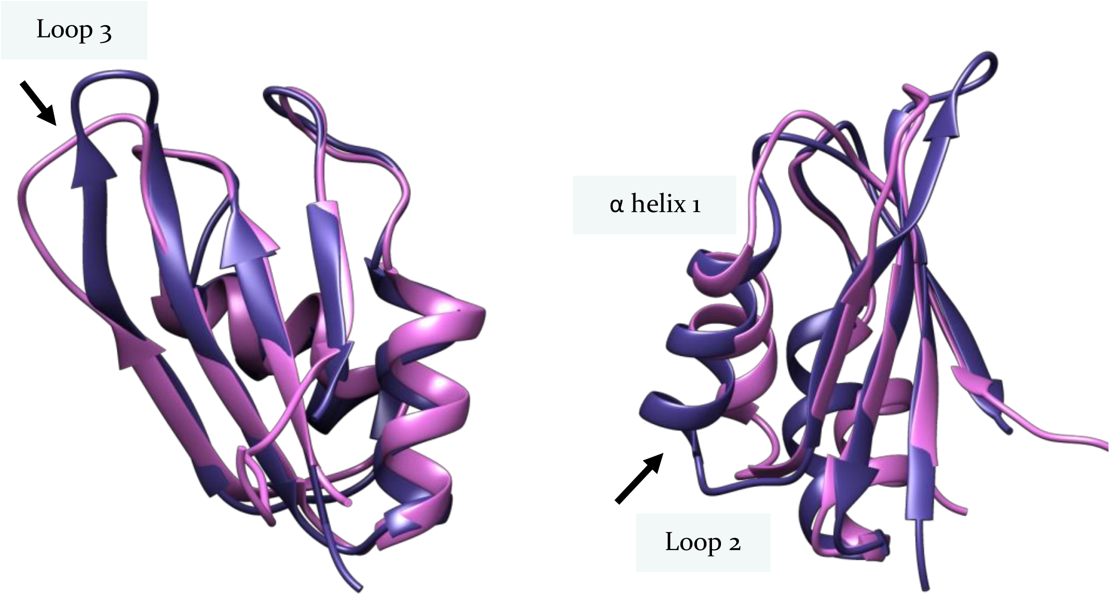


Figure 4.48. Superposition of RRM3 and RRM4 from PES4. The β2 andβ3 strands are more extended in RRM4 (in dark slate blue) than RRM3 (magenta), making loop 3 shorter and loop2 longer in Pes4 RRM4, while much longer and flexible loop3 seen in RRM3. This loop, however, along with ß2 and ß3 are said to be involved in the RNA binding in many RNA recognition motifs (RRMs).

4.12.3. Crystal structure of Pes4 RRM3/4

4.12.3.1. Crystallization and data processing

Additionally, some other crystals belonged to a distinct space group with larger unit cell dimensions. Some of these crystals were from the initially obtained crystals (figure 4.45) that were not unique. Growth condition amplifications and additive screens were carried out in an attempt to improve the crystals. The best crystals were obtained in 10% PEG 4000, 10% 2-Propanol, 100 mM Sodium citrate pH 5.6, or in 0.2M Ammonium sulfate, 0.1 M Sodium Cacodylate trihydrate pH 6.5, 30% PEG 8000, and diffracted up to 2.6 or 2.8 Å consecutively and belonged to the orthorhombic $P2_12_12_1$ space group. The data from the better diffracted crystal at 2.6Å were processed with *XDS* and further scaled with *Aimless* from CCP4. *Mathews coefficient* with solvent content analysis indicated unit cell volume sufficient for one molecule of Pes4 RRM3/4 in the asymmetric unit.

Before solving the individual structures of Pes4 RRM3 or Pes4 RRM4 (section 4.12.1 and 4.12.2), getting phases with molecular replacement using other PDB models was not successful. However, using the solved Pes4 RRM4 structure coordinates as a search model for molecular replacement was successful for obtaining phases and RRM4 could be modelled in the electron density map. Yet, obtaining phases for RRM3 and tracing it in the map was extremely challenging due to the flexibility and movement of RRM3 in the structure, especially that the diffraction data was poor. The electron density map around RRM3 was poor, especially for tracing long side chains. Additionally, few residues especially in the loop areas and in the linker connecting it to RRM4 were hard to trace or missing. RRM4 domain in the structure had more crystal packing contacts and therefore was stabilizing and maintaining the crystal, whereas RRM3 in the structure was involved in few crystal contacts and therefore more disordered and flexible. This flexibility was translated in high B-factors on RRM3, with average B-factors of 107.2 for RRM3 while 60.74 on RRM4 and also implicated high final R factors. TLS parameters were refined in multiple cycles with *Phenix refine* and manually tuned with *COOT*.

	Pes4 RRM3/4
Resolution (Å)	2.6
Space group	$P2_12_12_1$
Unit-cell parameters	
a,b,c (Å)	27.539, 50.29, 125.01
α, β, γ (°)	90, 90, 90
Wavelength (Å)	0.97921
Total reflections	70646 (7225)
Unique reflections	5766 (563)
Completeness (%)	100
Multiplicity	12.3 (12.8)
Mean I/Sigma(I)	15.01 (2.11)
R_{merge}	0.1331 (1.851)
R_{meas}	0.1391
R_{work}/R_{free}	0.2566/0.3402
CC1/2	0.999(0.763)
Protein residues	159
Ramachandran plot	
Favoured (%)	72
Outliers (%)	9

Average B factors	95.20

Table 4.8. Data collection and refinement statistics of Pes4 RRM3/4 solved structure.

4.12.3.2. Pes4 RRM3 and RRM4 have no contact between each other

The structure of Pes4 RRM3/4 as illustrated in figure 4.49. displayed two RRM domains, each consisting of a four-stranded anti-parallel β sheet backed by two α helices, arranged in tandem connected by a non-structured 15 residue long domain linker (residues 377-392). Few residues of the linker as it connects RRM3 are not visible in the structure though. Interestingly, the two domains do not interact with each other and are positioned in a way that there was no intra-molecular crystal contact between them. The two RRMs have their ß sheet platforms facing opposite directions in a V-shape like clamp formed by the linker between them. No intermolecular crystal contacts maintained between RRM3 and the linker neither. It is known from other few solved RRM structures, that the flexibility of the linker and that the subsequent positioning of both RRMs with respect to each other might change depending on the free form or the RNA bound form upon RNA induced conformational changes. This conformational change between multiple RRMs might also affect RNA topology upon binding which can be functionally important.

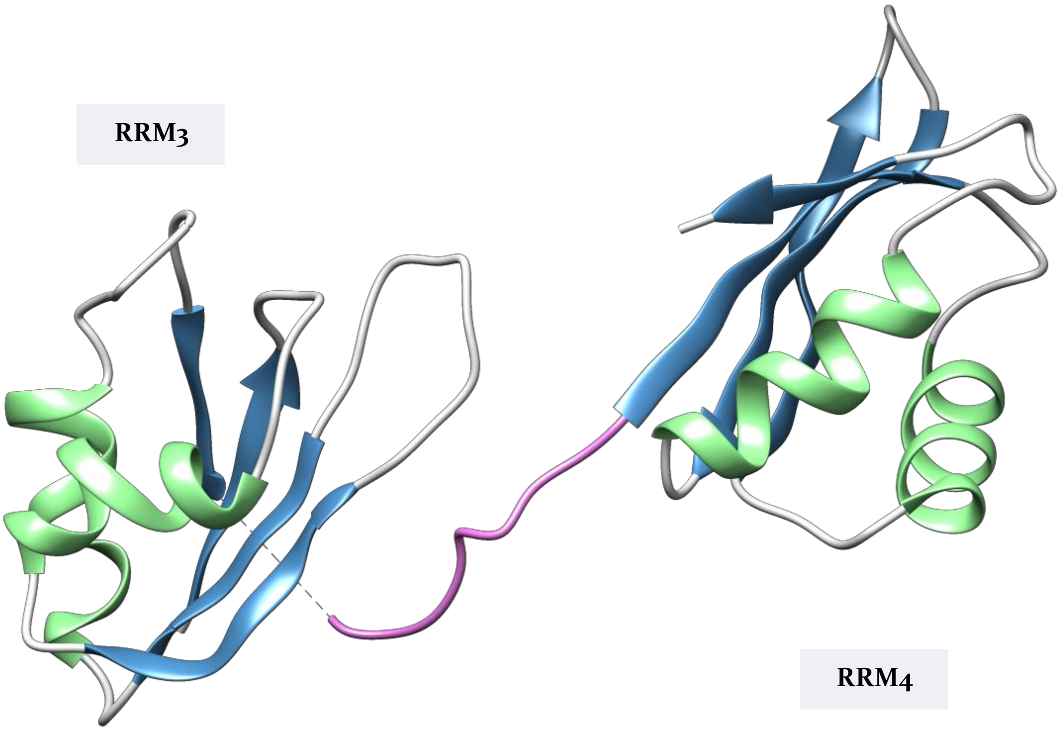


Figure 4.49. Overview of the solution structure of Pes4 RRM3/4 and its topology. Figure was generated using *Chimera*.

On the surface, the electrostatic potential calculated using the APBS server, showed a strongly electropositive inter-domain cleft on one side of the structure extending to the surface on one side of RRM3 and RRM4, while the other face of the structure was relatively neutral. This positivity on one side might point to a possible surface for RNA binding (Figure 4.50). Typically RRMs contain two conserved sequences called RNP1 and RNP2, on their ß2 and ß3 bearing hydrophobic and positively charged residues that respectively recognize the bases and negatively charged backbones of nucleic acids.

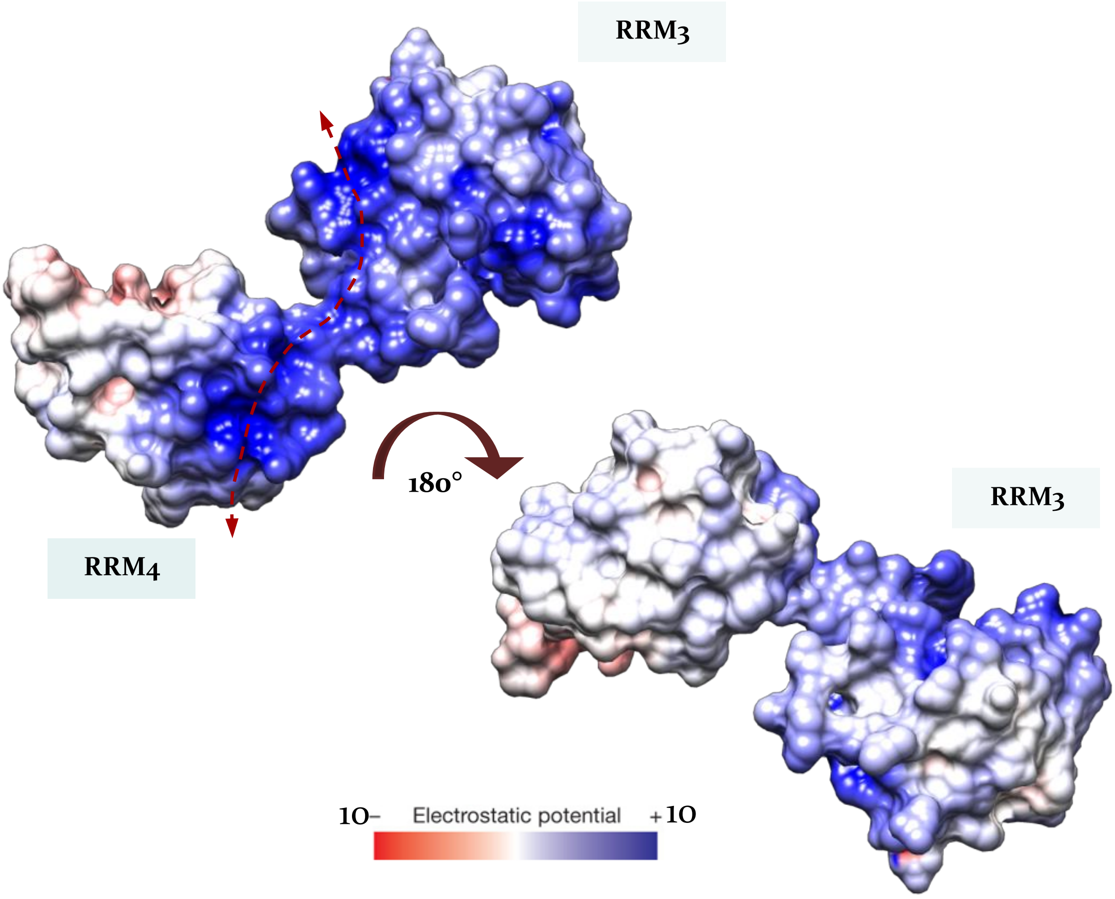


Figure 4.50. The surface electrostatic potential of Pes4 RRM3/4. The positive charge extends from RRM3 to RRM4 including the linker on one side, while the other side shows no positivity and is mostly neutral. The positive surface of the structure might point to the possible RNA binding interface. The electrostatic potential was calculated with PDB2PQR server followed by APBS and was visualized using *Chimera*.

4.12.4. Towards the crystal structure of Pes4 RRM3/4 bound to RNA

After demonstrating the ability of Pes4 RRM3/4 to bind RNA *in vitro* with good micro molar affinity, we attempted to crystallize Pes4 RRM3/4 with RNA bound. For this purpose, HPLC purified Poly(U) RNA of 11 residues was purchased from Sigma Aldrich and mixed in 1:1 or 1.2:1 molar ratio with Pes4 RRM3/4 before setting crystallization plates or mixed directly in drop. It has to be noted that mixing the protein with RNA before setting the crystallization plates caused a significant white precipitation. In that case, the sample was centrifuged before being added to the sitting drops. Few crystals in the form of very small, fragile plates appeared in a range of conditions at 21°C (figure 4.51, table 4.9).

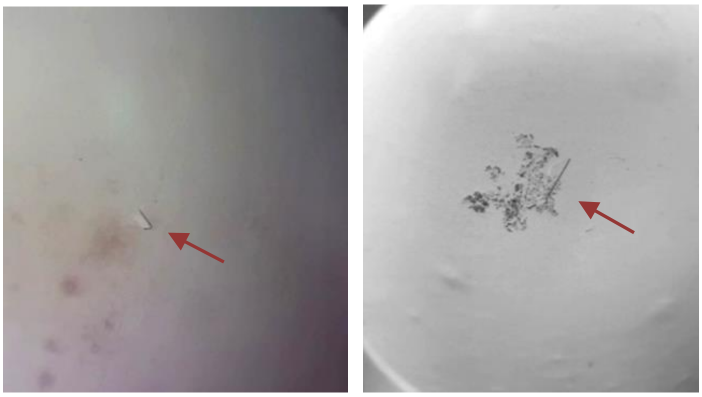


Figure 4.51. Crystals from Pes4 RRM3/4 when co-crystallized with Poly(U) RNA. The small plate like crystals obtained are pointed to by an arrow.

Growth conditions	Screen	Resolution	Unit cell dimensions
10% PEG 8000, 100 mM MES sodium salt pH 6.5, 200 mM Zinc acetate	JBSI (Jena bioscience)	1.95 Å	a,b,c (Å): 86.89, 94.32, 111.44 α, β, γ (°): 90, 94.76, 90
0.1 M Succinid acid, pH 7, 15% PEG 3350	JCSG-plus (molecular dimensions)	2.7 Å	a,b,c (Å): 87.79, 93.84, 110.50 α, β, γ (°): 90, 94.99, 90
0.15 m DL-Malic acid, pH 7, 20% PEG 3350	JCSG-plus (molecular dimensions)	2.3 Å	a,b,c (Å): 146.82, 83.29, 184.46 α, β, γ (°): 90, 90, 90
0.1 M Bicine pH 9, 10% PEG 6000	JCSG-plus (molecular dimensions)	3 Å	a,b,c (Å): 86.73, 93.504, 110.215 α, β, γ (°): 90, 94.718, 90

Table 4.9. Growth conditions where crystals were obtained in crystallization drops containing Pes4 RRM3/4 mixed with RNA, and the resolution to which the crystals were collected.

Some crystals diffracted poorly than others and data processing and scaling was challenging due to the low quality data. Both *XDS* and *iMosflm* were used for indexing and integration of the data. The data from the crystals mentioned in table 4.9 were integrated in either the monoclinic $P12_11$ (no. 3) or in I121 (c2, no. 5) space groups. Table 4.10 summarizes the data integration and scaling results from two different crystals.

	Crystal 1	Crystal 2
Resolution (Å)	1.95	2.7
Space group	$P12_11$	I121
Unit-cell parameters		
a,b,c (Å)	87.18, 94.63, 111.69	88.36, 94.58, 111.50
α, β, γ (°)	89.93, 94.82, 90.05	90, 94.99, 90
Wavelength (Å)	0.97925	1.07178
Total reflections	1828025 (40520)	110082 (15702)
Unique reflections	131270 (6509)	24831 (3322)
Completeness (%)	99.6	99.5
Multiplicity	6.3 (6.2)	4.4 (4.7)
Mean I/Sigma(I)	15.3 (2.1)	16.1 (2.3)
R_{merge}	0.351 (2.752)	0.802 (1.848)
R_{meas}	0.405	1.016
CC1/2	0.995(0.069)	0.994(0.205)

Table 4.10. Data collection and processing of Pes4 RRM3/4/RNA crystals. Statistics from the outer shell are indicated in parenthesis.

Preliminary Data analysis *Mathews coefficient* calculations suggested a unit cell where many copies of the RNA bound Pes4 RRM3/4 may be present in the asymmetric unit but it was not conclusive especially due to the uncertainty of the space group and whether the protein in the crystal was indeed bound to RNA. *Mathews* calculations indicated a high probability for 4-8 molecules in the asymmetric unit depending on whether it's a $P12_11$ or I121 crystal. This number of molecules was compatible with the results of the self rotation function calculated by *Molrep* from *CCP4* with a resolution cut off at 6 Å (figure 4.52). The data from either the $P12_11$ or I121 space group showed similar stereographic projections of the self rotation function that might be of two binary axis in addition to a 5 axis perpendicular to the crystallographic binary axis. However, the self rotation did not provide a definitive answer on the number of molecules in asymmetric unit neither, but might be indicating the presence of at least four or five molecules per asymmetric unit.

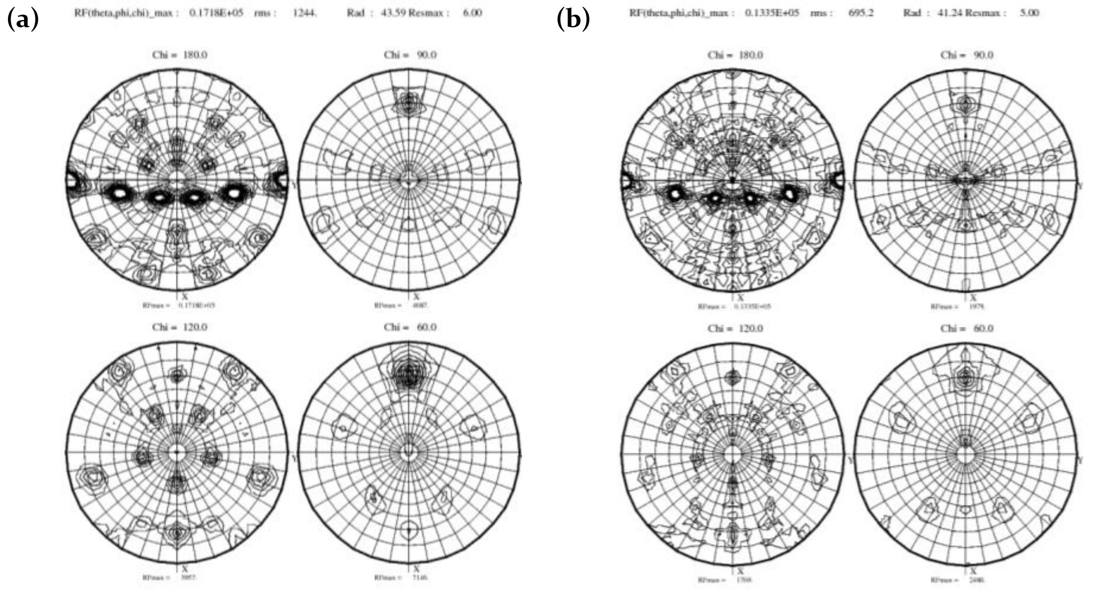
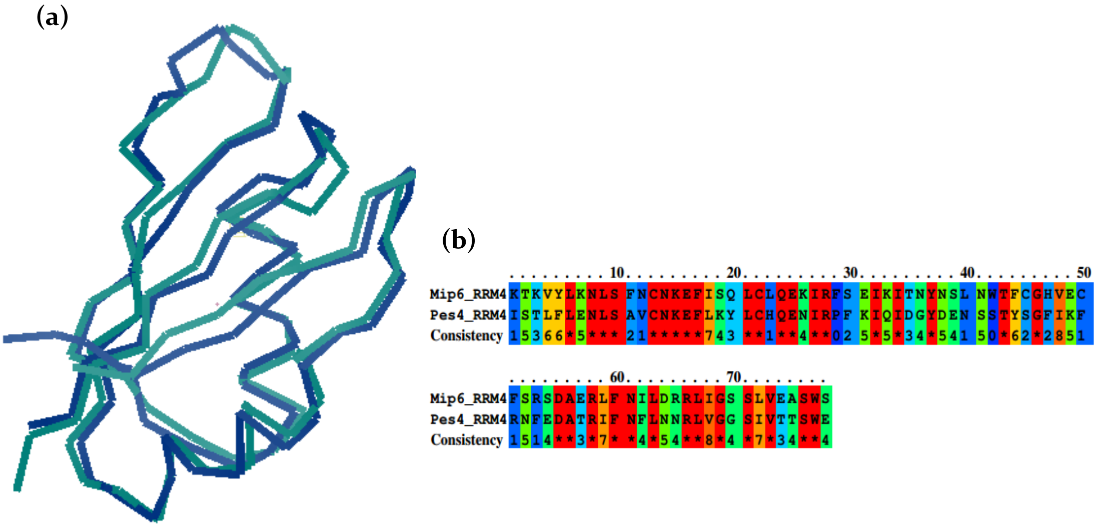


Figure 4.52. Stereographic projections of the self orientation function of the monoclinic crystals integrated in either $P12_11$ (**a**) and I121 space groups (**b**)

Unfortunately, the collected data did not allow us to obtain phases using molecular replacement even when the previously solved Pes4 RRM3/4 was used as a model possibly due to a different conformation or folding inside the crystal and the uncertainty about the angle between both domains in the crystal caused by the flexible loop, especially if it was indeed RNA bound and thus our effort for structure solution of Pes4 RRM3/4 bound to RNA was not until the moment successful. Meanwhile, attempts to obtain phases for structure determination are still ongoing and amplification plates for better crystals for future experimental phasing using heavy metals are attempted.

4.13. Mip6 RRM4 model building

Since our interest was more focused on the structure of Mip6 RRM4 as the binding platform for Mex67, and considering that our attempts to get Mip6 RRM4 crystals were not in the meantime productive, the structure of Pes4 RRM4 we solved earlier, as both are RRMs and share a good 42% of identical sequence, was used as a model to build a prediction model for Mip6 RRM4 structure. For this purpose, *I-TASSER* protein structure predication server was utilized, where Pes4 RRM4 structure was specified as a template. The overall predicted structure model showed a very similar fold as the Pes4 RRM4 with an R.M.S.D. value of 1.29 when both structures were superposed (figure 4.53).



```
          .........10.........20.........30.........40.........50
Mip6_RRM4 KTKVYLKNLS FNCNKEFISQ LCLQEKIRFS EIKITNYNSL NWTFCGHVEC
Pes4_RRM4 ISTLFLENLS AVCNKEFLKY LCHQENIRPF KIQIDGYDEN SSTYSGFIKF
Consistency 15366*5*** 21*****743 **1**4**02 5*5*34*541 50*62*2851

          .........60.........70........
Mip6_RRM4 FSRSDAERLF NILDRRLIGS SLVEASWS
Pes4_RRM4 RNFEDATRIF NFLNNRLVGG SIVTTSWE
Consistency 1514**3*7* *4*54**8*4 *7*34**4
```

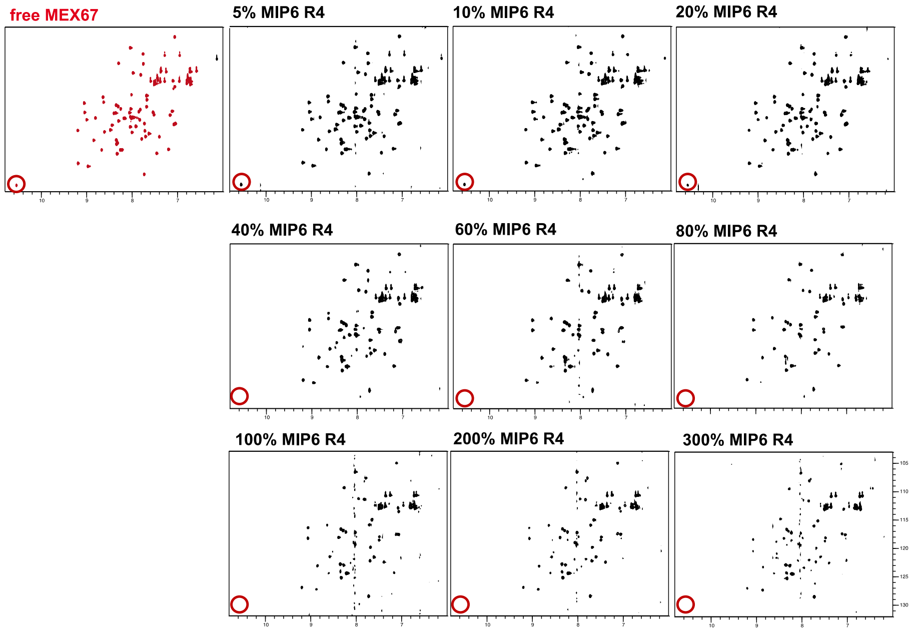
Figure 4.53. overview of Mip6 RRM4 predicted structure and its sequence conservation with Pes4 RRM4. (a) Superposition of Pes4 RRM4 structure (in cyan) with the Mip6 RRM4 structure predicted by I-TASSER (in blue). Illustration was generated with *COOT* **(b)** Sequence alignment of Mip6 RRM4 and Pes4 RRM4 using PRALINE multiple sequence alignment server where the results are color coded for amino acid conservation. The scoring scheme works from 0 for the least conserved alignment position, up to 10 (*) for the most conserved alignment position.

4.14. NMR titration analysis of Mip6 RRM4 upon complexation with Mex67$_{(528\text{-}599)}$

To elucidate further on the details of Mip6 RRM4 binding to Mex67, a chemical shift perturbation mapping by Nuclear Magnetic Resonance (NMR) spectroscopy was employed in collaboration with Dr. José Manuel Pérez Cañadillas (ICFR-CSIC, Madrid). Analysis of the chemical shift changes on the protein upon the addition of the protein partner and thus complex formation was used as a sensitive method to give us information on the binding interface between the two protein and defining residues that might be directly involved in the binding.

The heteronuclear single quantum coherence (HSQC) spectra assignment for the ^{15}N labelled Mex67$_{(528-599)}$ alone resulted in an good monomeric NMR structure calculations. Moreover, titration with increasing concentrations of Mip6 RRM4 resulted in obvious chemical shifts; some signals were lost due to chemical exchange while the remaining visible resonance peaks map was indicating amino acids not involved in the interaction. Our previously obtained crystal structure of Mex67$_{(528-599)}$ (figure 4.26) was used to interpret the obtained HSQC spectra and to map the obtained chemical shifts perturbations on the structure. The most notable resonance signal was that of Tryptophan Trp 575, the unique tryptophan residue in the Mex67$_{(528-599)}$ sequence, which was gradually lost as the concentration of Mip6 RRM4 increased, which signifies a high probable involvement of this amino acid in the interaction with Mip6 RRM4 (figure 4.54). Analysis of the chemical shift perturbation map indicated that these changes were mainly affecting the core alpha helix 2 of the Mex67 UBA domain and its N-term to a lower degree (figure 4.54) which coincides with the results obtained from the cross-linking experiment (figure 4.36) indicating that helix 2 of the UBA domain of Mex67 is central and crucial for maintaining the interaction with Mip6 RRM4.

(a)



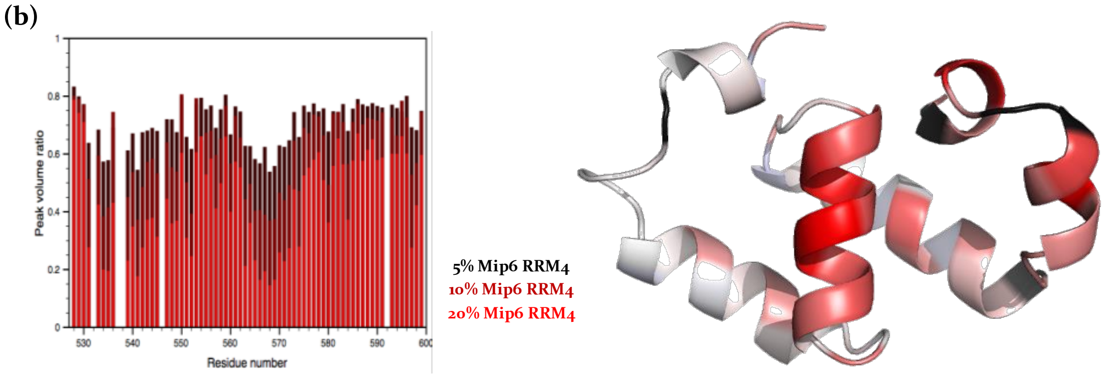


Figure 4.54. (a) NMR titration experiment showing the changes in the 1H-15N HSQC spectrum of free Mex67$_{(528\text{-}599)}$ upon the addition of increasing concentrations of Mip6 RRM4 until 300%. One of the most pronounced resonance changes was that of Trp 575 highlighted with a red circle. **(b)** Histogram of the overall chemical shift changes on Mex67$_{(528\text{-}599)}$ upon increasing Mip6 RRM4 concentrations, with the corresponding possible area of binding shown on Mex67$_{(528\text{-}599)}$ structure we solved shown in red.

On the other hand, GST-Mex67$_{(528\text{-}599)}$ was also titrated into Mip6 RRM4. The HSQC spectra of the ^{15}N labelled Mip6 RRM4 were then assigned and recorded in the free form of the protein and in its bound form upon the addition of increasing concentrations of the unlabelled Mex67$_{(528\text{-}599)}$ protein. During the titration, Shifting and displacement of amino acids resonance of the unbound state of Mex67$_{(528\text{-}599)}$ was visible. Figure 4.55 shows plots of chemical shift changes between free and Mex67-bound Mip6 RRM4. In the lower left part of the resonance spectrum, the resonance of two tryptophan indoles, that correspond to the two tryptophan residues in the Mip6 RRM4 sequence, are highlighted. A comparison of the spectra chemical shifts in the absence and presence of Mex67 clearly indicated that one tryptophan did not show any modification and its resonance from the free and Mex67-bound overlapped. On the other hand, the other tryptophan was unmistakably affected by the interaction as it showed a chemical shift changes and resonance position was displaced in the bound form.

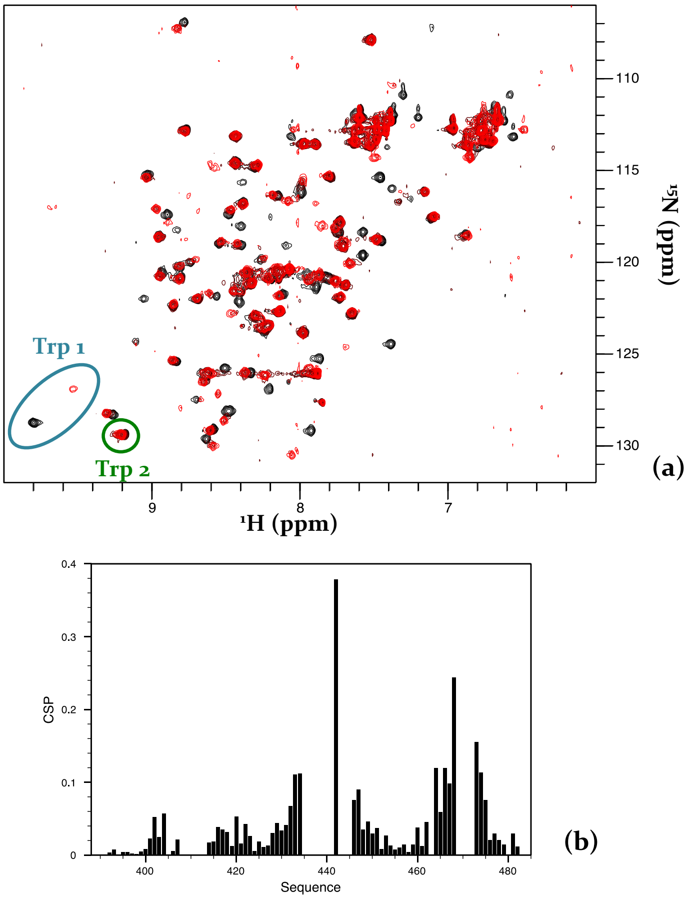


Figure 4.55. NMR titration experiment of Mip6 RRM4 titrated with Mex67$_{(528-599)}$. (a) Superposition of 1H-15N HSQC spectra of Mip6 RRM4 alone in the free form (in black) and upon the addition of increasing concentrations of GST- Mex67$_{(528-599)}$ (in red) where resonance displacements are visible. A selected area in the lower left region of the spectra is highlighted. It indicates the resonance of the two tryptophan indoles, Trp1 and Trp2 where one of them showed no shift at all while the other was clearly displaced. **(b)** Histogram representing the quantification of the chemical shift perturbations of Mip6 RRM4 individual amino acid resonances upon binding Mex67$_{(528-599)}$. Some resonances could not be observed in the spectrum of free and peptide bound Mip6 RRM4.

4.15. A Mip6 mutant that disrupts the interaction with Mex67

To investigate the effect of disrupting Mex67/Mip6 complex formation, we searched for a Mip6 RRM4 mutant that disrupts the interaction. Mip6 RRM4 sequence had two tryptophan residues, where one of them was clearly involved in binding with Mex67 as the NMR titration experiment have shown (figure 4.55). Aided by the information provided by the NMR titration experiment, and also by the information obtained from the chemical cross-linking experiment (section 4.8) that where in turn compatible with the NMR results, Trp 477 was excluded as a residue involved in binding, while Trp 442 was within the probable site of binding expected from the cross-linking/MASCOT experiment (figure 4.36). Additionally, Trp 442 was not conserved between Mip6 RRM4 and Pes4 RRM4 sequence, which does not bind Mex67, while the other Trp 477 was conserved as the structural and sequence alignment in figure 4.53 indicates and structural analysis of the Mip6 RRM4 structural model provided by I-TASSER (figure 4.56) showed Trp 442 as an exposed residue located in loop3 more probable to be involved in protein-protein binding. Therefore, we identified the Tryptophan 442 on Mip6 as a possible amino acid involved in the interaction with Mex67 and a good candidate for mutagenesis.

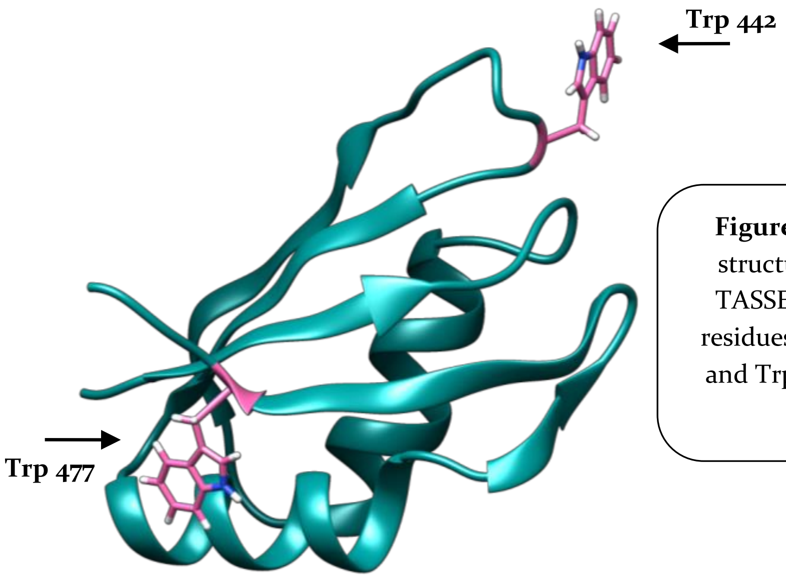


Figure 4.56. Mip6 RRM4 structure predicted by I-TASSER with tryptophan residues Trp 442 (in loop 3) and Trp 477 highlighted in pink.

4.15.1. Expression and purification of Mip6 W442A mutant

Subsequently, a mutant of Mip6 RRM4 where Trp442 is substituted by Alanine was constructed. The His-tagged mutant Mip6 RRM4 W442A was expressed in *E-coli* cells using IPTG induction in LB media as a protein expression system. The protein was expressed at 12 °C for 72 hours before the cells were collected. The soluble fraction of the protein was then purified with affinity chromatography on a 5 ml HisTrap column where the eluted fractions of the protein were dialyzed and the tag digested using TEV protease overnight. After that, the protein was additionally purified with size exclusion chromatography on Superdex 75 (16/60) gel filtration column. The Mip6 W442A mutant eluted from gel filtration in a single peak in the volume expected for its size. The W442A mutation on Mip6 RRM4 did not seem to affect or disrupt the folding of the domain as the yield of production of the protein did not decrease. The protein obtained was more than 90% pure as visualized on coomassie stained acrylamide gel (figure 4.57).

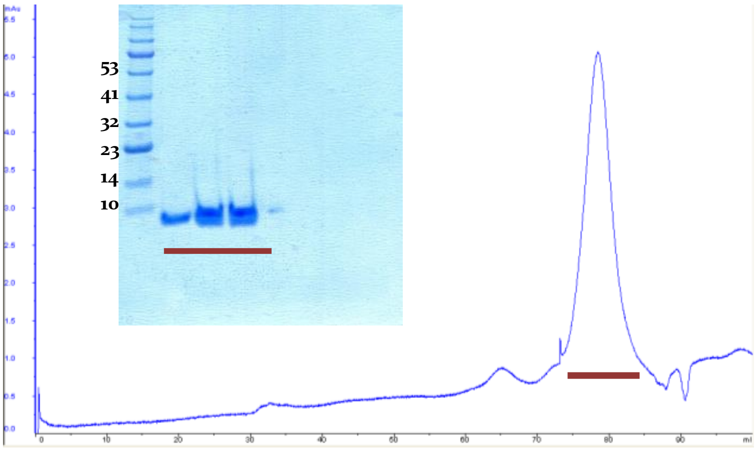


Figure 4.57. Gel filtration chromatogram representing the elution profile of Mip6 RRM4 W442A. The protein eluted in a single peak. The protein fractions corresponding to the peak were visualized on 10% SDS-PAGE gel and highlighted with a red bar.

4.15.2. Mip6 RRM4 W442A do not bind Mex67$_{(528-599)}$

Next, we tested the Mip6 RRM4 W442A mutant for its interaction with Mex67$_{(528-599)}$ using ITC technique. 50 µM of the purified Mip6 RRM4 W442A protein (section 4.15.1) in the sample cell was titrated with the same concentration of 690 µM of Mex67$_{(528-599)}$ used previously to assess its binding to the wild type Mip6 RRM4. No binding was detected and the peaks of heat were only those resulting from the dilution effect (figure 4.58). The experiment was repeated and the result was confirmed, the point mutation on Mip6 RRM4 where the Tryptophan 442 was mutated into Alanine was enough to abolish the complex formation and render Mip6 incapable of interacting with Mex67 or at least decrease its binding affinity to undetectable values.

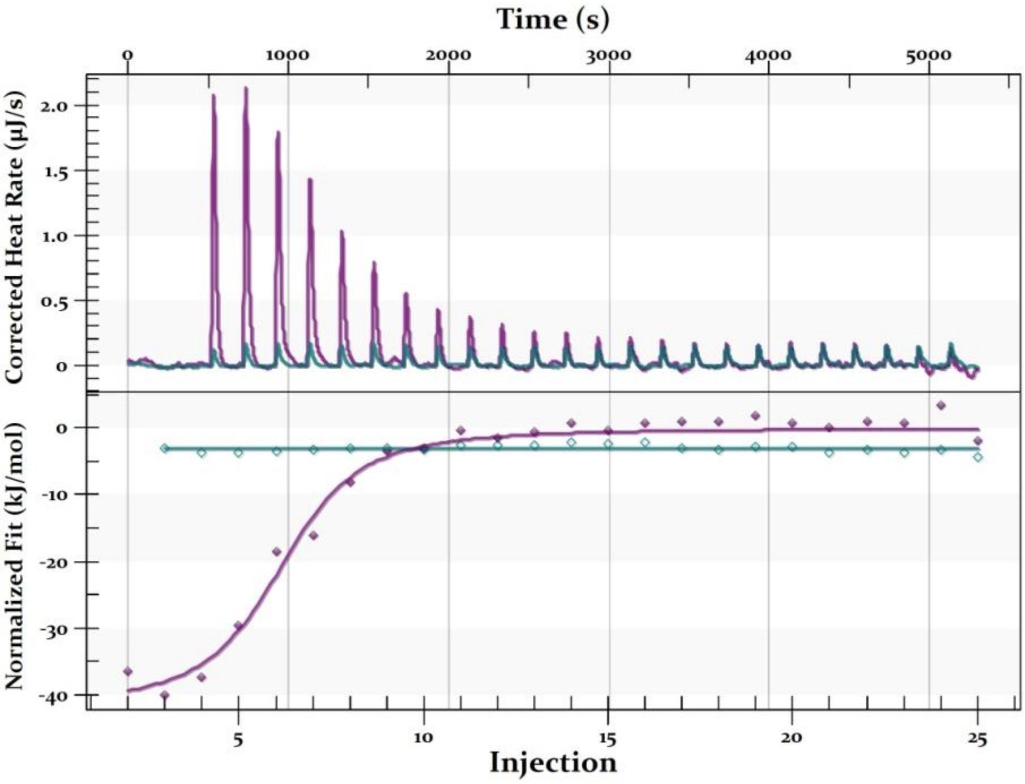


Figure 4.58. Mip6 RRM4 W442A mutant lost the ability to bind Mex67$_{(528-599)}$. The graph shows the ITC raw heat peaks from Mex67$_{(528-599)}$ when titrated into Mip6 RRM4$_{(390-480)}$ in violet or when titrated into Mip6 RRM4 W442A in green, and the corresponding curves obtained from fitting and analyzing the obtained data using Nanoanalyze. A flat no-binding line corresponds to the data obtained from Mip6 RRM4 W442A.

Additionally, the NMR titration analysis and HSQC spectrum assignment when the Mip6 RRM4 W442A was titrated into Mex67$_{(528-599)}$ confirmed the result where no resonance displacements were observed on the residues of Mex67$_{(528-599)}$ structure, previously observed with Mip6 RRM4 wild type titration (figure 4.55). The resonance peaks overlapped perfectly between the free form and the Mip6 RRM4 W442A titrated form of Mex67$_{(528-599)}$ (figure 4.59).

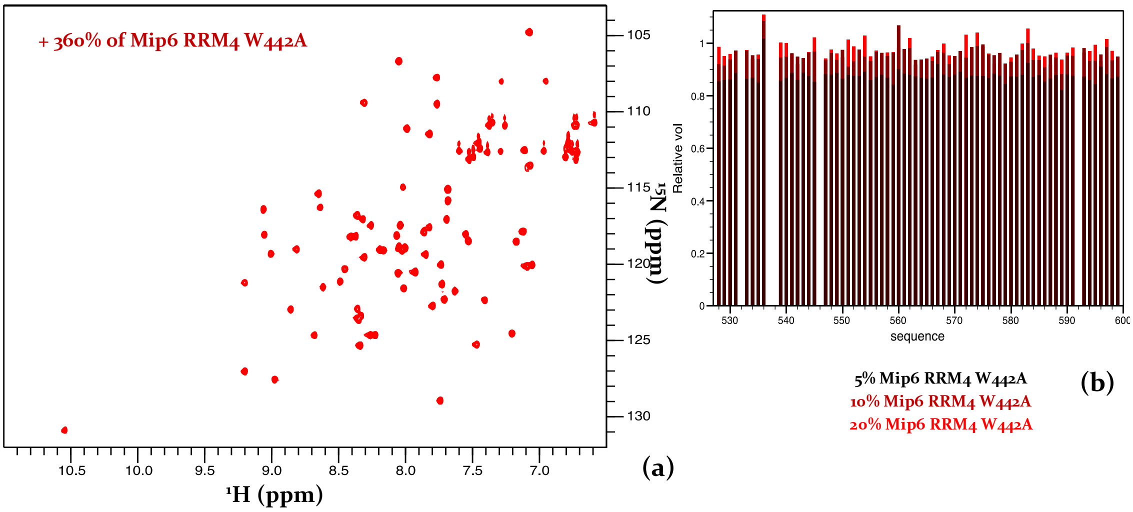
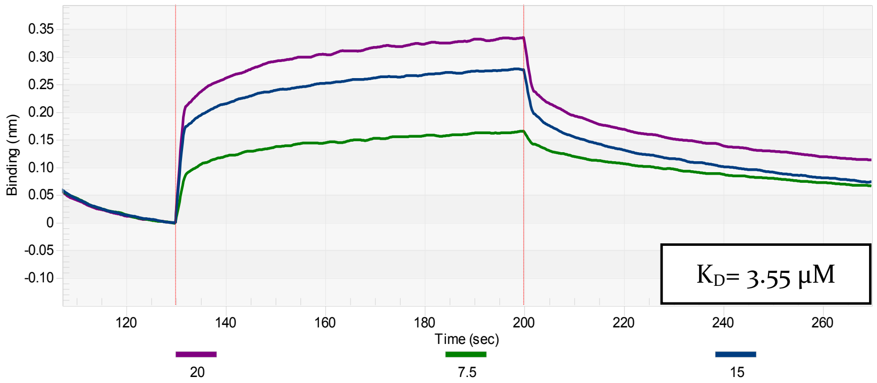


Figure 4.59. NMR titration assignment of Mip6 RRM4 W442A titrated into Mex67$_{(528-599)}$. (a) 1H-15N HSQC spectra of the free form of Mex67$_{(528-599)}$ (in black) and upon the addition of increasing concentrations of Mip6 RRM4 W442A up to 360% (in red) showed no resonance displacements and the peaks perfectly coincided. **(b)** Histogram representing the chemical shift perturbations of Mex67$_{(528-599)}$ individual amino acid resonances upon increasing concentrations of Mip6 RRM4 W442A.

4.15.3. Mip6 RRM4 W442A mutant disrupts the ability to bind Mex67 but not RNA

Finally, to inspect whether the mutation that rendered Mip6 RRM4 unable to bind Mex67$_{(528-599)}$ might also affect its ability to bind RNA we resorted to the Bio-layer interferometry system where a biotinlylated 15 residues of Poly(U) were immobilized on streptavidin sensors, while different concentrations of Mip6 RRM4 W442A was used as prey protein. Interestingly, the mutation did not affect the ability of Mip6 RRM4 to bind RNA as the Mip6 RRM4 W442A protein was still able to interact with the Poly(U) with comparable affinity even when it has lost the ability to interact with Mex67. The binding affinity of Mip6 RRM4 W442A to Poly(U) calculated was K_D ~3.55 µM, a similar binding affinity of that of Mip6 RRM4 indicating no loss or decrease in binding affinity to RNA (figure 4.60). Additionally, no big difference in either association or dissociation constant was evident.



Construct	**Binding affinity (K_D)**	**Ka/Kd**
Mip6 RRM4$_{(401-480)}$	2.8 μM	3.47 e4/ 9.75 e-2 M
Mip6 RRM4 W442A	3.55 μM	1.31 e4/ 4.67 e-2 M

Figure 4.60. Mip6 RRM4 W442A binds RNA. (a) Bio-layer interferometry association and dissociation steps shows binding profile to the sensor immobilized Poly(U) when using increasing concentrations of Mip6 RRM4 W442A shown in the figure in different colours. **(b)** Table comparing the binding affinities calculated for Mip6 RRM wild type or Mip6 RRM4 W442A mutant to Poly(U) RNA with the association (ka) and dissociation constant (Kd) values in each case indicted.

5. Discussion

The proteins subject of this study: Mex67, Mip6, and Pes4 are yeast (*Saccharomyces cerevisiae*) proteins. Mex67 is a well-studied protein, due to its evolutionary conserved cellular importance where its function is critical to proper mRNA export specifically and to rRNA export to a lower degree as recently discovered and consequently cell survival. However though, much about Mex67 interaction dynamics and the overall nuclear RNA export mechanism is still to be learned and better understood. On another hand, practically very little, if anything, is known about the structure and function of either Mip6 or Pes4. Therefore our focus in this study was on learning more about Mip6, Pes4, and studying the Mip6/Mex67 relation for a better understanding not only of these proteins, but also of the consequent role in cellular function and the dynamics of mRNA export mechanism.

5.1. Mip6 and Pes4 protein expression and solubility

In the course of this study, we faced difficulties in the aim to express and purify the full length Mip6 and Pes4 proteins. For Mip6, the presence of the N- and C- terms, outside the ordered RRM domains, displayed very poor solubility and also aggregation in standard buffers. Both prokaryotic and eukaryotic expression systems were thus approached involving the addition of a range of additives, but only tagging the protein with a MBP tag known to enhance protein solubility was helpful in getting a slight amount of soluble Mip6 FL protein. However, even the small soluble quantity obtained was aggregating after size exclusion chromatography. But with our over-expression and purification approaches, we successfully obtained soluble protein from shorter constructs of both Mip6 and Pes4 proteins containing a combination of the ordered RRM domains or even the four RRM domains all together in case of Mip6 which constitutes around 56% of the full length sequence in good quantities in the prokaryotic system. Thus, it is tempting to think that the disordered, unstructured, regions on Mip6 especially on its N-term and C-term (figure 4.2) were greatly affecting its solubility and tendency to aggregate or it might be simply that the *in vivo* expression and folding of the protein is aided by chaperones that we could not reconstruct *in vitro*. Additionally, the long disordered linkers connecting the small RRM domains of the protein made it more prone to proteolysis and degradation. However, our expression systems and purification techniques used in this study were suitable and proved successful for obtaining good amounts of soluble truncated proteins which allowed us to analyze structural and functional properties of both proteins even if it was not the full length protein.

5.2. Mip6 and Pes4 are RNA binding proteins with multiple RNA recognition motifs

A common feature in the diverse RNA binding proteins (RBPs) is containing RNA binding domains (RBDs) also known as RNA recognition motifs (RRMs) or ribonucleoprotein domain (RNP). This domain, which recognizes RNA sequences (canonically single stranded RNA), and binds them, is the most abundant RNA-binding domain in eukaryotes (Clé Ry, Blatter, Dé Ric, & Allain, 2008; Maris, Dominguez, & Allain, 2005). The diversity of RBPs and the number of highly dynamic RNA binding domains they contain renders them multifunctional, playing dynamic roles in the different steps of RNA biogenesis and thus regulating gene expression. When it comes to mRNA export, this feature is of extreme importance as the fate of mRNA, whether to be exported and therefore translated or not, is dependent on the proper formation of mature messenger ribonucleoprotein (mRNP).

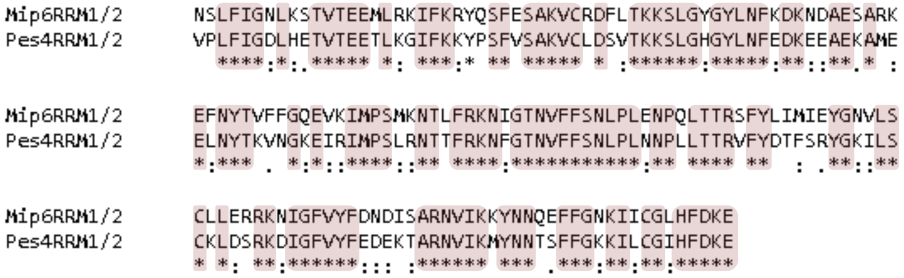
In the very few literature references on Mip6 or Pes4, the proteins were usually referred to as putative RNA binding proteins containing three RRMs (Bolognesi *et al.*, 2016; Jin, Zhang, Sternglanz, & Neiman, 2017; Segref *et al.*, 1997). In the only two papers on the subject, The “predicted” RRM domains of Mip6 or both Mip6 and Pes4, are said to be critical for their function: formation of cytoplasmic foci containing RNA when Mip6 is over-expressed (Bolognesi *et al.*, 2016), and the role of both Mip6 and Pes4 in mRNA protection and localization in stress conditions (Jin *et al.*, 2017). However though, no one had actually demonstrated the ability of these proteins to bind RNA *in vitro* or *in vivo* or looked closer at the number and diversity of RRMs they contained.

We are the first to actually demonstrate, by using biochemical and biophysical techniques (Poly(U) agarose beads pull downs and Bio-layer interferometry system), the RNA binding properties of Mip6 and Pes4 proteins where they physically bind RNA with high affinity. Although we used a general Poly(U) RNA to measure the binding, since until the moment the recognition or preference for any specific RNA motif sequence is not yet known, it still showed a good micro-molar or even nano-molar affinities. That indicates not only that these proteins can physically bind RNA, but also that they are capable of binding non-specific RNA with high affinity, that might differ in case of any specific RNA sequence recognition knowing that RRMs are usually highly plastic domains.

Although we did not put much emphasis on the first two RRMs of Mip6 or Pes4, since we were more interested with the later two as the interaction platform of Mex67, these first two RRM sequences showed the most conservation in terms of sequence with 66% comparing sequences of Mip6 and Pes4 (figure 5.1). A PDB search looking for sequence similarity in the data bank solved structures gave us the highest score with percentage identity (36%) with Mip6 RRM1/2 for RRMs 1 and 2 of the human Poly(A) binding protein (PABC) (UniProtKB - P11940). For this reason, crystallization trial of Mip6 RRM1/2 were carried out with Poly(U) and Poly(A) but none was successful until the moment, and in many occasions Poly(A) double helix crystals were obtained.

However, it seems that the addition/mix of the Mip6 RRM1/2 protein with Poly(A) in the crystallization drop was crucial for the crystallization of the Poly(A) in the form of a double helix as crystals were obtained only upon the addition of the protein and never with Poly(A) on its own. A similar case is seen in the paper where the Poly(A) double helix structure was previously published. Just like in our case the Poly(A) duplex structure was obtained while the group was attempting to crystallize a similar RRM containing fragment of a Poly(A) binding protein (PABP) bound to RNA, and also in their case obtaining crystals required the presence of the protein fragment. It might be therefore suggested that *in vitro* and possibly *in vivo*, the occurrence and function of Poly(A) duplexes might be regulated by the added protein that may be playing the role of a chaperone for the Poly(A) thus shaping it and enhancing its duplex formation, possibly in a concentration related manner. This might be relevant under cellular conditions where accumulation of Poly(A) occurs such as cytoplasmic granules formation upon cellular stress.

Additionally, Safaee *et al.* suggested that the $NH4^+$ ions in their structure had a role in stabilizing the duplex and provided evidence for the duplex increased thermal stability upon the addition of $NH4^+$ (Safaee *et al.*, 2013). It is true that also in our case, the Poly(A) double helix crystals were obtained mostly in crystallization conditions where high concentration of Ammonium was present, however it did not seem to make part of our structure. On the contrary, the electron densities located where Safaee group suggested ammonium ions, were more probably for water molecules instead.



```
Mip6RRM1/2      NSLFIGNLKSTVTEEMLRKIFKRYQSFESAKVCRDFLTKKSLGYGYLNFKDKNDAESARK
Pes4RRM1/2      VPLFIGDLHETVTEETLKGIFKKYPSFVSAKVCLDSVTKKSLGHGYLNFEDKEEAEKAME
                  ****:*:.***** *: ***:* ** ***** * :******:*****:**::**.* :

Mip6RRM1/2      EFNYTVFFGQEVKIMPSMKNTLFRKNIGTNVFFSNLPLENPQLTTRSFYLIMIEYGNVLS
Pes4RRM1/2      ELNYTKVNGKEIRIMPSLRNTTFRKNFGTNVFFSNLPLNNPLLTTRVFYDTFSRYGKILS
                *:*** . *:*::****::** ****:***********:** **** **  : .**::**

Mip6RRM1/2      CLLERRKNIGFVYFDNDISARNVIKKYNNQEFFGNKIICGLHFDKE
Pes4RRM1/2      CKLDSRKDIGFVYFEDEKTARNVIKMYNNTSFFGKKILCGIHFDKE
                * *: **:******::: :****** *** .***:**:**:*****
```

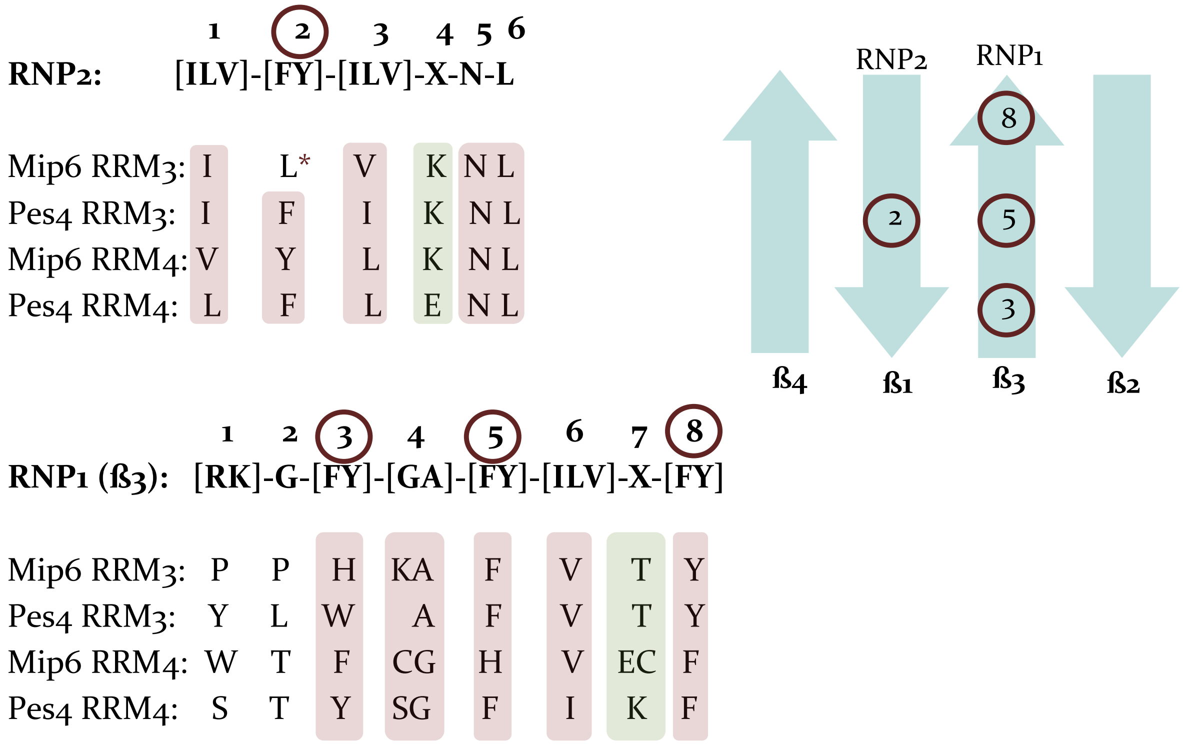
Figure 5.1. Sequence alignment between Mip6 RRM1/2 and Pes4 RRM1/4 with 66% identical sequence identity.

On the other hand, we managed to solve several structures from Mip6 or Pes4: Mip6 RRM3$_{(313-389)}$, Pes4 RRM3$_{(377-384)}$, Pes4

RRM4$_{(391-469)}$, Pes4 RRM3/4$_{(303-473)}$. All turned out to be a typical ordered RRM structures that adopt a typical ß1α1ß2ß3α2ß4 topology formed by four anti-parallel ß sheets and two α helices packed against the ß sheets.

Canonically, RRMs contain two conserved sequences located on their central ß sheet of the domain, named RNP1 (on ß3) and RNP2 (on ß1) exposing three aromatic side chains on the surface that would form the most common RNA binding platform of an RRM. The two aromatic rings in RNP2 (position 2) and RNP1 (position 1) are traditionally involved in the stacking of two deoxynucleotides, while the third aromatic residue located in RNP1 (position 3) often interacts with the sugar rings of a dinucleotide. However though, not all residues of RNP2 and RNP1 are conserved in all RRMs. Some RRMs will have only one or two exposed aromatic residues on their central ß sheet (Allain, Bouvet, Dieckmann, & Feigon, 2000). The most frequent conserved residues, though, are the aromatic residues in RNP2 position 2 and RNP1 position 5.

Even though we couldn't get any RRM structure with RNA bound yet, looking closer at the RNP2 and RNP1 sequences in RRMs 3 and 4 from Mip6 or Pes4 we see that the RNP2 sequence in all is conserved in all 6 positions except for Mip6 RRM3 missing a key aromatic residue in position 2 (figure 5.2). As for RNP1, it also showed a good level of general conservation. Additionally, loops in an RRM can be sometimes also involved in RNA recognition and can be used to modulate the binding affinity and specificity of the RNA, and as we have seen in the structures we solved (Mip6 RRM3, Pes4 RRM3), these loops are usually long and flexible, adopting variable conformations (figure 4.17). But how these Mip6 or Pes4 RRMs exactly bind RNA, whether on the most preferred canonical beta sheet surface or not-so-common non-canonical way involving other mechanism, is still to be learned and we are working on it.



RNP2:	[ILV]	[FY]	[ILV]	X	N	L
	1	2	3	4	5	6
Mip6 RRM3:	I	L*	V	K	N	L
Pes4 RRM3:	I	F	I	K	N	L
Mip6 RRM4:	V	Y	L	K	N	L
Pes4 RRM4:	L	F	L	E	N	L

RNP1 (ß3):	[RK]	G	[FY]	[GA]	[FY]	[ILV]	X	[FY]
	1	2	3	4	5	6	7	8
Mip6 RRM3:	P	P	H	KA	F	V	T	Y
Pes4 RRM3:	Y	L	W	A	F	V	T	Y
Mip6 RRM4:	W	T	F	CG	H	V	EC	F
Pes4 RRM4:	S	T	Y	SG	F	I	K	F

Figure 5.2. A closer look on the RNP2 and RNP1 sequences in RRM3 and RRM4 of both Mip6 and Pes4, and a scheme showing the organization of the key conserved aromatic residues on the central ß sheet of the RRM. A general conservation in the sequences is learned, except for Mip6 RRM3 missing the key aromatic residue on the second position of RNP2.

5.3. Pes4 RRM3/4 structure and the role of multiple RRMs

Although the RRM domain is one of the most existent domains in eukaryotes, only a handful of structures for tandem RRMs in complex with nucleic acid are solved. A big part of the reason might be the fact that individual RRMs not only comprise flexible loops (figure 4.17), but also usually connected by long flexible linkers, which in general contribute to these domains high plasticity and dynamicity but also makes them more difficult to crystallize.

In the Pes4 RRM3/4 structure we solved, the RRM3 and RRM4 do not contact each other. The linker connecting both domains is extended and disordered keeping the two RRMs tumbling independently in a V-shaped cleft (figure 4.49). Of course, RRMs can function alone, or in tandem giving them an extended binding interface for the recognition of longer RNA sequences cooperatively.

From the few crystal structures available for free or RNA bound two tandem RRMs, the free tandem RRMs frequently convene an open conformation with no-interdomain contacts with a disordered linker as seen also in Pes4 RRM3/4 structure (figure 4.49). Yet, this conformational change when RNA is bound takes a more closed and fixed conformation bringing the two RRMs closer together and making the linker more ordered. This is true for most protein structures where tandem RRMs were solved bound to RNA, like in sex lethal (Sxl) protein (PDB ID: 3SXL; 1B7F), human RNA binding HuR protein (PDB ID: 4EGL; 4ED5), or even the Poly U binding (Pub) protein (Crowder, Kanaar, Rio, & Alber, 1999; Li *et al.*, 2010; Wang *et al.*, 2013; Yokoyama *et al.*, 1999) .

This is interesting since it gives the domains plasticity due to the range of new possible conformations, as the protein recognizes different RNA molecules with different sequence specificity, length, or binding affinity. However, when RRMs are separated by a very long disordered linker making two patch or RRMs clearly independent, like in the case of the big distance between RRM1/2 and RRM3/4 of Mip6 or Pes4, might indicate a possible recognition of different RNA sequences and perhaps increase the probability of the protein encountering their different binding sequences (Cléry & Allain, 2013). It is plausible that the presence of a very long flexible linker between RRM1/2 on one side and RRM3/4 on the other, in addition to the other disordered inter-domain linkers and flexible domain loops might also be a contributing factor to the fact that we couldn't get crystals of the Mip6 construct containing all its four RRMs (Mip6$_{(111\text{-}480)}$) until the moment.

Bearing this in mind, we tried to crystallize Mip6 RRM3/4 and Pes4 RRM3/4 with added RNA in an attempt to get a more ordered stable structure when RNA bound. Crystals from Pes4 RRM3/4 bound to Poly(U) were obtained. The crystals shape was distinct from those of free Pes4 RRM3/4 (needles, monoclinic) and belonged to another space group (orthorhombic). Even so, we were not able to obtain phases to solve the structure especially that the resolution was poor at 2.7 Å in some crystals and many copies were to be expected in the asymmetric unit. Nevertheless, ongoing amplifications to obtain better crystals for heavy metal phasing are being carried out.

Additionally, sometimes the cooperation between two RRMs to bind RNA might mean more contribution in RNA binding of one RRM than the other. For example, As seen in the RRM-bound HuD RRM1 and RRM2 protein structure (Wang *et al.*, 2013), the RRM1 primarily recognizes the RNA (PDB ID: 4ED5). The consequent RNA-binding induced conformational changes in the structure bring RRM2 and the inter-domain linker in contact to the RNA substrate. So, while RRM1, the initial recognition RRM, interacts with five RNA nucleotides making it the critical RNA binding domain, RRM2 only recognizes two nucleotides stabilizing the complex.

This, however, might also be the case of Mip6 RRM3 as it could explain the weaker binding of individual Mip6 RRM3 seen to the Poly(U) agarose beads with correspondence to the strong Mip6 RRM4 band co-eluting with Poly(U) from the beads as visualized on SDS-PAGE (figure 4.11,c). Also, what supports this hypothesis is, as discussed in figure 5.2, the fact that Mip6 RRM3 sequence is missing a key, typically conserved, aromatic residue of RNP2 position 2 (ß2) while having only one fully exposed aromatic residue in RNP1 (ß3) regardless of having other exposed residues on ß2 and ß4 respectively (figure 5.3). So in case RRM3 binds RNA in a canonical manner, this might have an effect on the number of nucleotides it would be able to stack on its beta surface. A similar case is seen in the complex structure of RNA-bound nucleolin RRM1/2 (Allain *et al.*, 2000), where nucleolin RRM2 is also missing an exposed aromatic residue in RNP2 position 2. Upon RNA binding, while RRM1 contributed to most of the nucleotide interactions (six nucleotides), RRM2 in the structure only interacted with two nucleotides. On the other hand, looking at Mip6 RRM4 sequence and its predicted structure we obtained using I-TASSER, it had three exposed residues of RNP1 and RNP2 respectively.

This is also consistent with the high affinity of Mip6 RRM3/4 to RNA we measured using the Bio-layer interferometry that calculated a binding affinity of 0.318 µM, with an RNA binding affinity of 5.7 µM for Mip6 RRM4 alone (figure 4.12) (Wang *et al.*, 2013).

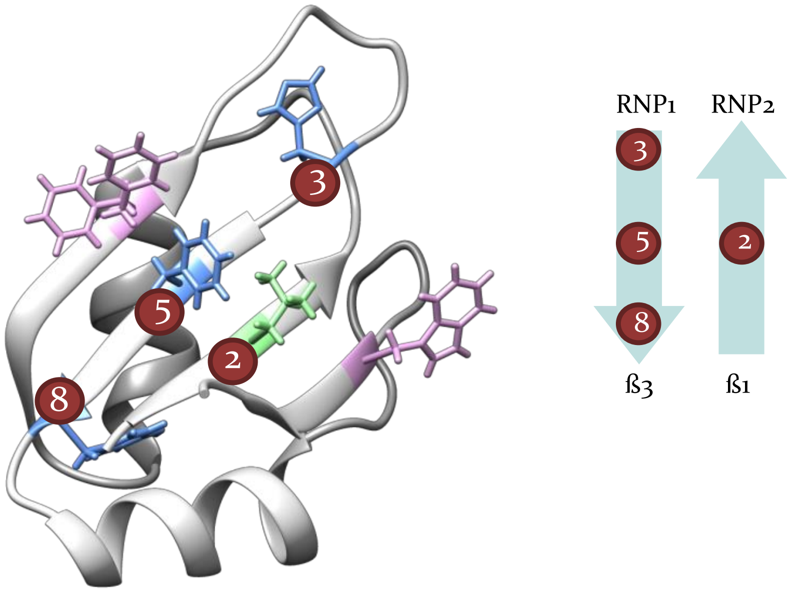


Figure 5.3. Aromatic residues exposed on Mip6 RRM3 beta sheet. The three conserved aromatic residues on RNP1 and their position are shown in blue, where only one of them exposed to the solvent (position 5). On the other hand, no aromatic residues are found on ß1 in the RNP2 sequence, instead RNP2 position 2 harbours a leucine instead of the typical conserved aromatic residue. Other aromatic residues found on ß2 and ß4 not part of the RNP1 and RNP2 are shown in pink.

5.4. Mip6 RRM4 is the binding surface for Mex67

Contrary to the previously held thought, we demonstrated that Mip6 has a fourth RNA binding domain instead three using different techniques: with Poly(U) agarose beads pull down experiments and also with Bio-layer interferometry (section 4.3.4 and 4.3.5). Thus, we characterized a novel RRM4 on Mip6. Moreover, we proved this particular RRM4 is also the site of binding of a protein, Mex67 (section 4.5.3). In the unique study where the Mip6/Mex67 interaction was first tackled using the yeast two hybrid system, they suggested that the RNA binding of Mip6 has nothing to do with its interaction with Mex67 since a Mip6 construct lacking all the putative RNA binding domains was still able to interact with Mex67 (Segref *et al.*, 1997).

However, now we know that Mex67 actually binds an RNA binding domain of Mip6 that was not previously recognized. Also, the fact that $\text{Mex67}_{(528\text{-}599)}$ eluted together with Mip6 RRM4 from size-exclusion chromatography column, a technique rather used for protein separation, in a single peak indicated that the two proteins were holding together tightly in a stable complex with rather good affinity (figure 4.24).

Additionally, using ITC biophysical technique we showed that the obtained low micro molar binding affinity of Mex67 to Mip6 RRM4 was almost equivalent to that of $\text{Mip6}_{(111\text{-}480)}$ containing all four RRMs (figure 4.25), so we conclude that Mip6 RRM4 is sufficient to achieve the interaction with Mex67 . We then looked closer at the ITC kinetic parameters to get an overview on the type of interaction we would expect between the two proteins. In the ITC interaction between Mip6 RRM4 and Mex67, neither enthalpy nor entropy appeared to be dominating the interaction as the enthalpy (ΔH) and entropy (ΔS) changes both gave negative values although ΔH was slightly more negative (figure 4.25). However, In ITC experiments where $\text{Mip6}_{(111\text{-}480)}$ was used instead of Mip6 $\text{RRM4}_{(390\text{-}480)}$ in the cell, or when GST-Mex67 UBA was used instead of $\text{Mex67}_{(528\text{-}599)}$ in the syringe, the entropy change (ΔS) was positive while ΔH was negative. So it seems that the interaction might probably be a more polar protein-protein interaction or maybe a mixture of polar and hydrophobic interactions.

Although RRMs are known for their RNA binding abilities, recently structural and functional studies showed that some RRMs are capable of protein binding as well. Yet, the protein recognition had no general mechanism of interaction. In general, the RNA-binding RRMs that are involved in protein-protein binding use different platforms on their structure to bind both the RNA and the protein simultaneously (Nagai, Price, & Evans, 1998; Rideau *et al.*, 2006).

On another hand, a class of RRM domains is involved in protein binding but not RNA binding. This class of non-RNA binding RRMs either have their ß sheet surface exposed but generally lack the conserved aromatic residues on RNP1 and RNP2 sequences important for RNA connection like the non-canonical U2AF homology motif family (UHM) (Kielkopf, Lücke, & Green, 2004), or have similar sequences to canonical RRMs but achieves protein interaction through its entire ß sheet instead of RNA interaction . This is not the case of Mip6 RRM4 as it has conserved aromatic residues in RNP1 and RNP2 sequences like canonical RRMs and we also proved that it is indeed involved in RNA binding and protein binding *in vitro* with high affinity.

5.5. Pes4 RRM4 binds RNA but not Mex67

Mip6 and Pes4 are homologs sharing ⍰ 40% sequence identity (figure 4.39), and are said to be paralogs that arose due to an ancient yeast whole genome duplication (Byrne & Wolfe, 2005; Wolfe & Shields, 1997). While Mip6 was binding to Mex67 through its RRM4 as we showed and discussed earlier, the homologous Pes4 RRM4 did not have the ability to interact with Mex67. It is not clear, however, whether Mip6 RRM4 have evolved its ability to interact with Mex67 separately or is it that Pes4 that have lost it. On all cases our finding suggest, that although both Mip6 and Pes4 have all four ordered RRM domains in the structure, and share the function of RNA binding, they did not share the ability of their RRM4 to bind Mex67 which went exclusively to Mip6 RRM4 domain. It might be criticized that Mex67 perhaps may bind on another region on Pes4 not necessarily its RRM4. This, of course, could be taken as a possibility that we consider but hugely doubt if we take into account the high sequence identity between the two proteins, the high conservation of their domain organization, their secondary structure organization, and the fact that RRM4 is a well ordered domain as seen in Pes4 RRM4 structure that shares 42% of identical sequence with Mip6, while other domain like RRM1 and RRM2 had an even higher degree of sequence conservation (figure 5.1). If this is indeed the case, it would suggest that Mip6 and Pes4 evolved their sequence for divergent functions through time. Future experiment using the full length Pes4 to check interaction with Mex67 would then set the case to rest, however the difficulty of getting a soluble full length Pes4 prevented us from carrying out the experiment at the moment.

Additionally, a protein cleavage due to proteolysis that led to the independent crystallization of Pes4 RRM3 and Pes4 RRM4 seems to have most probably occurred during the crystallization process and not before, thus excluding the possibility of any proteolytic event that might have affected the function of the protein in Mex67 binding, especially that from the same purified protein sample, crystals of the whole Pes4 RRM3/4 were obtained.

5.6. Mip6 RRM4 Mex67 binding affects its ability to bind RNA

Since Mip6 RRM4 can bind both nucleic acids (RNA) and Mex67 protein as we showed with similar affinity, the imperative question was whether the Mip6 RRM4 bind both partners simultaneously or perhaps competitively. For this reason, we checked the ability of a pre-formed Mip6 RRM4/Mex67$_{(528\text{-}599)}$ to bind to Poly(U) agarose beads *in vitro* (section 4.9, figure 4.37). The pre-formed complex did not elute with the Poly(U) from the beads indicating no binding of Mip6 RRM4 to the Poly(U) once the Mex67 was bound, although Mip6 RRM4 on its own, as a control, was clearly binding. The experiment was repeated several times with different boundaries of Mip6 RRM4 complex with Mex67$_{(528\text{-}599)}$ with the same results obtained. On the other hand, when Mip6 RRM4 was added first to the Poly(U) agarose before adding Mex67$_{(528\text{-}599)}$, only Mip6 RRM4 co-eluted with Poly(U). This finding was suggesting that whoever was binding first to Mip6 RRM4 was affecting its ability to bind the other. A native gel electrophoresis with a pre-formed Mip6 RRM4/Mex67$_{(528\text{-}599)}$ and increasing concentrations of RNA also pointed to the same direction, as no band shift was seen upon adding increasing concentration of RNA (figure 4.38).

A point mutation W442A in Mip6 RRM4 loop3, however, disrupted the interaction with Mex67 but not with RNA (section 4.15.3). This severe interaction disruption caused by a single mutated Tryptophan 442, indicates that this exposed polar and aromatic residue might be involved directly in the interaction with Mex67, which suggests its possible location in the direct site of binding. This Tryptophan however, is not located on an ordered alpha helix or beta sheet, instead it is located on the loop3 (connecting ß2 and ß3) which as we discussed earlier is usually flexible in movement and diverse in length that might have an effect on the function of the RRM domain. In some RRMs this loop might even be involved in RNA binding.

To our knowledge, there is not any known protein-binding RRM that binds protein in this manner. As discussed in section 5.4, the protein-binding RRMs bind the protein and RNA simultaneously, bind protein through RNA, or are protein-binding RRMs that lost the ability to bind RNA, but an RRM that interacts with both RNA or protein not competitively but rather where one conditions the binding of the other is not, to our knowledge, characterized yet in literature. In the first case, the protein and the RNA usually bind simultaneously on opposite interfaces of the RRM. The RNA canonically and stably interacts with the ß sheet platform as we explained earlier especially as the hydrophobic aromatic residues on RNP1 and RNP2 are conserved (as in the case of Mip6 RRM4), on the other hand the protein typically interacts with the α helices positioned behind the ß sheet as illustrated in Figure 5.4. For all the structures that we know of, this is the case for RRM structures binding both RNA and proteins as seen in the structure of the nuclear Cap-binding protein complex for example, or that of the splicesomal U2B”-U2A’ protein complex bound to small nuclear RNA illustrated in figure 5.4, 5.6) (Mazza, Segref, Mattaj, & Cusack, 2002; Nagai *et al.*, 1998) .

This does not seem to be the case of Mip6 RRM4 however. Mip6 RRM4 was certainly not binding Mex67 through RNA, as Mex67$_{(528-599)}$ was not binding RNA on its own as our results show (figure 4.29, 4.39), and both proteins were still physically interacting in the absence of added RNA or even when RNase was added. Moreover, looking at the Mip6 RRM4 predicted structure; the alpha helices interface is quite distant from the position of the Trp442 on the loop3 to the opposite direction. The measured distance between Trp 442 and the closest residue from the first alpha or the second alpha helix in the domain is approximately 16.74 and 19.44 Å respectively. So, apart from the non-simultaneous binding to Mex67 and RNA, it seems like the mechanism of binding itself might be different as the canonical RNA binding site on an RRM is very close to the expected Mex67 binding site than other characterized protein binding sites.

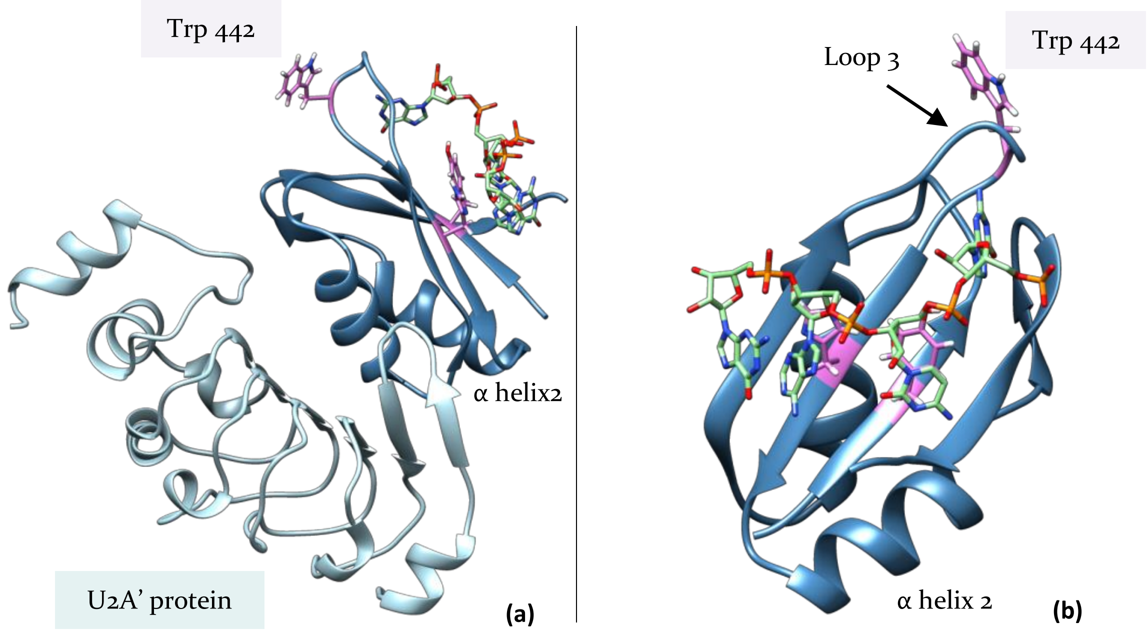


Figure 5.4. The protein-binding RRM interfaces for binding both a protein partner and RNA. **(a)** The crystal structure of U2B"-U2A' potein complex bound to a fragment of U2 small nuclear RNA. In the illustration, a model compisition where the U2B" RRM is replaced with Mip6 RRM4 (dark blue) in the structure to show how would it apply to Mip6 RRM4 as the folding in both is highly conserved. As shown in the structure, U2A' interacts with U2B" hydrophobically through its alpha helix 1 majorly, while the RNA binds to the RRMs' ß sheet on the opposite side. Additionally, the two exposed aromatic residues in RNP1 and RNP2 on the central beta sheet (highlighted in pink),conserved in both U2B" RRM and Mip6 RRM4 , as shown in the structure, are directly involved in the interaction with two deoxynucleotides of the RNA. **(b)** A closer look

on the RRM4 central ß sheet that constitute the binding platform for RNA. In both figures, the Tryptophan Trp 442 located on loop3 is highlighted in pink.

Combining the results of the chemical cross-linking experiment (figure 4.36), and the comparison of sequence alignment with Pes4 RRM4 in which the conserved residues with Pes4 RRM4 are excluded (figure 4.53), also comes in line with this hypothesis indicating a Mex67-binding interface on Mip6 RRM4 involving loop3, in close proximity to the expected RNA binding surface. Interestingly, this loop, loop3, as seen in Pes4 RRM4 structure which did not have the ability to bind Mex67 is very short and is stable with no room for flexibility. Additionally, as we demonstrated earlier that Trp442 mutation to Alanine would have a tremendous effect on the complex formation, which reflects its central role in the interaction, and Trp442 is not conserved in Pes4 RRM4 sequence. Actually, comparing the aligned Mip6 RRM4 and Pes4 RRM4 sequences, a distinctive overall low degree in sequence conservation is noticed in the area of ß2 and ß3 and loop3 connecting them as figure 5.5 shows. All this information given from different experimental results indicates the likelihood that the location of this residue on a loop may reflect that this loop flexibility might play a role in the ability of Mip6 RRM4 to recognize or bind Mex67 on one hand, and also the ability to interact with RNA on the other due to the close proximity of the expected binding areas for the partners.

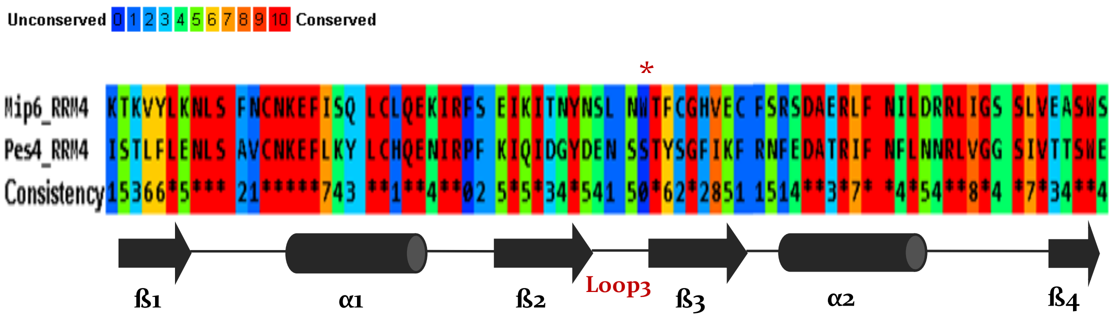

Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved
*
Mip6_RRM4 KTKVYLKNLS FNCNKEFISQ LCLQEKIRFS EIKITNYNSL NWTFCGHVEC FSRSDAERLF NILDRRLIGS SLVEASWS
Pes4_RRM4 ISTLFLENLS AVCNKEFLKY LCHQENIRPF KIQIDGYDEN SSTYSGFIKF RNFEDATRIF NFLNNRLVGG SIVTTSWE
Consistency 15366*5*** 21*****743 **1**4**02 5*5*34*541 50*62*2851 1514**3*7* *4*54**8*4 *7*34**4
ß1
α1
ß2
Loop3
ß3
α2
ß4

Figure 5.5. Sequence alignment between Mip6 RRM4 and Pes4 RRM4 with the secondary structure arrangement as seen in Pes4 RRM4 crystal structure. The alignment residues are coloured from the least conserved residues up to the most conserved ones. As visible, the area between ß2 and ß3 including loop3 shows very little sequence conservation, unlike the alpha

helices regions. The Trp 442 in Mip6 RRM4 is highlighted with a star where it shows no conservation.

Thus, If what we suggest is the case, this indicates that Mip6 RRM4 is a novel RRM demonstrating a distinctive RRM-protein interaction manner, in which the protein partner, not only binds the RRM on newly characterized interface, but also located on a close proximity to the RNA binding site, so that the binding of the RRM to either the protein or the RNA as a first partner conditions its ability to bind the other especially when both Mex67 and Poly(U) RNA had very similar micro molar binding affinities to Mip6 RRM4.

But how exactly Mex67 binding to Mip6 RRM4 is disrubting it interaction with RNA or vice versa? Of course, a future detailed structural studies would clarify this point but unfortunetly until the moment our crystallization trials to get a Mip6 RRM4/Mex67 complex structure were not yet succesful. Nevertheless, one possibility is that the close proximity of the expected binding surface for Mex67 on one hand, and RNA on the other hand as illustrated in the figure 5.4 and figure 5.6.b might indicate a partial overlapping of binding sites of both partners. This would be supported by the fact that loop3 in most RRM structures, and as seen also in Mip6 RRM3 and Pes4 RRM3 structure is a flexible, dynamic, and changing conformations, where some conformations might be favorable for specific RNA binding while others not. Also as seen in figure 5.6.a, sunstituting U2B” RRM in the structure with Mip6 RRM4 structure shows clearly the different conformations a long loop3 can adopt regardless of the conserved overall structure. Knowing the level of high plasticity RRMs generally have, we suggest that the binding of one partner whether it is Mex67 or RNA might induce some conformational changes, probably involving loop3, on Mip6 RRM4 rendering the binding of the other partener less favorable, espcially when the binding sites are such of close proximity (figure 5.6.b).

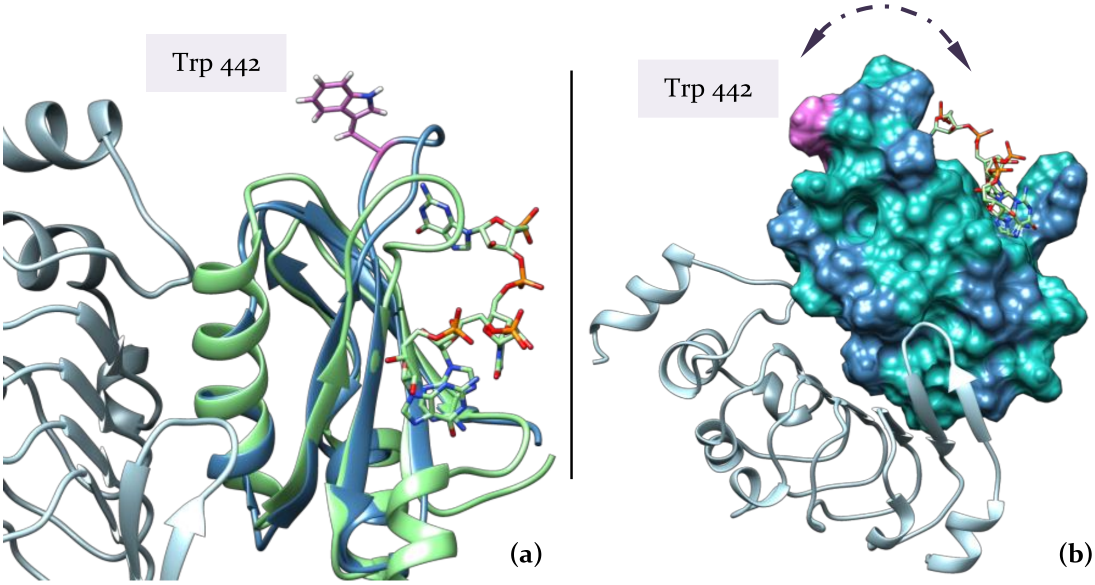


Figure 5.6. The flexible loop of Mip6 RRM4 and the close proximity to the canonical RNA-binding site. **(a)** Superposing U2B" structure with Mip6 RRM4 shows the different loop3 conformations **(b)** Mip6 RRM4 surface (in blue) with hydrophobicity shown in green. The hydrophobic site of canonical RRM protein-binding site where U2A' is binding, on the other hand the RNA binding site is also pointed to, whereas the position of loop3 with Trp442 highlighted in pink shows how it's possible different conformations may interfere with RNA binding or vice versa.

In addition to the dynamics of Mip6 RRM4 domain itself, another possibility would be an equilibrium between a monomer/homodimer association of Mip6 RRM4. This equilibrium may help regulate the function of the domain, increasing specificity or affinity or both for one partner but not the other. Some RRM domains are known to form homodimers generating extended binding surface to increase their specific RNA recognition properties or/and increasing the binding affinity too (Lunde, Moore, & Varani, 2007; Ye, Malinina, & Patel, 2003). A support for this hypothesis comes from the primarily NMR spectroscopy done in collaboration with Dr. Jose Manuel Perez Canadillas protein NMR group, used for assigning Mip6 RRM4 structure. The NMR assignment of the spectrum was poor and not complete, as seen on 15N-1H HSQC spectrum of the Mip6 RRM4 domain shown in figure 5.7. The profile of the resulting peaks might be indicating a possible dimerization or a possible monomer/dimer exchange. Future experimentation, however, might give us better answers.

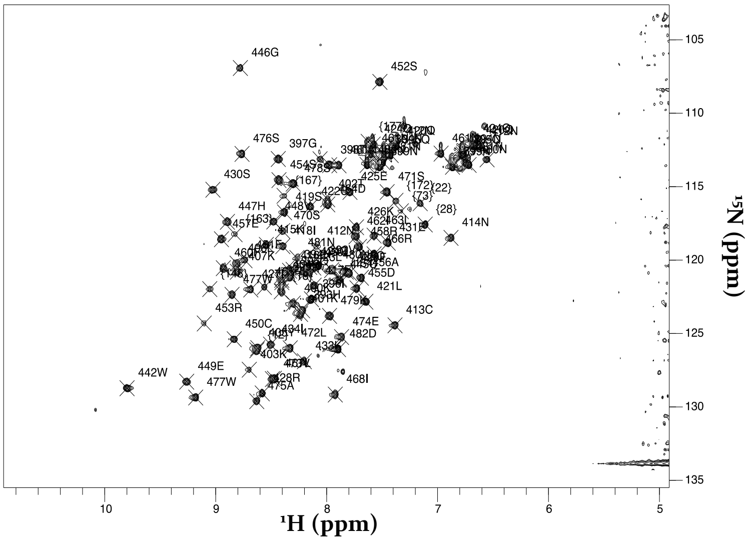


Figure. 5.7. 2D 1 H,15N HSQC NMR spectra of Mip6 RRM4. Resonance peaks are annotated to show the backbone amide assignments. Not all the residues have been successfully assigned.

Additionally, the ITC kinetics for Mip6 RRM4/Mex67$_{(528-599)}$ complex, where Mex67$_{(528-599)}$ was titrated into 50 µM of Mip RRM4 in sample cell was calculating a stoichiometry coefficient of n=0.965 ~ 1 which indicates a 1:1 binding ratio between both proteins. However, when the concentration of Mip6 RRM4 used was increased to 70 µM, the stoichiometry coefficient calculated was around n=0.72-0.8 which now indicates a different ratio than 1:1. Also, reversing the proteins location during the ITC experiment so that the higher concentration of Mip6 (700 µM) was titrated into 50 µM of Mex67 lead to a calculation of n ~1.6 which is consistent with the previous finding. In both cases the binding affinity for Mex67 was mildly lower than when lower concentration of Mip6 RRM4 was used. This, in addition to the NMR assignment profile of Mip6 RRM4 led us to suspect a possibility of a higher level of association in Mip6 RRM4 that might be regulating its function.

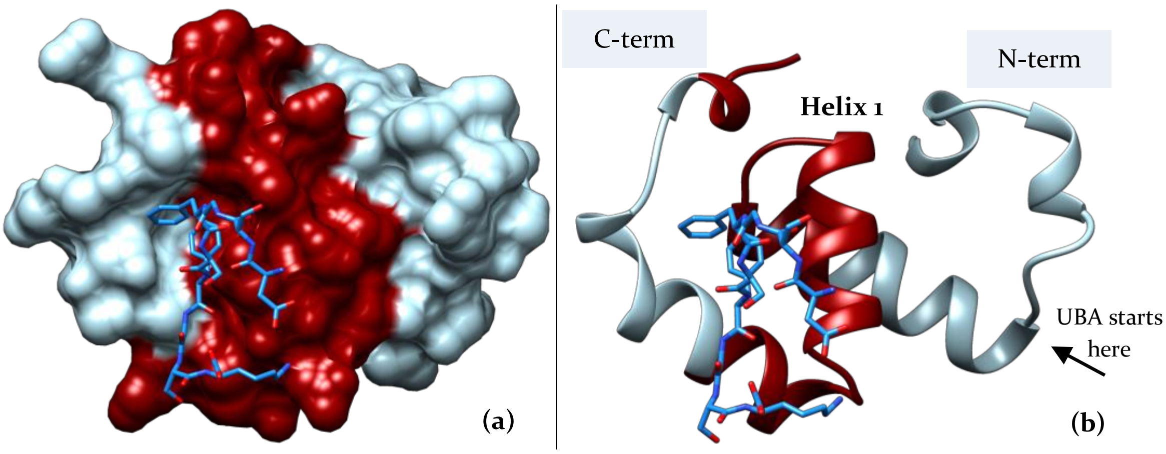
5.7. Mex67 UBA domain gets an additional partner

The survival of the cell depends on the proper function of Mex67, since it is the major mRNA exporter, conserved from yeast to humans (J. Katahira *et al.*, 1999). One critical domain of Mex67 is its C-terminal UBA not only required for a proper mRNA export, but also for the pre-ribosomal subunit export and also contributes to the recruitment of Mex67 to the TREX complex (Gwizdek *et al.*, 2006). Although the crystal structure of the human Mex67 UBA was previously solved (PDB ID: 1OAI), and recently from *Chaetomium thermophilum* fungal species (PDB ID: 4WP2), we succeeded in solving the crystal structure of yeast (*Saccharomyces cerevisiae*) Mex67 UBA with additional N-terminal 18 residues of the linking area connecting the NTF2 like domain to the UBA domain.

The structure showed a general conservation where the UBA domain was formed by four alpha helices, while its N-term was made of two additional small alpha helices. The highest degree of conservation was in the core alpha helices. Comparing Mex67 UBA domain to that of the other species, showed the most structural conservation with the human homolog. On the other hand, despite the conservation of the first three helices, the *C. thermophilum* UBA had significant differences with an N-terminal additional extension and a longer more protruding loop connecting α2 and α3 of the UBA domain.

Mex67 is known to interact with many adaptor proteins and previous studies have shown that Mex67 UBA is not only the site of binding for the FG nucleoporins, but also the site of binding of Hpr1, a core THO/TREX complex protein. Curiously, both compete for the binding to the same hydrophobic pocket on Mex67 UBA with comparable affinities (Hobeika *et al.*, 2009). Furthermore, previous study that used the yeast two-hybrid technique to check interaction between Mex67 and Mip6 saw no binding when a construct containing the UBA domain of Mex67 (residues 549-599) was used, and thus concluded that the UBA domain of Mex67 was not the direct site of binding of Mip6 and instead an intact Mex67 carboxy-terminal domain which includes a part of the UBA domain is required for the two hybrid interaction with Mip6 (Segref *et al.*, 1997).

Therefore we used different techniques to check whether Mex67 also binds to Mip6 through its UBA domain. Initially, using the Bio-layer interferometry technique, where a tagged protein is experimentally required, indicated that Mip6 was binding on Mex67 UBA$_{(545\text{-}599)}$ while no binding was evident with only the N-terminal region (481-544). Subsequent pull down assays confirmed that the N-terminal region was not sufficient for obtaining binding on its own. However, surprisingly, cleaving the fused N-terminal Mex67 UBA tag prevented the two proteins holding stably together in gel filtration thus indicating a disruption in the binding. Affirmative results from the ITC experiments confirmed that indeed N-terminally free Mex67 UBA lost the ability to interact and bind Mip6, while an N-terminal GST tagged Mex67 UBA restored the binding with a binding affinity comparable to that with endogenous Mex67$_{(528\text{-}599)}$ comprising the UBA plus its N-terminal 18 residues (section 4.7.3, 4.7.4). Besides, the MASCOT results following the cross-linking experiment, were also excluding the N-terminal region of the UBA as part of the interaction. Instead, both MASCOT results and NMR spectrometry titration experiment of Mip6 RRM4 into Mex67 were pointing to the central helix 2 of the UBA, involved in the binding of both the FG repeats of the nucleoporins and Hpr1 protein, as the main helix involved in binding of Mip6 as visible in figure 5.8.



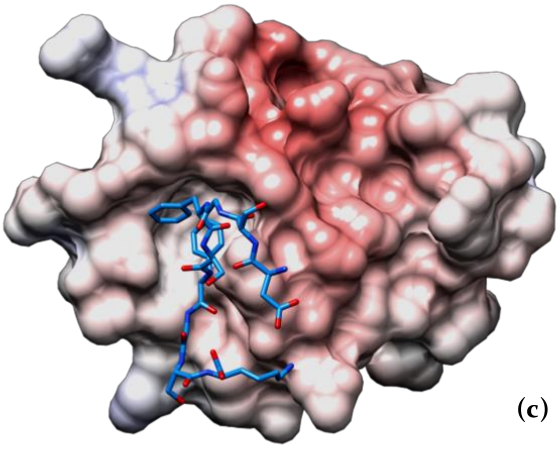


Figure 5.8. The site of binding of FG nucleoporins on Mex67 UBA domain. (**a**) An illustration showing the expected site of binding of FG nucleoporins on the surface of Mex67$_{(528-599)}$. In the illustration, a model composition where Mex67 is replacing the human homolog TAP or NXF1 UBA domain in the structure solved from it bound to FG nucleoporins (PDB: 1OAI). **(b)** The tertiary structure of Mex67$_{(528-599)}$ with the bound FG peptide. The parts highlighted in red are the ones identified by MASCOT as the parts involved in binding. The overlap of results from the cross-linking experiment and NMR titration suggests that helix 2 of the UBA involved in nucleoporins binding is majorly involved in Mip6 binding. (**c**) The surface electrostatic potential on Mex67$_{(528-599)}$ surface calculated by ABPS and visualized using *Chimera*.

Therefore, we here characterize Mip6 as a new Mex67 UBA binding protein in addition to the FG nucleoporins and the Hpr1 protein. However though, interestingly an N-terminal extension on the UBA is mandatory for the protein to be able to interact with Mip6. This mandatory N-terminal extension does not seem to be sequence specific as the replacement of the endogenous N-terminal region with the GST tag replenished the binding with a comparable affinity. Moreover, the risk of somehow the Mip6 protein binding non-specifically to the GST, or an effect of GST dimerization, is dismissed as both Mex67 UBA and Mex67 ΔUBA proteins were GST tagged.

So depending on the several lines of evidence, we suggest that additional residues on the N-terminus of Mex67 UBA are required for Mip6 binding, not necessarily sequence specific or because of a direct contribution in the binding, but more likely for certain conformational purposes putting the Mex67 UBA domain in a conformation required for Mip6 identification, or maybe for a role in stabilizing the binding once Mip6 is actually bound. Moreover, as visible in the Mex67$_{(528-599)}$

structure (figure 5.9), the N-terminal region was not floating in an open conformation; instead it was folding toward the first helix of the UBA domain, as also seen for the last helix 4 of the UBA. It is known, for example, that this last helix of the Mex67 UBA domain (H4) works as a switch required for stabilizing the binding of the FG repeats or Hpr1 to the UBA, as it packs against the bottom of their binding pocket and so seals it although it is not itself directly involved in the binding (Grant *et al.*, 2003; Hobeika *et al.*, 2009). Therefore, the deletion of H4 would greatly reduce the strength of its interaction to its Hpr1 or FG nucleoporins partners (Hobeika *et al.*, 2007). Additionally, although it is not directly involved in binding, its residues showed chemical shift changes during NMR spectrometry titration experiment with FG peptide consistent with an FG induced conformational changes in helix 4 (Grant *et al.*, 2003). This might also be the case of the N-terminal region of the UBA, but instead of being the C-term helix it would be the N-term region stabilizing and locking helix 2 in position required for binding, although it might not be itself majorly involved in binding (figure 5.9). This would also explain the NMR spectroscopy shift changes seen in the UBA N-terminal area when Mip6 RRM4 is titrated into Mex67$_{(528\text{-}599)}$ (figure 4.54), although our cross-linking/MASCOT and ITC results did not show any direct involvement of the N-terminal region. Furthermore, an interesting future experiment would be to check whether Mip6 binding interferes with the FG repeats or Hpr1 binding or vice versa, or whether Mex67 can carry both partners simultaneously.

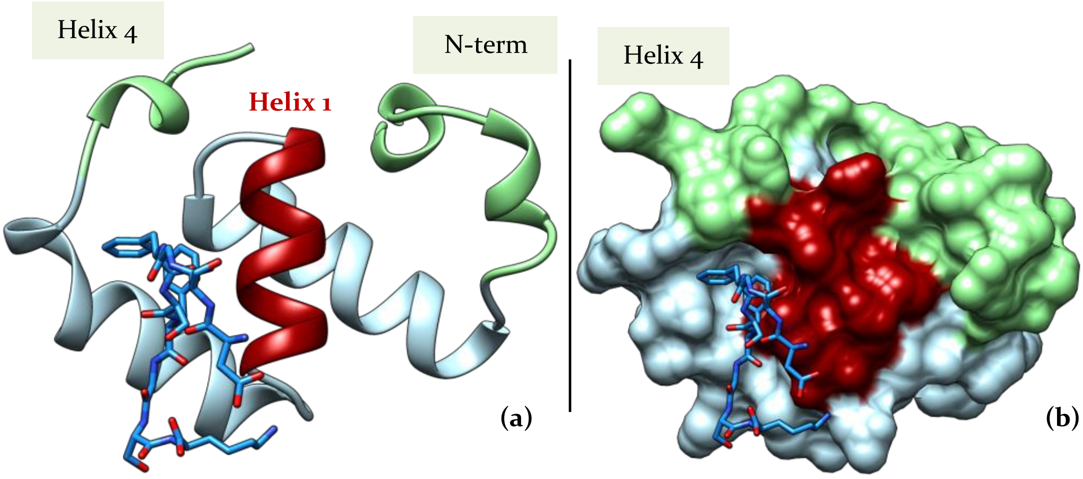


Figure 5.9. (a) In the illustration, the core helix 2 of the Mex67 UBA domain is highlighted in red while both the C-terminal helix 4 and the additional N-terminal region of the UBA are shown in green. As demonstrated, these two extremities form like a clamp folding toward the core helix 2 locking it in position. So while helix four is required for the FG nucleoporins binding (shown in blue) it is not itself directly involved in the binding. The N-terminal region might be behaving the same for any partner binding on the other sides of helix 2, thus locking it in position and stabilising its interaction. **(b)** The surface of the structure. The illustration was generated using *Chemira* where the human Mex67 UBA from the structure with PDB ID: 1OAI was replaced by our Mex67 yeast C-terminal structure in a model composition.

5.8. Mip6 function *in vivo* as a possible novel adaptor protein for Mex67

Until this day, only two published studies have looked into the *in vivo* function of Mip6 and only one of them looked at Pes4 function along with Mip6, so its exact role *in vivo* is not yet well understood awaiting future details. Carrying out *in vivo* assays in collaboration with Susana Rodriguez Navarro lab however was able to give us a better look on the possible Mip6 function. The results (unpublished) show that although a deletion of Mip6 (Mip6Δ) was not cell-lethal as Mex67Δ and did not show any evident cell growth phenotype, a double mutant of Mip6 with the thermo-sensitive Mex67-5 strain (a thermo-sensitive Mex67 mutant (Segref *et al.*, 1997)) had a negative sensitive synthetic growth phenotype where cells were unable to grow at 33°C. Interestingly, it was found that the Mip6 W442 mutant, not able to interact with Mex67 also exhibits a thermo-sensitive phenotype even at 35°C. These results in addition to the fact that Mip6 is an RNA binding protein capable of direct RNA binding *in vitro* as we demonstrated would lead us to suggest a mode of function for Mip6. Here we suggest a possible function for Mip6 as a novel adaptor protein for Mex67 in mRNA export, as it fits all the categories for being one.

As discussed in the introduction (section2), Mex67 as the major mRNA export receptor is known to interact with many RNA binding adaptor proteins, where some of these adaptor proteins are required for the recruitment of Mex67 to transcribing genes before export, while others further shuttle with Mex67 to the cytoplasm. It might be proposed then, that the RNA binding Mip6 is also playing that role, and while it binds Mex67 through its RRM4, the other RRMs are still binding RNA as we showed and thus would be holding RNA and further shuttling with Mex67 to the nucleus especially as Mip6 doesn't have the capacity to shuttle on its own. This hypothesis is also supported by the fact that *in vivo* Mip6 localizes homogeneously both in nucleus and in cytoplasm at normal growth conditions (Bolognesi *et al.*, 2016), however when Mex67 was inactivated *in vivo*, most Mip6-GFP was retained in the nucleus. Additionally, a non-published work from in Susana Rodriguez Navarro lab have also provided evidence for Mip6 co-purifying with proteins involved in mRNA export and known to interact with Mex67 as it exports the nucleus like Sac3 (a part of the TREX-2 complex on the nuclear face of the NPC), and Dbp5 (localizes at the cytoplasmic face of the NPC) using TAP purification. This proposition would fit and explain all what we learned about Mip6 till the moment from RNA binding, to its interaction with a major export receptor Mex67, and also its localization both in the nucleus and cytoplasm. Therefore, we characterize Mip6 as a novel adaptor protein for Mex67 exporter.

5.9. Mip6 additional role during cellular stress

On the other hand, previous published results also show that when stressing the cells to 39°C, Mip6 localization was altered and was seen evidently in cytoplasmic stress granules. One study that looked at the cellular toxicity caused by a concentration-dependent liquid phase separation upon increased protein expression, suggested that the strong over-expression of the dosage-sensitive Mip6 protein indeed changes its localization to cytoplasmic foci impairing cell growth, reducing the global translation rates, and causing toxicity (Bolognesi *et al.*, 2016). The study also found that these cytoplasmic foci contained RNA and demonstrated that at least two of the RRM domains of Mip6 are required for the cytoplasmic focus formation.

Another study that dealt with both Mip6 and Pes4 also suggested a role for Mip6 in regulating steady state expression, protection, and mRNA localization during stress conditions (Jin *et al.*, 2017). They also found that during stress induced sporulation, Mip6 localizes to cytoplasmic foci that contained mRNA transcripts which Mip6 protected from degradation. Additionally, the RRM domains of Mip6 were also proven required for the function.

Taking our *in vivo* results in collaboration with Susana Rodriguez Navarro laboratory and the literature studies into account, we then propose another function of Mip6 in stress conditions, so while Mip6 binds mRNA and may shuttle with Mex67 under normal conditions thus playing the role of an adaptor protein in mRNA export and gene expression regulation, when the cells suddenly get stressed, Mip6 carries mRNA transcripts through its RRM domains to cytoplasmic granules for protection until the growth conditions goes back to normal so the gene translation of non-stress specific genes then continues.

Recently, a very interesting study also suggested an additional role for Mex67 during stress conditions (Zander *et al.*, 2016). It demonstrated that unlike Mex67 role under normal growth conditions, upon stress Mex67 dissociates from its known adaptors in the nucleus such as Npl3 and Nab2 and therefore from regular mRNA in favour of binding and rapid export of heat shock, stress-specific mRNA, by passing mRNA quality control checkpoints. On the other hand the cytoplasmic mRNA pool accumulates with bound proteins in stress granules. However, it was not yet clear whether in this case, Mex67 directly binds mRNA on its own or to unknown stress-specific adaptor proteins. But primarily results from Susana Rodriguez Navarro laboratory suggest that Mip6 is probably also an adaptor protein for Mex67 in stress conditions. The results show that in stress conditions Mip6 preferentially binds transcripts connected to stress response. Additionally, the export of a certain stress induced transcript SSA4 that depends exclusively on Mex67 for transcripts during stress was seen to be less efficiently exported in the absence of Mip6 compared to the wild-type strain (unpublished work).

This is consistent with our hypothesis, as the protein might have different functions not only in normal conditions or stress conditions, but also in different pools of the protein with different function. We then suggest a model where the cytoplasmic pool of Mip6 re-localizes with its bound RNA transcripts to stress granules under stress, while the nuclear pool of Mip6 acts as an adaptor protein for the nuclear pool of Mex67 to rapidly export stress-specific mRNA transcripts. This is also consistent with the fact that other studies have also shown that a pool of Mex67 re-localizes to stress-granules or P-bodies under stress conditions (Mitchell, Jain, She, & Parker, 2013). Probably, the pool of Mex67 that re-localizes to the stress granules is also the cytoplasmic pool of the protein, since it is proven recently that the nuclear pool was the one exporting stress-specific mRNAs (Zander *et al.*, 2016). This stress granules localization of the cytoplasmic pool of the proteins, however, seems to be independent of the Mip6/Mex67 interaction, as the Mip6 W442A mutant was still localizing to cytoplasmic granules upon stress.

5.10. Pes4 evolutionary divergence of function

Our results also indicate that Mip6 and Pes4 may have evolved separate functions through time, which is usually the case of paralogs. Although both had a role during cellular stress induced sporulation

(Jin *et al.*, 2017), and both re-localize to cytoplasmic foci, however each protein displayed distinct cytoplasmic protein localization. Moreover, while the protection for few mRNA transcripts by either Mip6 or Pes4 during sporulation was redundant, many mRNA transcripts were uniquely protected by either Mip6 or Pes4. These information in addition to our results showing that Pes4 did not have the ability to interact with Mex67, therefore implicates that whatever function Mip6 was accomplishing in complex with Mex67, it is not relevant or shared with Pes4. Additionally, although both are RNA binding proteins having similar roles in stress conditions, but evolved to recognize different mRNA transcripts uniquely although some might be still recognized by both.

5.11. Concluding remarks

Our work that combines structural, functional and biophysical analysis hopefully adds to our knowledge and understanding of the dynamics of cellular functions especially mRNA export. However, scientific knowledge is the fruit of accumulated work, and our work here is just a part contribution for a better understanding and future possible implications involving the vital interactions and mechanisms in which a major exporter like Mex67 is involved in during different cellular conditions.

6. Conclusions

1. Mip6 is an RNA binding protein that binds RNA *in vitro* with high nano-molar affinity.
2. Pes4, the prologue of Mip6, is also an RNA binding protein capable of binding general RNA *in vitro* with high affinity.
3. Mip6 contains four RNA Recognition Motifs (RRMs), all having the capability of binding RNA with the fourth RRM not been previously characterized.
4. Mip6 RRM1/2 when mixed with Poly(A) RNA chaperones the growth of crystals formed by a parallel Poly(A) duplex structure.
5. Mip6 RRM3 is a canonical RRM with a flexible loop3 adopting different conformations in different structures. Mip6 RRM3 is missing the conserved exposed aromatic residues in the RNP2 sequence and binds with less affinity to RNA possibly due to a lower contribution in cooperative RNA binding.
6. Pes4 RRM3 and RRM4 structures adopt a canonical RRM fold, with Pes4 RRM4 showing some structural differences in terms of longer ß2 and ß3 with a much shortened and more stable loop3. Pes4 RRM3 and RRM4 are connected with a disordered 15 residues inter-domain linker. Pes4 RRM3 and RRM4 have no contacts between each other in their free form.
7. Mip6 RRM4 binds both RNA and protein, and it is the physical site of interaction with Mex67 protein, the major mRNA nuclear exporter conserved from yeast to humans.
8. Pes4 RRM4 does not bind Mex67 despite having more than 40% sequence conservation with Mip6 RRM4.
9. The binding of Mip6 to either Mex67 or RNA first conditions its ability to bind the other partner: a characteristic not seen before for an RRM domain.
10. A point mutation of Tryptophan W442A on Mip6 RRM4 loop3 disrupts its interaction with Mex67, but not its interaction with RNA. This Tryptophan is not conserved in Pes4.

11. Mex67 C-terminal crystal structure includes the UBA domain with additional N-terminal 18 residues packing toward the UBA domain core forming a clamp like shape.
12. The Mex67 UBA domain from *Saccharomyces cerevisiae* is more structurally conserved with the human Mex67 homolog UBA domain, than that from *Chaetomium thermophilum*.
13. Mex67 binds Mip6 RRM4 through its UBA domain. However, a non-specific N-terminal region of the UBA is needed not as a main contributor in the interaction but most probably to stabilize the binding.
14. *In vivo* experiments suggest a role for Mip6 protein as a novel Mex67 adaptor especially under stress growth conditions.

7. References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., … Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, *66*(2), 213–221.

Aguilera, A. (2005). Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. *Current Opinion in Cell Biology*, *17*(3), 242–250.

Aguilera, A., & Klein, H. L. (1988). Genetic control of intrachromosomal recombination in Saccharomyces cerevisiae. I. Isolation and genetic characterization of hyper-recombination mutations. *Genetics*, *119*(4), 779–790.

Aibara, S., Bai, X.-C., & Stewart, M. (2016). The Sac3 TPR-like region in the Saccharomyces cerevisiae TREX-2 complex is more extensive but independent of the CID region. *Journal of Structural Biology*, *195*(3), 316–324.

Aibara, S., Valkov, E., Lamers, M., & Stewart, M. (2015). Domain organization within the nuclear export factor Mex67:Mtr2 generates an extended mRNA binding surface. *Nucleic Acids Research*, *43*(3), 1927–1936.

Allain, F. H.-T., Bouvet, P., Dieckmann, T., & Feigon, J. (2000). Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *The EMBO Journal*, *19*, 6870–6881. Retrieved from

Allen, B. L., & Taatjes, D. J. (2015). The Mediator complex: a central integrator of transcription. *Nature Reviews Molecular Cell Biology*, *16*(3), 155–166.

Askjaer, P., Bachi, A., Wilm, M., Bischoff, F. R., Weeks, D. L., Ogniewski, V., … Fornerod, M. (1999). RanGTP-regulated interactions of CRM1 with nucleoporins and a shuttling DEAD-box helicase. *Molecular and Cellular Biology*, *19*(9), 6276–85.

Bachi, A., Braun, I. C., Rodrigues, J. P., Panté, N., Ribbeck, K., von Kobbe, C., … Izaurralde, E. (2000). The C-terminal domain of TAP interacts with the nuclear pore complex and promotes export of specific CTE-bearing RNA substrates. *Rna*, *6*(1), 136–158.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., & Leslie, A. G. W. (2011). iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography*, *67*(4), 271–281.

Bayliss, R., Littlewood, T., & Stewart, M. (2000). Structural Basis for the Interaction between FxFG Nucleoporin Repeats and Importin-ß in Nuclear Trafficking. *Cell*, *102*, 99–108.

Bolognesi, B., Gotor, N. L., Dhar, R., Cirillo, D., Baldrighi, M., Tartaglia, G. G., & Lehner, B. (2016). A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression. *Cell Reports*, *16*(1), 222–232.

Bradatsch, B., Katahira, J., Kowalinski, E., Bange, G., Yao, W., Sekimoto, T., … Hurt, E. (2007). Arx1 Functions as an Unorthodox Nuclear Export Receptor for the 60S Preribosomal Subunit. *Molecular Cell*, *27*(5), 767–779.

Byrne, K. P., & Wolfe, K. H. (2005). The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Research*, *15*(10), 1456–61.

Castanotto, D., Lingeman, R., Riggs, A. D., & Rossi, J. J. (2009). CRM1 mediates nuclear-cytoplasmic shuttling of mature microRNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(51), 21655–9.

Chlebowski, A., Lubas, M., Jensen, T. H., & Dziembowski, A. (2013). RNA decay machines: The exosome. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1829*(6), 552–560.

Chook, Y. M., & Süel, K. E. (2011). Nuclear import by karyopherin-βs: Recognition and inhibition. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1813*(9), 1593–1606.

Chung, J. K. H. & Y. D. (2016). A novel model of THO/TREX loading onto target RNAs in metazoan gene expression. *BMB Reports*, *49*(7), 355–356.

Clé Ry, A., Blatter, M., Dé Ric, F., & Allain, H.-T. (2008). RNA recognition motifs: boring? Not quite. *Current Opinion in Structural Biology*, *18*, 290–298.

Cléry, A., & Allain, F. H.-T. (2013). FROM STRUCTURE TO FUNCTION OF RNA BINDING DOMAINS. Retrieved from

Cowtan, K. (2006). The *Buccaneer* software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D Biological Crystallography*, *62*(9), 1002–1011.

Crowder, S. M., Kanaar, R., Rio, D. C., & Alber, T. (1999). Absence of interdomain contacts in the crystal structure of the RNA recognition motifs of Sex-lethal. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(9), 4892–7.

Cullen, B. R. (2003). Nuclear mRNA export: insights from virology. *Trends in Biochemical Sciences*, *28*(8), 419–424.

Dargemont, C., & Kühn, L. C. (1992). Export of mRNA from microinjected nuclei of Xenopus laevis oocytes. *The Journal of Cell Biology*, *118*(1).

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, *6*(3), 277–293.

Dimitrova, L., Valkov, E., Aibara, S., Flemming, D., McLaughlin, S. H., Hurt, E., & Stewart, M. (2015). Structural Characterization of the Chaetomium thermophilum TREX-2 Complex and its Interaction with the mRNA Nuclear Export Factor Mex67:Mtr2. *Structure (London, England : 1993)*, *23*(7), 1246–57.

Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., & Baker, N. A. (2007). PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research*, *35*(SUPPL.2), W522–W525.

Dong, S., Li, C., Zenklusen, D., Singer, R. H., Jacobson, A., & He, F. (2007). YRA1 autoregulation requires nuclear export and cytoplasmic Edc3p-mediated degradation of its pre-mRNA. *Molecular Cell*, *25*(4), 559–73.

Dong, X., Biswas, A., Süel, K. E., Jackson, L. K., Martinez, R., Gu, H., & Chook, Y. M. (2009). Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature*, *458*(7242), 1136–1141.

Dufu, K., Livingstone, M. J., Seebacher, J., Gygi, S. P., Wilson, S. A., & Reed, R. (2010). ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. *Genes & Development*, *24*(18), 2043–53.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, *66*(4), 486–501.

Eshleman, N., Liu, G., McGrath, K., Parker, R., & Buchan, J. R. (2016). Defects in THO/TREX-2 function cause accumulation of novel cytoplasmic mRNP granules that can be cleared by autophagy. *RNA (New York, N.Y.)*, *22*(8), 1200–14.

Fairman, M. E., Maroney, P. A., Wang, W., Bowers, H. A., Gollnick, P., Nilsen, T. W., & Jankowsky, E. (2004). Protein Displacement by DExH/D &quot;RNA Helicases&quot; Without Duplex Unwinding. *Science*, *304*(5671).

Fasken, M. B., Stewart, M., & Corbett, A. H. (2008). Functional significance of the interaction between the mRNA-binding protein, Nab2, and the nuclear pore-associated protein, Mlp1, in mRNA export. *The Journal of Biological Chemistry*, *283*(40), 27130–43.

Faza, M. B., Chang, Y., Occhipinti, L., Kemmler, S., Panse, V. G., Warner, J., … Léger-Silvestre, I. (2012). Role of Mex67-Mtr2 in the Nuclear Export of 40S Pre-Ribosomes. *PLoS Genetics*, *8*(8), e1002915.

Faza, M. B., Kemmler, S., Jimeno, S., González-Aguilera, C., Aguilera, A., Hurt, E., & Panse, V. G. (2009). Sem1 is a functional component of the nuclear pore complex–associated messenger RNA export machinery. *The Journal of Cell Biology*, *184*(6), 833–846.

Fischer, T., Strässer, K., Rácz, A., Rodriguez-Navarro, S., Oppizzi, M., Ihrig, P., … Reed, R. (2002). The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. *The EMBO Journal*, *21*(21), 5843–52.

Fornerod, M., Ohno, M., Yoshida, M., & Mattaj, I. W. (1997). CRM1 Is an Export Receptor for Leucine-Rich Nuclear Export Signals. *Cell*, *90*(6), 1051–1060.

Fribourg, S., Braun, I. C., Izaurralde, E., & Conti, E. (2001). Structural basis for the recognition of a nucleoporin FG repeat by the NTF2-like domain of the TAP/p15 mRNA nuclear export factor. *Molecular Cell*, *8*(3), 645–656.

Fujiwara, N., Shiki, T., & Masu, S. (2012). mRNA Biogenesis in the Nucleus and Its Export to the Cytoplasm. In *Current Frontiers and Perspectives in Cell Biology*. InTech.

Gallouzi, I.-E., & Steitz, J. A. (2001). Delineation of mRNA Export Pathways by the Use of Cell-Permeable Peptides. *Science*, *294*(5548).

Grant, R. P., Neuhaus, D., & Stewart, M. (2003). Structural basis for the interaction between the Tap/NXF1 UBA domain and FG nucleoporins at 1 Å resolution. *Journal of Molecular Biology*, *326*(3), 849–858.

Green, D. M., Johnson, C. P., Hagan, H., & Corbett, A. H. (2003). The C-terminal domain of myosin-like protein 1 (Mlp1p) is a docking site for heterogeneous nuclear ribonucleoproteins that are required for mRNA export. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(3), 1010–5.

Güttler, T., Madl, T., Neumann, P., Deichsel, D., Corsini, L., Monecke, T., … Görlich, D. (2010). NES consensus redefined by structures of

PKI-type and Rev-type nuclear export signals bound to CRM1. *Nature Structural & Molecular Biology*, *17*(11), 1367–1376.

Gwizdek, C., Iglesias, N., Rodriguez, M. S., Ossareh-Nazari, B., Hobeika, M., Divita, G., … Dargemont, C. (2006). Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(44), 16376–81.

Hackmann, A., Gross, T., Baierlein, C., & Krebber, H. (2011). The mRNA export factor Npl3 mediates the nuclear export of large ribosomal subunits. *EMBO Reports*, *12*(10), 1024–31.

Hautbergue, G. M., Hung, M.-L., Golovanov, A. P., Lian, L.-Y., & Wilson, S. A. (2008). Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(13), 5154–9.

Hector, R. E., Nykamp, K. R., Dheur, S., Anderson, J. T., Non, P. J., Urbinati, C. R., … Swanson, M. S. (2002). Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. *The EMBO Journal*, *21*(7), 1800–10.

Herold, A., Suyama, M., Rodrigues, J. P., Braun, I. C., Kutay, U., Carmo-Fonseca, M., … Izaurralde, E. (2000). TAP (NXF1) belongs to a multigene family of putative RNA export factors with a conserved modular architecture. *Molecular and Cellular Biology*, *20*(23), 8996–9008.

Hobeika, M., Brockmann, C., Gruessing, F., Neuhaus, D., Divita, G., Stewart, M., & Dargemont, C. (2009). Structural requirements for the ubiquitin-associated domain of the mRNA export factor Mex67 to bind its specific targets, the transcription elongation THO complex component Hpr1 and nucleoporin FXFG repeats. *The Journal of Biological Chemistry*, *284*(26), 17575–83.

Hobeika, M., Brockmann, C., Iglesias, N., Gwizdek, C., Neuhaus, D., Stutz, F., … Dargemont, C. (2007). Coordination of Hpr1 and ubiquitin binding by the UBA domain of the mRNA export factor

Mex67. *Molecular Biology of the Cell*, *18*(7), 2561–8.

Hocine, S., Singer, R. H., & Grünwald, D. (2010). RNA processing and export. *Cold Spring Harbor Perspectives in Biology*, *2*(12), a000752.

Hoelz, A., Glavy, J. S., & Beck, M. (2016). Toward the atomic structure of the nuclear pore complex: when top down meets bottom up. *Nature Structural & Molecular Biology*, *23*(7), 624–630.

Huang, Y., Gattoni, R., Stévenin, J., & Steitz, J. A. (2003). SR Splicing Factors Serve as Adapter Proteins for TAP-Dependent mRNA Export. *Molecular Cell*, *11*(3), 837–843.

Huertas, P., & Aguilera, A. (2003). Cotranscriptionally Formed DNA:RNA Hybrids Mediate Transcription Elongation Impairment and Transcription-Associated Recombination. *Molecular Cell*, *12*(3), 711–721.

Iglesias, N., Tutucci, E., Gwizdek, C., Vinciguerra, P., Von Dach, E., Corbett, A. H., … Stutz, F. (2010). Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. *Genes & Development*, *24*(17), 1927–38.

Jankowsky, E., & Bowers, H. (2006). Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. *Nucleic Acids Research*, *34*(15), 4181–8.

Jimenez, R. M., Polanco, J. A., Lupt??k, A., & Lupták, A. (2015). Chemistry and Biology of Self-Cleaving Ribozymes. *Trends in Biochemical Sciences*, *40*(11), 648–61.

Jin, L., Zhang, K., Sternglanz, R., & Neiman, A. M. (2017). Predicted RNA Binding Proteins Pes4 and Mip6 Regulate mRNA Levels, Translation, and Localization during Sporulation in Budding Yeast. *Molecular and Cellular Biology*, *37*(9), e00408-16.

Johnson, S. A., Cubberley, G., & Bentley, D. L. (2009). Cotranscriptional Recruitment of the mRNA Export Factor Yra1 by Direct Interaction with the 3′ End Processing Factor Pcf11. *Molecular Cell*, *33*(2), 215–226.

Kabachinski, G., & Schwartz, T. U. (2015). The nuclear pore complex –

structure and function at a glance. *Journal of Cell Science*, *128*(3), 423–429.

Kabsch, W. (2010). XDS. *Acta Crystallographica Section D: Biological Crystallography*, *66*(2), 125–132.

Katahira, J. (2012). mRNA export and the TREX complex. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1819*(6), 507–513.

Katahira, J., Strässer, K., Podtelejnikov, A., Mann, M., Jung, J. U., Hurt, E., … Jung, J. U. (1999). The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human. *The EMBO Journal*, *18*(9), 2593–609.

Kielkopf, C. L., Lücke, S., & Green, M. R. (2004). U2AF homology motifs: protein recognition in the RRM world. *Genes & Development*, *18*(13), 1513–1526.

Kilchert, C., & Vasiljeva, L. (2013). mRNA quality control goes transcriptional. *Biochemical Society Transactions*, *41*(6).

Knockenhauer, K. E., & Schwartz, T. U. (2016). The Nuclear Pore Complex as a Flexible and Dynamic Gate. *Cell*, *164*(6), 1162–1171.

Köhler, A., & Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. *Nature Reviews Molecular Cell Biology*, *8*(10), 761–773.

Kuhlmann, S. I., Valkov, E., & Stewart, M. (2014). Structural basis for the molecular recognition of polyadenosine RNA by Nab2 Zn fingers. *Nucleic Acids Research*, *42*(1), 672–80.

Kurshakova, M. M., Georgieva, S. G., & Kopytova, D. V. (2016). Protein complexes coordinating mRNA export from the nucleus into the cytoplasm. *Molecular Biology*, *50*(5), 639–644.

Le Hir, H., Gatfield, D., Izaurralde, E., & Moore, M. J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *The EMBO Journal*, *20*(17), 4987–97.

Le Hir, H., Izaurralde, E., Maquat, L. E., Moore, M. J., Alexiadis, V., Waldmann, T., … Reed, R. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *The EMBO Journal*, *19*(24), 6860–9.

Lei, E. P., Krebber, H., & Silver, P. A. (2001). Messenger RNAs are recruited for nuclear export during transcription. *Genes & Development*, *15*(14), 1771–82.

Li, H., Shi, H., Wang, H., Zhu, Z., Li, X., Gao, Y., … Teng, M. (2010). Crystal structure of the two N-terminal RRM domains of Pub1 and the poly(U)-binding properties of Pub1. *Journal of Structural Biology*, *171*(3), 291–297.

Libri, D., Graziani, N., Saguez, C., & Boulay, J. (2001). Multiple roles for the yeast SUB2/yUAP56 gene in splicing. *Genes & Development*, *15*(1), 36–41.

Lin, D. H., Stuwe, T., Schilbach, S., Rundlet, E. J., Perriches, T., Mobbs, G., … Hoelz, A. (2016). Architecture of the symmetric core of the nuclear pore. *Science*, *352*(6283).

Long, F., Vagin, A. A., Young, P., & Murshudov, G. N. (2007). BALBES: A molecular-replacement pipeline. In *Acta Crystallographica Section D: Biological Crystallography* (Vol. 64, pp. 125–132). International Union of Crystallography.

Lund, M. K., & Guthrie, C. (2005). *The DEAD-Box Protein Dbp5p Is Required to Dissociate Mex67p from Exported mRNPs at the Nuclear Rim*. *Molecular Cell* (Vol. 20).

Lunde, B. M., Moore, C., & Varani, G. (2007). {RNA}-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*, *8*(6), 479–490.

Maris, C., Dominguez, C., & Allain, F. H. T. (2005, April 21). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS Journal*.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., & Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. *Genes & Development*, *19*(13), 1512–7.

Mazza, C., Segref, A., Mattaj, I. W., & Cusack, S. (2002). Large-scale induced fit recognition of an m7GpppG cap analogue by the human nuclear cap-binding complex. *EMBO Journal*, *21*(20), 5548–5557.

Mitchell, S. F., Jain, S., She, M., & Parker, R. (2013). Global analysis of yeast mRNPs. *Nature Structural & Molecular Biology*, *20*(1), 127–33.

Monecke, T., Güttler, T., Neumann, P., Dickmanns, A., Görlich, D., & Ficner, R. (2009). Crystal Structure of the Nuclear Export Receptor CRM1 in Complex with Snurportin1 and RanGTP. *Science*, *324*(5930).

Montpetit, B., Thomsen, N. D., Helmke, K. J., Seeliger, M. A., Berger, J. M., & Weis, K. (2011). A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. *Nature*, *472*(7342), 238–242.

Müller-McNicoll, M., & Neugebauer, K. M. (2013). How cells get the message: dynamic assembly and function of mRNA–protein complexes. *Nature Reviews Genetics*, *14*(4), 275–287.

Nagai, K., Price, S. R., & Evans, P. R. (1998). Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature*, *394*(6694), 645–650.

Neumann, B., Wu, H., Hackmann, A., & Krebber, H. (2016). Nuclear export of pre-ribosomal subunits requires Dbp5, but not as an RNA-helicase as for mRNA export. *PLoS ONE*, *11*(2), e0149571.

Neville, M., & Rosbash, M. (1999). The NES-Crm1p export pathway is not a major mRNA export route in Saccharomyces cerevisiae. *The EMBO Journal*, *18*(13), 3746–56.

Niño, C. A., Hérissant, L., Babour, A., & Dargemont, C. (2013). mRNA Nuclear Export in Yeast. *Chemical Reviews*, *113*(11), 8523–8545.

Noble, K. N., Tran, E. J., Alcázar-Román, A. R., Hodge, C. A., Cole, C. N., & Wente, S. R. (2011). The Dbp5 cycle at the nuclear pore complex during mRNA export II: Nucleotide cycling and mRNP remodeling by Dbp5 are controlled by Nup159 and Gle1. *Genes* *and Development*, *25*(10), 1065–1077.

Oeffinger, M., Wei, K. E., Rogers, R., DeGrasse, J. A., Chait, B. T., Aitchison, J. D., & Rout, M. P. (2007). Comprehensive analysis of diverse ribonucleoprotein complexes. *Nature Methods*, *4*(11), 951–956.

Okamura, M., Inose, H., & Masuda, S. (2015). RNA Export through the NPC in Eukaryotes. *Genes*, *6*(1), 124–149.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612.

Ramanathan, A., Robb, G. B., & Chan, S.-H. (2016). SURVEY AND SUMMARY mRNA capping: biological functions and applications. *Nucleic Acids Research*, *44*(16), 7511–7526.

Reed, R., Clouse, K. N., Luo, M., & Zhou, Z. (2001). A Ran-independent pathway for export of spliced mRNA. *Nature Cell Biology*, *3*(1), 97–99.

Rehwinkel, J., Herold, A., Gari, K., Köcher, T., Rode, M., Ciccarelli, F. L., … Izaurralde, E. (2004). Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster. *Nature Structural & Molecular Biology*, *11*(6), 558–566.

Ren, Y., Schmiege, P., Blobel, G., Abruzzi, K., Lacadie, S., Rosbash, M., … Stutz, F. (2017). Structural and biochemical analyses of the DEAD-box ATPase Sub2 in association with THO or Yra1. *eLife*, *6*, 2620–2631.

Rideau, A. P., Gooding, C., Simpson, P. J., Monie, T. P., Lorenz, M., Hüttelmaier, S., … Smith, C. W. J. (2006). A peptide motif in Raver1 mediates splicing repression by interaction with the PTB RRM2 domain. *Nature Structural & Molecular Biology*.

Robertson, M. P., & Joyce, G. F. (2014). Highly Efficient Self-Replicating RNA Enzymes. *Chemistry & Biology*, *21*(2), 238–245.

Rodríguez, D. D., Grosse, C., Himmel, S., González, C., de Ilarduya, I.

M., Becker, S., … Usón, I. (2009). Crystallographic ab initio protein structure solution below atomic resolution. *Nature Methods*, *6*(9), 651–653.

Rodríguez, D., Sammito, M., Meindl, K., de Ilarduya, I. M., Potratz, M., Sheldrick, G. M., & Usón, I. (2012). Practical structure solution with ARCIMBOLDO. *Acta Crystallographica. Section D, Biological Crystallography*, *68*(Pt 4), 336–43.

Rodríguez-Navarro, S., Fischer, T., Luo, M.-J., Antúnez, O., Brettschneider, S., Lechner, J., … Hurt, E. (2004). Sus1, a Functional Component of the SAGA Histone Acetylase Complex and the Nuclear Pore-Associated mRNA Export Machinery. *Cell*, *116*(1), 75–86.

Safaee, N., Noronha, A. M., Rodionov, D., Kozlov, G., Wilds, C. J., Sheldrick, G. M., & Gehring, K. (2013). Structure of the Parallel Duplex of Poly(A) RNA: Evaluation of a 50 Year-Old Prediction. *Angewandte Chemie International Edition*, *52*(39), 10370–10373.

Saguez, C., & Jensen, T. H. (2009). Assembly of Export-Competent mRNP: It's All about Being Connected. *Molecular Cell*, *33*(2), 139–140.

Sainsbury, S., Bernecky, C., & Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, *16*(3), 129–143.

Santos-Pereira, J. M., Herrero, A. B., Moreno, S., & Aguilera, A. (2014). Npl3, a new link between RNA-binding proteins and the maintenance of genome integrity. *Cell Cycle (Georgetown, Tex.)*, *13*(10), 1524–9.

Santos-Rosa, H., Moreno, H., Simos, G., Segref, A., Fahrenkrog, B., Panté, N., & Hurt, E. (1998). Nuclear mRNA export requires complex formation between Mex67p and Mtr2p at the nuclear pores. *Molecular and Cellular Biology*, *18*(11), 6826–38.

Schneider, M., Hellerschmied, D., Schubert, T., Amlacher, S., Vinayachandran, V., Reja, R., … Köhler, A. (2015). The Nuclear Pore-Associated TREX-2 Complex Employs Mediator to Regulate

Gene Expression. *Cell*, *162*(5), 1016–1028.

Segref, A., Sharma, K., Doye, V., Hellwig, A., Huber, J., Lührmann, R., … Hurt, E. (1997). Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores. *The EMBO Journal*, *16*(11), 3256–71.

Sengupta, J., Bussiere, C., Pallesen, J., West, M., Johnson, A. W., & Frank, J. (2010). Characterization of the nuclear export adaptor protein Nmd3 in association with the 60S ribosomal subunit. *The Journal of Cell Biology*, *189*(7), 1079–86.

Smith, C., Lari, A., Derrer, C. P., Ouwehand, A., Rossouw, A., Huisman, M., … Montpetit, B. (2015). In vivo single-particle imaging of nuclear mRNA export in budding yeast demonstrates an essential role for Mex67p. *Journal of Cell Biology*, *211*(6), 1121–1130.

Snay-Hodge, C. A., Colot, H. V, Goldstein, A. L., Cole, C. N., Aitchison, J., Blobel, G., … Ricupero-Hovasse, S. (1998). Dbp5p/Rat8p is a yeast nuclear pore-associated DEAD-box protein essential for RNA export. *The EMBO Journal*, *17*(9), 2663–76.

Sträßer, K., Baßler, J., & Hurt, E. (2000). Binding of the Mex67p/Mtr2p Heterodimer to Fxfg, Glfg, and Fg Repeat Nucleoporins Is Essential for Nuclear mRNA Export. *The Journal of Cell Biology*, *150*(4).

Sträßer, K., & Hurt, E. (2001). Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature*, *413*(6856), 648–652.

Strässer, K., Hurt, E., Bailer, S., Siniossoglou, S., Podtelejnikov, A., Hellwig, A., … Hurt, E. (2000). Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *The EMBO Journal*, *19*(3), 410–20.

Sträßer, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., … Hurt, E. (2002). TREX is a conserved complex coupling transcription with messenger RNA export. *Nature*, *417*(6886), 304–308. https://doi.org/10.1038/nature746

Teplova, M., Wohlbold, L., Khin, N. W., Izaurralde, E., & Patel, D. J. (2011). Structure-function studies of nucleocytoplasmic transport of retroviral genomic RNA by mRNA export factor TAP. *Nature Structural & Molecular Biology*, *18*(9), 990–8.

Tran, E. J., Zhou, Y., Corbett, A. H., & Wente, S. R. (2007). The DEAD-Box Protein Dbp5 Controls mRNA Export by Triggering Specific RNA:Protein Remodeling Events. *Molecular Cell*, *28*(5), 850–859.

Tutucci, E., & Stutz, F. (2011). Keeping mRNPs in check during assembly and nuclear export. *Nature Reviews Molecular Cell Biology*, *12*(6), 377–384.

Umlauf, D., Bonnet, J., Waharte, F., Fournier, M., Stierle, M., Fischer, B., … Tora, L. (2013). The human TREX-2 complex is stably associated with the nuclear pore basket. *Journal of Cell Science*, *126*(12), 2656–2667.

Viphakone, N., Hautbergue, G. M., Walsh, M., Chang, C.-T., Holland, A., Folco, E. G., … Wilson, S. A. (2012). TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. *Nature Communications*, *3*, 1006.

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., … Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function and Genetics*, *59*(4), 687–696.

Wang, H., Zeng, F., Liu, Q., Liu, H., Liu, Z., Niu, L., … Li, X. (2013). The structure of the ARE-binding domains of Hu antigen R (HuR) undergoes conformational changes during RNA binding. *Acta Crystallographica Section D Biological Crystallography*, *69*(3), 373–380.

Weirich, C. S., Erzberger, J. P., Flick, J. S., Berger, J. M., Thorner, J., & Weis, K. (2006). Activation of the DExD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP 6 is required for mRNA export. *NATURE CELL BIOLOGY*, *8*(7).

Wickramasinghe, V. O., & Laskey, R. A. (2015). Control of mammalian gene expression by selective mRNA export. *Nature Reviews*

Molecular Cell Biology, *16*(7), 431–442.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., … Wilson, K. S. (2011, April 1). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography.

Wolfe, K. H., & Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, *387*(6634), 708–713.

Woychik, N. A., & Hampsey, M. (2002). The RNA Polymerase II Machinery: Structure Illuminates Function. *Cell*, *108*(4), 453–463.

Yang, Z. R., Thomson, R., McNeil, P., & Esnouf, R. M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, *21*(16), 3369–3376.

Yao, W., Roser, D., Köhler, A., Bradatsch, B., Baßler, J., & Hurt, E. (2007). Nuclear Export of Ribosomal 60S Subunits by the General mRNA Export Receptor Mex67-Mtr2. *Molecular Cell*, *26*(1), 51–62.

Ye, K., Malinina, L., & Patel, D. J. (2003). Recognition of small interfering RNA by a viral suppressor of RNA silencing. *Nature*, *426*(6968), 874–878.

Yokoyama, S., Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., … Muto, Y. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature*, *398*(6728), 579–585.

Zander, G., Hackmann, A., Bender, L., Becker, D., Lingner, T., Salinas, G., & Krebber, H. (2016). mRNA quality control is bypassed for immediate export of stress-responsive transcripts. *Nature*, *540*(7634), 593–596.

Zenklusen, D., Vinciguerra, P., Wyss, J.-C., & Stutz, F. (2002). Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. *Molecular and Cellular Biology*, *22*(23), 8241–53.

Zhang, Y., French, S. L., Beyer, A. L., & Schneider, D. A. (2016). The Transcription Factor THO Promotes Transcription Initiation and Elongation by RNA Polymerase I. *The Journal of Biological Chemistry*, *291*(6), 3010–8.

Zhang, Y., & Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, *33*(7), 2302–9.

Zhao, Y., Chapman, D. A. G., & Jones, I. M. (2003). Improving baculovirus recombination. *Nucleic Acids Research*, *31*(2), e6–e6.