



**HAL**  
open science

# Insights into the control of mRNA decay by YTH proteins during the transition from meiosis to mitosis in yeasts.

Ditipriya Hazra

► **To cite this version:**

Ditipriya Hazra. Insights into the control of mRNA decay by YTH proteins during the transition from meiosis to mitosis in yeasts.. Biochemistry, Molecular Biology. Université Paris Saclay (COMUE), 2019. English. NNT : 2019SACLX041 . tel-03769919

**HAL Id: tel-03769919**

**<https://theses.hal.science/tel-03769919v1>**

Submitted on 6 Sep 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Insights into the control of mRNA decay by YTH proteins during the transition from mitosis to meiosis in yeasts

Thèse de doctorat de l'Université Paris-Saclay  
préparée à l'École Polytechnique

École doctorale n°573 : interfaces : approches interdisciplinaires,  
fondements, applications et innovation (Interfaces)  
Spécialité de doctorat: BIOLOGIE

Thèse présentée et soutenue à Palaiseau, le 5 Septembre 2019, par

**Ditipriya HAZRA**

## Composition du Jury :

<b>M. Damien HERMAND</b> Directeur de recherche, University of Namur (Department of Biology)	President
<b>M. Cyril DOMINGUEZ</b> Associate Professor, University of Leicester (Department of Molecular and Cell Biology)	Rapporteur
<b>M. Dominique FOURMY</b> Directeur de recherche, I2BC –CNRS (UMR9198)	Examineur
<b>Mme. Claudine MAYER</b> Professeure des Universités, Université Paris Diderot (UMR3528)	Examineur
<b>Mme. Cécile BOUSQUET-ANTONELLI</b> Directrice de recherche, CNRS (UMR5096)	Examineur
<b>M. Marc GRAILLE</b> Directeur de recherche, École Polytechnique, CNRS (UMR7654)	Directeur de thèse



# Acknowledgement

I would like to express deepest gratitude for my PhD supervisor Dr Marc Graille for accepting me as a PhD student in his team and for his continuous support and guidance throughout. I learned a lot of things from you which immensely helped me to develop an insight for research. Working under your supervision was truly a great learning opportunity for me.

I'd like to thank my jury members, Cyril Dominguez, Damien Hermand, Claudine Mayer, Cécile Bousquet-Antonelli and Dominique Fourmy for accepting to evaluate my thesis. Their insightful comments and detailed discussions helped me to improve my thesis.

I would like to thank all the past and present team members, it had been a pleasure knowing you all and working with you. Clement (Charenton), it was inspiring to work with you and thanks for all the scientific tips. Nhan, you were a good friend, always accommodative and nice. Nathalie, I cannot thank you enough for organizing everything in the lab and for always being there when I needed help. Clement (Chapat), I learned new information about cell and molecular biology from you. Clara, you have always been nice and helpful, it was a pleasure knowing you. Irene and Can, both of you just started your PhD and I know at this moment you are facing hurdles but I can assure you that you'll be able to overcome these obstacles one day. I wish you both good luck with your research. Niki, I wish you good luck for future endeavours.

I'm grateful to all my colleagues of BIOC, for their generous support.

I worked in close collaboration with Dr Mathieu Rougemaille's lab from I2BC for the *S. pombe* project. I express my sincere gratitude to Mathieu and Vedrana for their effort and for all the scientific discussions. For *S. cerevisiae* project I worked in collaboration with Dr Bertrand Seraphin's lab from IGBMC. I'd like to thank Jeremy and Melody for their hard work. Your results were very helpful.

There are some people who actively shared my workload and without them everything would have been very complicated. I'd like to thank Pascal and Pascale for washing tons of glassware that I had used and Guillaume for preparing litres of media. I'm grateful to Melanie, Caroline, Lydia and Catherine, for their help with all the administrative stuff.

I'd like to thank Professor Amit K Das (IIT Kharagpur) for giving me a platform to learn protein crystallography in details during my Masters and Dr Madhurima Roy for patiently teaching me the techniques with me when I was a novice.

My friends and family have always supported me. I was lucky to have some amazing friends who were with me through every thick and thin. Angelina, I am glad that I worked in this department because I met you! I miss our long conversations. Radhika and Silvia, it has been amazing to know you guys and when I leave France I'm going to miss you a lot.

Last but most important is my family without their love and encouragement I couldn't have achieved anything. I thank my parents for always standing by my side. Your sincerity, passion for work and integrity are a constant inspiration for me. I am blessed to have you as my parents. My friend, colleague and life-partner, Amlan, I cannot thank you enough for the way you support me, always putting me before yourself. You have an infectious passion for science which always motivates me. Thank you for everything!

# Table of Contents

## **Chapter 1: Introduction**

1.1 Cell Cycle	3
1.1.1 Stages of cell cycle	3
1.1.2 Meiosis	6
1.2 Meiosis in yeast	8
1.2.1 Meiosis in fission yeast	10
1.2.1.1 Entry to meiosis	10
1.2.1.2 Transcriptional regulation of meiosis	12
1.3 Mei2- the master regulator of meiosis	12
1.3.1 Structure and RNA binding mode of an RRM domain	13
1.3.2 Localisation of Mei2	13
1.3.3 Substrates of Mei2	15
1.3.4 Function of Mei2	15
1.3.5 Regulation of Mei2	15
1.4 Selective elimination of meiosis specific mRNAs	16
1.5 DSR mediated RNA degradation by Mmi1	18
1.5.1 The MTREC complex	18
1.5.2 Ccr4-Not complex	20
1.6 Erh1	21
1.6.1 Role of EMC in Heterochromatin island formation	21
1.7 Regulation of Mmi1 in meiotic cells	25
1.8 Mmi1-the exceptional YTH domain	25
1.9 Comparative study of human YTH domains	28
1.9.1 Discovery of m <sup>6</sup> A modification and YTH domain	28
1.9.2 The YTH Domain, an m <sup>6</sup> A RNA Grip	31
1.9.3 YTH, a Building Block Governing the Fates of m <sup>6</sup> A Containing mRNAs	34
1.9.3.1 YTHDC1	35
1.9.3.2 YTHDC2	37
1.9.3.3 YTHDF Family	38
1.10 YTH domain of Plants	42
1.11 Role of YTH domain protein in RNA degradation in <i>S. cerevisiae</i>	43
1.11.1 Structural comparison of Pho92 to human YTH domains	43
1.11.2 Function of Pho92 in phosphate metabolism	43
1.11.3 Pho4 degradation by Pho92	45

1.11.4 Role of <i>m<sup>6</sup>A</i> in meiosis of budding yeast	48
--	----

## **Chapter 2: Objectives**

2.1 Characterization of <i>Mmi1</i> regulation	51
2.2 Characterization of <i>Pho92</i> function	52

## **Chapter 3: Methods**

3.1 Cloning	57
3.1.1 Site directed mutagenesis	58
3.2 Solubility profiling	59
3.3 Protein over-expression	60
3.3.1 Selenomethionine labelling	60
3.4 Purification	60
3.4.1 Affinity chromatography	62
3.4.1.1 GST-purification	62
3.4.1.2 Ni-NTA purification	63
3.4.2 Ion exchange chromatography	63
3.4.3 Size exclusion chromatography	64
3.5 Fluorescence anisotropy	64
3.6 ITC	66
3.7 Limited proteolysis	67
3.8 Pull down	68
3.9 Thermal shift assay	68
3.10 SEC-MALLS	69
3.11 Crystallization, and diffraction data collection	71
3.11.1 Crystallization	71
3.11.2 Diffraction data collection	73
3.12 Structure solution	75
3.12.1 Data processing	75
3.12.2 Space group determination	77
3.12.3 Phasing	77
3.12.3.1 Data processing and Phase determination of <i>Mei2-RRM3</i>	78
3.12.3.2 Data processing and phasing of <i>Erh1</i>	80
3.13 Model building and Structure refinement	81
3.14 Materials	82

## **Results**

### **Chapter 4: Structural and functional characterization of *Mmi1*, *Mei2* and *Erh1***

4.1 <i>S. pombe</i> <i>Mei2</i> , the Master regulator of meiosis	93
---	----

4.1.1 Cloning	93
4.1.2 Expression assays	93
4.2 Study of <i>Mei2</i> -RRM1-2	95
4.2.1 Protein expression and purification	95
4.2.2 Crystallization of <i>Mei2</i> -RRM1-2	97
4.3 Study of the RRM3 domain from <i>Mei2</i>	97
4.3.1 Protein expression and purification	97
4.3.2 Crystallization of <i>Mei2</i> -RRM3	99
4.3.3 Diffraction data collection and structure solution	101
4.3.3.1 Estimating the number of molecules in asymmetric unit	101
4.3.3.2 Structure solution	103
4.3.4 Structure analysis of <i>Mei2</i> -RRM3	106
4.3.5 Determination of optimal RNA sequence for <i>Mei2</i> -RRM3 binding by isothermal Titration Calorimetry	108
4.4 Study of RRM3-RNA	110
4.4.1 Co-crystallization trial of RRM3-RNA	110
4.4.2 Diffraction data collection and structure solution	110
4.4.2.1 Space group determination and calculation of the number of molecules in asymmetric unit	112
4.4.2.2 Structure solution	112
4.4.3 Structure analysis of <i>Mei2</i> -RRM3-RNA	112
4.4.3.1 Description of the structure	112
4.4.3.2 Protein-RNA interaction	115
4.4.3.3 Recognition of bases	116
4.4.3.4 Identification of RNA binding residues	116
4.5 <i>Mmi1</i> - a putative partner of <i>Mei2</i>	118
4.5.1 Cloning	118
4.5.2 Solubility profiling	118
4.5.3 Protein production and purification of <i>Mmi1</i> -YTH	121
4.5.4 Determination of interaction between <i>Mmi1</i> -YTH and <i>Mei2</i> -RRM3	121
4.5.4.1 Co-purification	121
4.5.4.2 ITC	122
4.5.4.3 Pull-down	122
4.5.5 Crystallization of <i>Mmi1</i> - <i>Mei2</i> -RNA	122
4.5.6 <i>Mmi1</i> structure	124
4.6 <i>Erh1</i>	124
4.6.1 Cloning	124
4.6.2 Interaction between <i>Mmi1</i> & <i>Erh1</i>	124
4.6.2.1 Expression assays	124

4.6.2.2 Protein expression and purification of Mmi1-Erh1 complex	127
4.7 Mmi1-Erh1-Mei2	127
4.8 Crystallization of Erh1	129
4.9 Draft of Manuscript: Formation of <i>S. pombe</i> Erh1 homodimer mediates gametogenic gene silencing and meiosis progression	131

## **Chapter 5: Structural and functional characterization of Pho92 & Not1**

5.1 Cloning	167
5.2 Solubility profiling	167
5.3 Construction of a recombinant plasmid for improving the yield of Not1	167
5.4 Purification	168
5.4.1 Not1 protein production and purification	168
5.4.2 Pho92 purification	168
5.4.3 Identification of interacting domains	168
5.4.4 Boundary determination for Not1	168
5.4.5 A conserved region of Pho92	173
5.4.6 Reconstitution of the Not1-Pho92 complex	174
5.4.6.1 Co-purification of Not1 & Pho92	174
5.4.6.2 Anisotropy	174
5.4.6.3 Crystallization trial	176
5.4.7 Reconstitution of the complex from different organisms- <i>C. glabrata</i> and <i>Z. rouxii</i>	176
5.4.8 Limited proteolysis	178
5.4.9 Co-crystallization trial with Not1-PM193 peptide	178
5.4.10 A Not1-Pho92 fusion	180

## **Discussion**

### **Chapter 6 - Discussion of Mmi1, Mei2, Erh1**

6.1 Mmi1-Erh1 and Mei2	187
6.1.1 Mei2-RRM3 is a novel RNA binding domain	187
6.1.2 Mmi1-Mei2 interaction	188
6.1.3 Mmi1 purification trial	188
6.1.4 Role of Erh1	190
6.1.5 Erh1 and Mei2	191
6.2 Conclusion	191
6.3 Future prospect	192

**Chapter 7 – Discussion of Pho92 & Not1**

<i>7.1 Pho92 and Not1</i>	<i>197</i>
<i>7.2 Conclusion</i>	<i>198</i>
<i>7.3 Future prospect</i>	<i>198</i>

**References**

<i>m6A mRNA Destiny: Chained to the rhYTHm by the YTH-Containing Proteins</i>	<i>Annex I</i>
<i>PDB validation report for Erh1</i>	<i>Annex II</i>



**CHAPTER I**  
**INTRODUCTION**



## ***1.1 Cell Cycle***

A cell must arise from a previous cell, this concept was proposed by a German pathologist, Rudolf Virchow (1858). Cell cycle is a series of events to produce a progenitor cell. To explain in simple words, cell cycle is a process where cell duplicates all its genetic material and divides in two, it is the essential process for reproduction of all kind of organisms. For single cell organisms, like bacteria, each cell division leads to a new organism. In multicellular organism, reproduction is more complex and multiple steps of complex cell division are required to produce an organism from a unicellular zygote. Eukaryotic cells have an intricate process regulating cell cycle, where multiple proteins are involved in regulations at different stages of progression through cell cycle.

The main function of cell cycle is to accurately duplicate all the genetic material of a cell and then to divide it precisely in two daughter cells. Replication of genetic material occurs in S-phase (synthetic phase) followed by chromosome segregation and cell division or mitosis in M-phase (mitotic phase). In eukaryotic cell division S-phase takes much longer time than M-phase. The cells require longer time to grow and replicate the organelles and proteins than to synthesize DNA and divide. This period when a cell prepares for division is called Gap-phase. There are two gap-phases, G1 between a mitotic phase and the S-phase of next cycle and G2 phase between S-phase and consecutive M-phase (Fig.1a). The two gap phases allow cell to have time to grow and monitor the environment to determine if the conditions are suitable for cell division. Thus eukaryotic cell cycle can be broadly divided in two phases, interphase that combines G1, S, and G2 phases and M phase (Fig.1a). In human beings, cell division is approximately a 24 hours long process with 23 hours consisting of interphase and one hour for mitosis.

### ***1.1.1 Stages of cell cycle***

**G1:** The G1 phase represents the interval between mitosis and initiation of DNA replication for the next round of cell division. In this phase, cell is metabolically active and growing. At a certain point, known as the restriction point, the cell commits to the next round of cell cycle and enters S phase.

In adult animals sometimes, some cells cannot proceed to cell division (due to injury). These cells cannot exit G1 phase and they arrest at a phase known as G0 stage or quiescent stage. These cells remain metabolically active but they do not enter cell cycle unless induced to do so.

**S phase:** This is the phase where replication of genetic material takes place. At the end of this phase each chromosome has two sister chromatids. If initial amount of DNA is denoted as  $2C$ , it will become  $4C$  after S-phase but the number of chromosomes stays fixed. In a diploid organism, if the chromosome number is denoted as  $2n$ , it stays  $2n$  at end of the cycle.

**G2:** G2 corresponds to the gap between S-phase and M-phase. In this phase, all the proteins are produced and the cell prepares for division. This is a crucial stage as DNA replication errors are corrected during this step, before chromosome segregation and cell division occur.

### ***Mitosis or M-phase***

In this phase, the chromosomes that were synthesized in the preceding S-phase, have to be accurately segregated and divided in two daughter cells (Fig.1b). M-phase is divided in five phases:

- i. Prophase: Prophase is marked by condensation of replicated chromosomes. A cell organelle, named centrosome, replicates to produce two daughter centrosomes. The two daughter centrosomes position to opposite ends of the cell and form mitotic spindle. The nucleolar membrane breaks down.
- ii. Prometaphase: this phase is marked by complete disintegration of the nuclear envelope. The chromosomes are scattered throughout the cell now. By this time, condensation of chromosomes is complete and each chromosome is made up by two sister chromatids joined together by a centromere. The chromosomes line up in the equator of the cell and kinetochores, small disc-shaped structures, on the surface of centromere connect each sister chromatid, to one pole of the spindle apparatus and the other one to the other pole. The plane of alignment of chromatids is called, the metaphase plane.
- iii. Metaphase: Chromosomes align themselves along the metaphase plane.
- iv. Anaphase: Anaphase is the initiated alignment of all chromosomes on the equatorial plate and splitting of centromere. Each sister chromatid (future chromosome of the daughter cell), now separated, migrates towards the opposite pole.
- v. Telophase: This is the last phase of Mitosis, where the sister chromatids have reached opposite poles, decondense and lose their individual form. The chromosomes gather

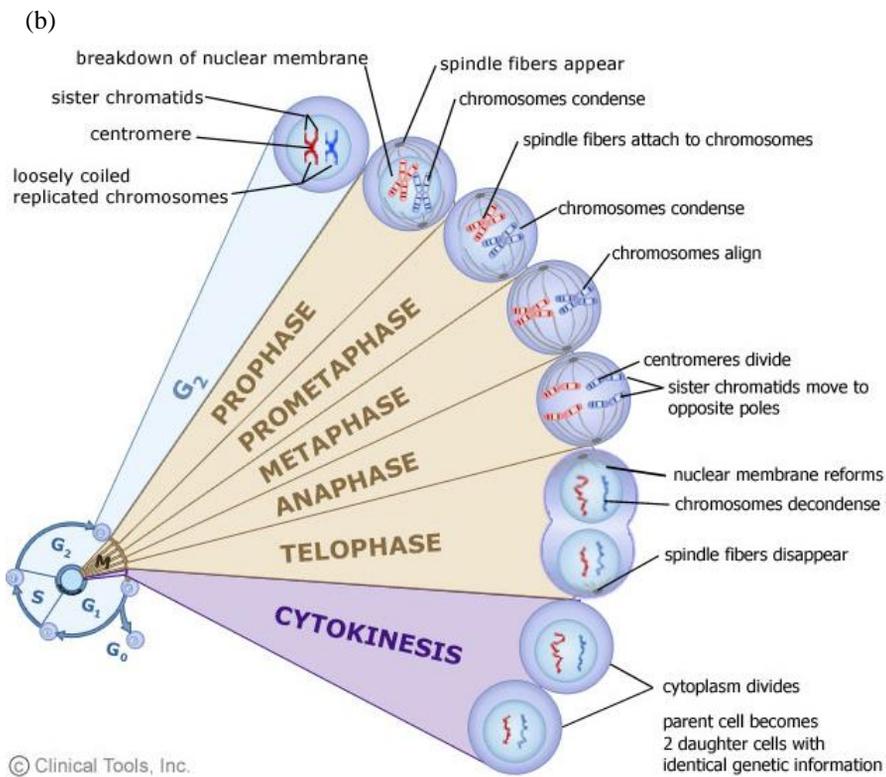
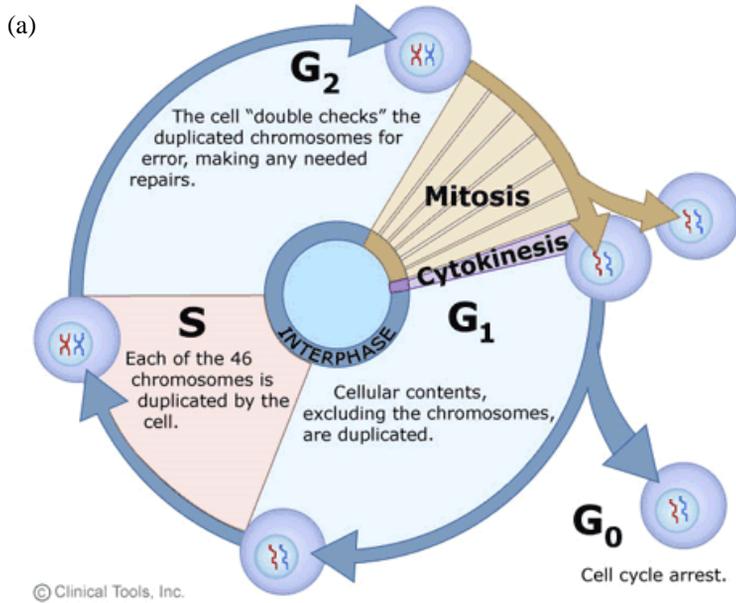


Fig. 1: Schematic representation of eukaryotic cell cycle. (a) different stages of human cell cycle; (b) detailed representation of mitosis steps (image from <https://www2.le.ac.uk/projects/vgec/highereducation/topics/cellcycle-mitosis-meiosis>)

as a mass and the nuclear envelope forms around this cluster. Finally, nucleolus, Golgi apparatus and endoplasmic reticulum (ER) form at the end of this stage.

Cytokinesis: at the end of mitosis, the cell physically divides by a process called cytokinesis. In animal cells, this is performed by the constriction of an actin-myosin ring in cytoplasm while in plant cells, a cell plate forms along the line of the metaphase plate. The newly generated daughter cells then enter interphase, the interval between two mitotic phases.

### ***1.1.2 Meiosis***

Sexual reproduction is carried out by fusion of two gametes each containing a complete haploid set of chromosome. Meiosis is the cellular process that produces haploid gametes from special diploid germ cells maintaining constant number of chromosomes in each generation. If the chromosome number of a diploid organism is denoted as  $2n$ , after meiosis it will be reduced to  $n$  in each daughter cell. Meiosis is the key process that governs production of gametes in higher vertebrates and after fertilization, the diploid phase is restored. Meiosis is a mechanism to conserve the chromosome number in each species. Without meiosis, in every generation the number of chromosomes would have doubled. In mitosis, a round of DNA replication is followed by one round of chromosome segregation and cell division. Meiosis is a specialized cell division process, in which one round of DNA replication is followed by two rounds of subsequent chromosome segregation and cell division. Meiosis is divided in two main phases, meiosis I and meiosis II with each round containing prophase, metaphase, anaphase and telophase. The steps are described by schematic representation (Fig.2). Albeit its biological importance, the intricate network that regulates the process is yet to be elucidated, especially in higher organisms. Over the last few decades however, a substantial amount of information about meiosis was gained by studies conducted in yeasts, specially the *Saccharomyces cerevisiae* budding yeast and the *Schizosaccharomyces pombe* fission yeast. Although yeasts are small unicellular fungi, the regulation of cell cycle in yeast is very similar to cell cycle regulation in higher eukaryotes. The yeasts multiply rapidly, their generation time (around 90 minutes) is closer from that of bacteria (around 30 minutes) than that of human cells. The size of their genome is 1% of mammals and they can exist in haploid state. Therefore, it is easy to study the effect of genetic manipulation in these organisms, this makes yeasts, especially *Schizosaccharomyces pombe* a good model organism to study meiosis.

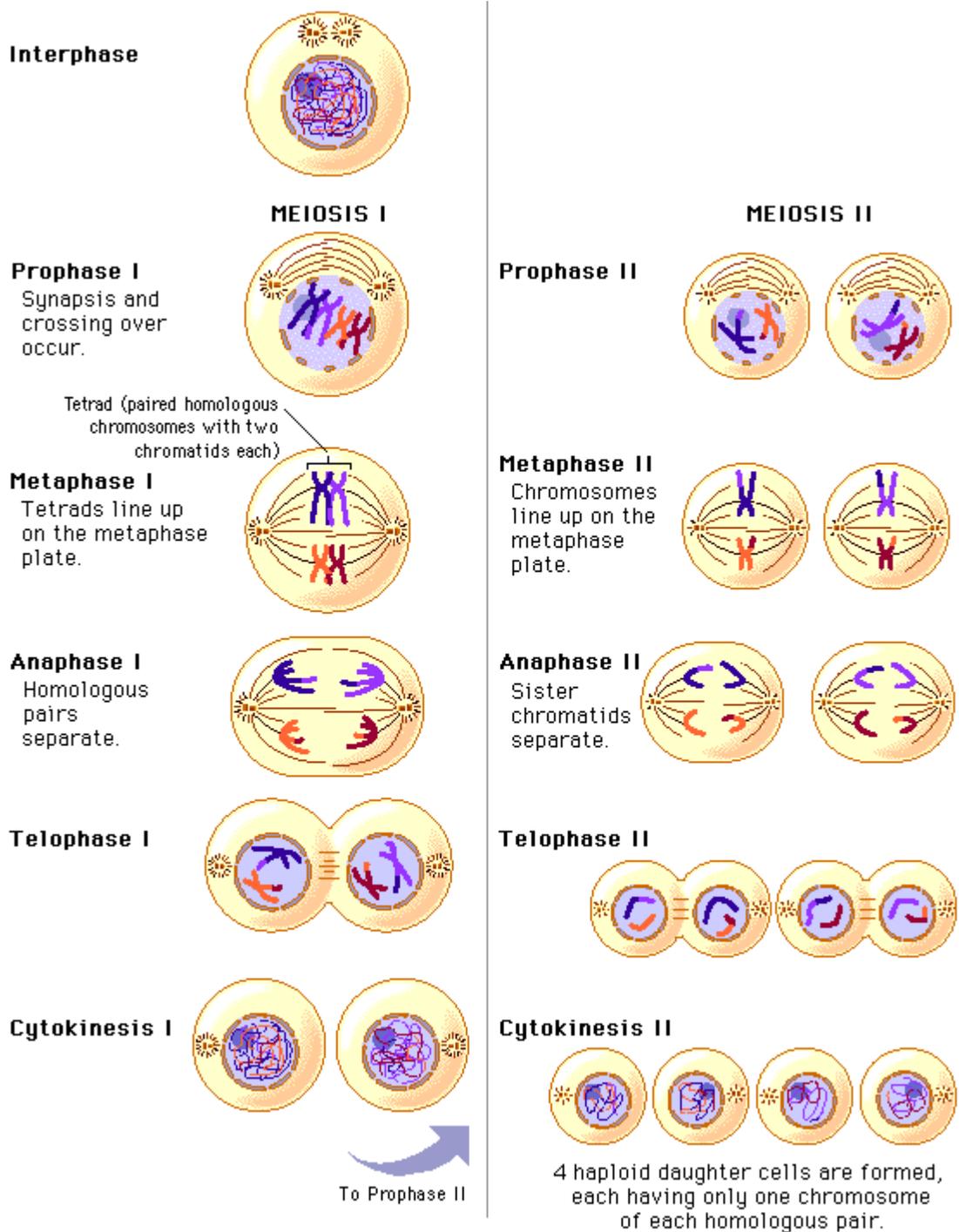


Fig.2: schematic representation of different steps of occurring during meiosis (image source: <https://celldivisionandreproduction.weebly.com/meiosis.html>)

## 1.2 Meiosis in yeast

In yeasts, meiosis leads to production of spores under conditions unsuitable for vegetative growth. Both budding and fission yeasts can grow in haploid or diploid states. Under unfavourable circumstances, two haploid cells of opposite type,  $a$  and  $\alpha$  for *S. cerevisiae* and *mat1-P* and *mat1-M* for *S. pombe* conjugate to form a zygote, containing both mating type genes, and then meiosis leads to generation of spores from the zygote. Albeit their biological importance, the networks and signalling cascades that control meiosis are yet to be elucidated.

Meiosis in lower eukaryotes, like yeasts is generally triggered by environmental signals. The environmental signals induce a cascade of signal transduction leading to meiosis. Usually nutrient deprivation can act as meiosis triggering signal in both type of yeasts. Onset of meiosis in *S. cerevisiae* depends on three conditions: (i) nitrogen starvation, (ii) presence of a non-fermentable carbon source, like acetate and (iii) absence of glucose. The third one is probably most important, since presence of glucose in media can inhibit entry into meiosis even if the other two conditions are met. In fission yeast, depletion of nutrient in the medium, usually nitrogen is enough to signal meiosis. Even if these conditions are met, in both type of yeasts, only those cells that contain both types of mating genes, *MATa/MAT $\alpha$*  for budding yeast and *mat1-M/mat1-P* for fission yeast, can enter into meiosis. The diploid cells that contain same type of mating genes, are arrested at G1 phase of cell cycle (Nurse et al., 1976; Forsburg and Nurse, 1991).

For *S. pombe*, conjugation, mating and spore formation is a response to an emergency situation, such as nutrient depletion or stress. Under unfavourable conditions, haploid fission yeast cells of different mating type go through conjugation to form a diploid which rapidly undergoes meiosis and forms ascospores (Fig.3a). The fission yeast diploids are quite unstable but they can be maintained in a diploid state by transferring them in a rich media before meiosis occurs. The budding yeast can mate even in rich medium. If there are haploid cells of two different mating types, they can conjugate and form a diploid. This diploid can't stably grow by mitosis. In case of nutrient deprivation, the diploid carrying both types of mating genes undergoes meiosis and forms spores (Fig.3b). So, although in both budding and fission yeasts, mating is driven by a peptide pheromone, budding yeasts normally grow in diploid state and conjugation takes place even in rich medium while fission yeasts usually

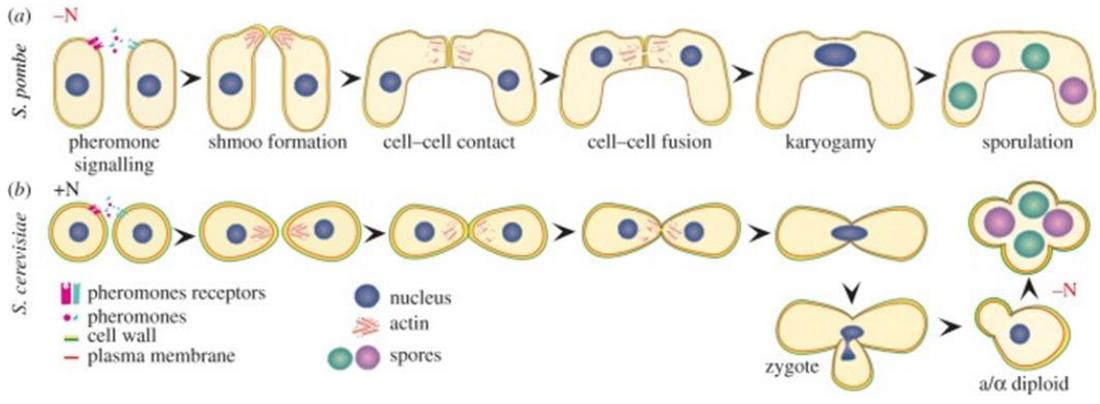


Fig.3: Sequential steps of mating in (a) *S. pombe* and in (b) *S. cerevisiae* (Merlini et al., 2013)

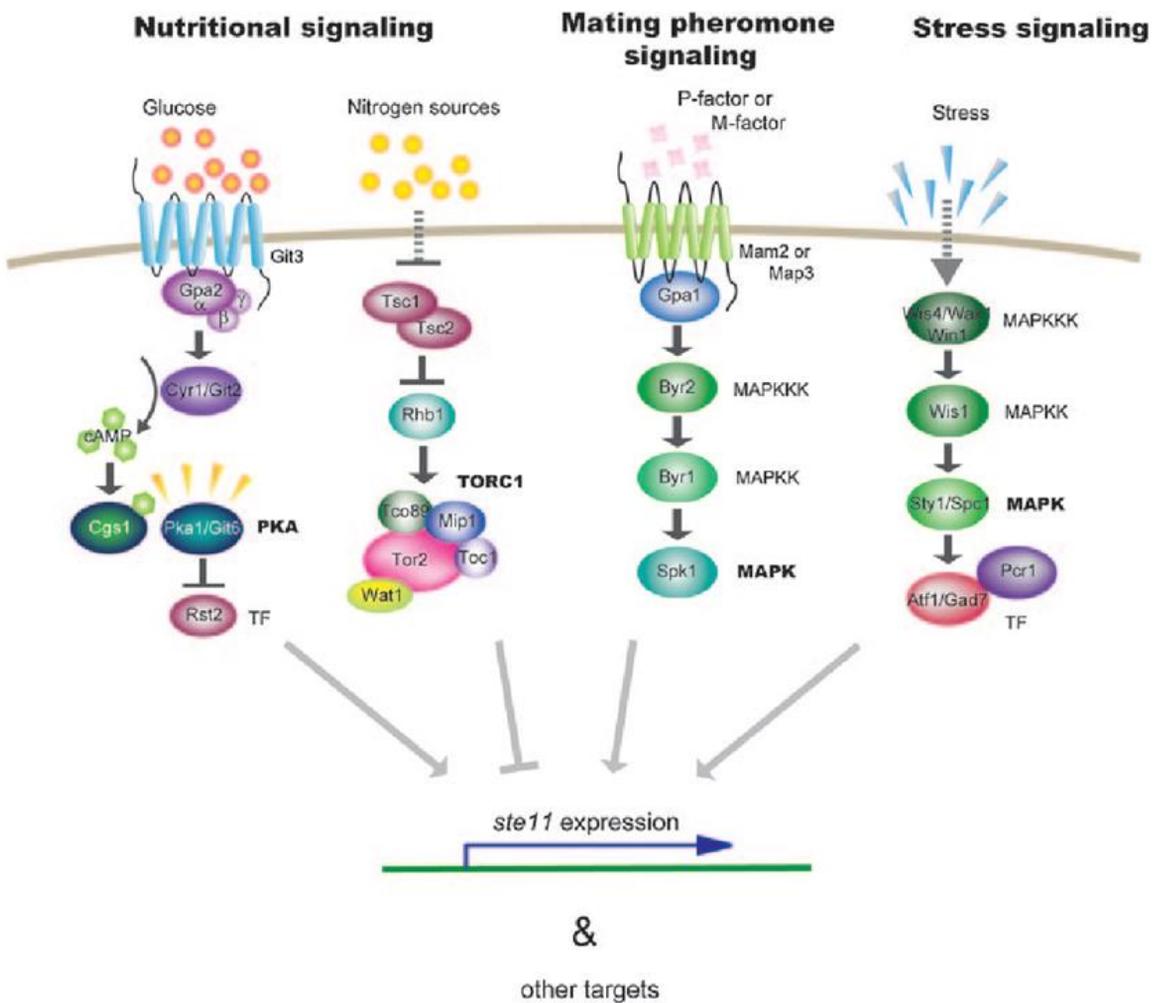


Fig.4: Signal transduction pathways of meiosis in fission yeast (Yamamoto 2010)

proliferate as haploid organism and hence conjugation and sporulation are triggered by nutrient depletion.

### ***1.2.1 Meiosis in fission yeast***

As mentioned earlier, meiosis in *S. pombe* can be triggered by introducing a nutritional shift. When heterozygous cells carrying opposite mating type genes (mat1-P/mat1-F) are exposed to nutrient depletion (especially nitrogen starvation) or stress, they conjugate, forming a zygote (2n). When this zygote enters meiosis, after one round of DNA replication and two rounds of meiotic division followed by cell divisions, asci are produced, each containing four haploid spores (n). Meiosis is regulated by ordered transcription of numerous genes.

#### ***1.2.1.1 Entry to meiosis***

In *S. pombe*, mitosis to meiosis switch starts with expression of *ste11*<sup>+</sup>. Ste11 is an HMG (high mobility group) protein that increases expression of some meiosis specific genes by binding to a consensus motif, TR box, present on the promoters of these genes. Ste11 regulates transcription of many genes, one of the most important target of Ste11 is an RNA binding protein, *mei2*<sup>+</sup>. In vegetatively growing cells (nutrient rich media), ATP is converted to cAMP (cyclic AMP). cAMP triggers a signalling cascade that up-regulates PKA (protein kinase A). PKA phosphorylates a transcription activator Rst2 leading to its inactivation and subsequent down-regulation of *ste11*<sup>+</sup>, this pathway is described in details in the next section (Otsubo and Yamamoto, 2012). Additionally, it has been shown that Ste11 is active only in G1 phase and phosphorylation of Ste11 at T82 by cyclin dependent kinase Cdk renders it incapable of binding to the target DNA (Kjærulff et al., 2007). Pat1 kinase is also known to inactivate Ste11 by phosphorylation (Li and McLeod, 1996), resulting in increased affinity of Ste11 for Rad24, a 14-3-3 protein causing inhibition in DNA binding (Kitamura et al., 2001).

Meiosis in fission yeast has been thoroughly studied by Yamamoto group. Four signal transduction pathways trigger the mitosis to meiosis switch in *S. pombe*, they are: glucose starvation, nitrogen starvation, mating pheromone signalling and stress signalling (Otsubo and Yamamoto, 2012) (Fig. 4). All these regulations converge to up-regulation of *ste11*<sup>+</sup> gene.

#### ***Glucose starvation***

Glucose starvation regulates the cAMP pathway. cAMP is an important second messenger which is essential for vegetative growth of *S. pombe* (Maeda et al., 1990). In nutrient rich

media, ATP is converted to cAMP (Kawamukai et al., 1991). The cAMP in turn suppresses *ste11* induction by phosphorylating a zinc finger transcription factor, Rst2 by PKA (protein kinase A) (Higuchi et al., 2002). Rst2 induces transcription of *ste11* by binding to its upstream region (Kunitomo et al., 2000) and phosphorylation of Rst2 renders it incapable of this action, thereby suppressing *ste11+* expression. Lack of glucose in the media decreases cAMP production, preventing activation of PKA. As a result, Rst2 is not phosphorylated and it can induce transcription of *ste11+*.

### ***Nitrogen starvation***

Nitrogen depletion also leads to up-regulation of *ste11+* in a TOR (Target of Rapamycin) kinase dependent manner. TOR kinases are a group of Ser/Thr kinase that are highly conserved among eukaryotes and they play an important role in regulation of cell growth and metabolism. Fission yeast has two copies of Tor kinases, Tor1 and Tor2. Tor1 is necessary for sexual development under conditions unfavourable for vegetative growth whereas Tor2 is indispensable for vegetative growth. Like other organisms, *S. pombe* Tor kinases are part of two different complexes: TORC1 and TORC2. TORC1 contains Tor2 as the major catalytic subunit. TORC1 regulates propagation from G1 to S-phase and inhibits sexual development in rich media (Otsubo and Yamamoto, 2010). Tor1 is the main component of TORC2, which regulates cell growth and metabolism during vegetative growth. Further studies showed that mutation in Tor2 can mimic nitrogen starvation even in nutrient rich media (Alvarez and Moreno, 2006; Uritani et al., 2006; Weisman et al., 2007; Matsuo et al., 2007). Additionally, it was found that the targets of Ste11, including *ste11+* itself and *mei2+* are up-regulated in absence of Tor2 and down-regulated when Tor2 is over-expressed (Matsuo et al., 2007; Valbuena and Moreno, 2010). Tor2 is also known to down-regulate Mei2 by phosphorylation.

### ***Mating signal***

Mating in fission yeast is regulated by pheromones. Mating pheromone, a peptide of 23 amino acids, known as P-factor, is produced by *mat1-P* type cells. The *mat1-M* type cells produce another pheromone, a farnesylated peptide of 9 amino acids, known as the M-factor. Upon nitrogen starvation, cells of opposite mating types produce the pheromones. Each pheromone then binds to their specific receptors present on the cell surface of opposite mating type, P-factor binds to Mam2 receptor on *mat1-M* type cells and M-factor binds to Map3 on *mat1-P* type cells (Tanaka et al., 1993; Kitamura and Shimoda, 1991; Toda et al.,

1991). Upon binding of the pheromones, the two receptors dock to the same protein, Gpa1 and activate it. Activated Gpa1 triggers a MAP kinase signalling cascade, that finally leads to up-regulation of the expression of *ste11+*.

### ***Stress signal***

Stress response leads to activation of different MAP kinase pathway. A product of this pathway is MAPK Sty1 that accumulates in the nucleus and phosphorylates several targets. Sty1 targets include Atf1-Gad7 and Pcr1 leading to activation of *ste11+* (Sukegawa et al., 2011).

#### ***1.1.2.2 Transcriptional regulation of meiosis***

Meiosis in fission yeast starts with transcription of *ste11+*. Therefore, transcriptional regulation of *ste11+* plays an important role in regulation of meiosis. The *ste11+* transcript has a long (>2kb) 5' UTR that functions as binding sites of many regulators. The most important of these regulators is probably Rst2, a zinc finger motif containing protein that can recognize a stress response element (STRE) in the *ste11+* promoter site (*i.e.* 5'-CCCCTC-3' motif) (Kunitomo et al., 2000). Activation of Rst2 depends on the nutritional condition of the media. As described earlier in nutrient depleted media, Rst2 is in its active form thereby inducing expression of *ste11+* gene (Kunitomo et al., 2000; Higuchi et al., 2002). Targets of Ste11 include *ste11+* itself, genes responsible for mating and the master regulator meiosis, *mei2+*. Ste11 up-regulates its own expression by binding to a TR box present in the promoter of *ste11+* gene.

The other important target of Ste11 is the *mei2+* gene codes for Mei2, an RNA binding protein, indispensable for pre-meiotic DNA synthesis as well as entry into Meiosis I (Watanabe and Yamamoto, 1994). It plays a key role in switching from mitosis to meiosis.

### ***1.3 Mei2- the master regulator of meiosis***

The *mei2+* gene was identified by screening for mutants that are incapable of entering meiosis (Bresch et al., MGG, 1968). Functional characterization revealed that Mei2 is required for pre-meiotic DNA synthesis and the first meiotic division (Watanabe and Yamamoto, 1994). Mei2 orthologues are found in fungi from the *Schizosaccharomyces* phylum but evolutionarily conserved Mei2-like proteins are also highly abundant in plants.

According to bioinformatics analyses, the *S. pombe* Mei2 protein is composed of three RRM (RNA recognition motifs) domains (Fig.5). The N terminal RRM domains (RRM1 and RRM2) are contiguous (residues 194-265 and 287-362) while the third RRM domain (RRM3) is located at the C-terminal extremity of Mei2 (596-685). This later RRM domain is absolutely necessary for entry into meiosis (Watanabe et al., 1997).

### ***1.3.1 Structure and RNA binding mode of an RRM domain***

The RRM domain is the most abundant RNA binding domain in vertebrates, consequently it is the most thoroughly studied RNA binding domain as well (Mariset al., 2005; Cléry and Allain, 2012). This domain is responsible for single-stranded RNA binding. RRM domain is an approximately 90 amino acid long domain that typically folds as a four-stranded antiparallel  $\beta$ -sheet (with a  $\beta_2\beta_3\beta_1\beta_4$  topology) stacked against two  $\alpha$ -helices ( $\alpha_1$  and  $\alpha_2$  in the back) (Fig.6a). In classical RRM domains, the  $\beta$ -sheet surface is involved in RNA recognition via two conserved motifs of respectively eight and six residues, RNP1 and RNP2 (Fig. 6b). The hydrophobic residues from RNP1 (positions 3 and 5) and RNP2 (position 2) are located in the  $\beta$ -sheet and are involved in interaction with the nucleic acid. The bases of the nucleotides are stacked on the aromatic ring on strands  $\beta_1$  (position 2 of RNP2) and  $\beta_3$  (position 5 of RNP1). The other conserved aromatic amino acid on  $\beta_3$  (position 3 of RNP1) is often inserted between the two sugar rings. The R/K (position 1 of RNP1) forms salt bridge with the RNA backbone. This conserved binding mode can accommodate two nucleotides in the centre of a  $\beta$ -sheet and one/two nucleotides on either side. The side chains of aromatic residues of  $\beta$ -sheet specifically recognize a WG (W: A/C) dinucleotide. In nature, RRM domain containing proteins are able to bind divergent sequence and length of RNA with high affinity. This is made possible by different binding mechanism outside the  $\beta$  sheet.

### ***1.3.2 Localisation of Mei2***

Mei2 is capable of shuttling between cytoplasm and nucleus. During vegetative growth, Mei2 is expressed in very low amount and accumulated only in the cytoplasm. During meiosis, Mei2 forms a single nuclear dot, known as the Mei2 dot (Fig. 7). This dot is comprised of Mei2 protein and a long non-coding RNA, meiRNA which is encoded by *sme2+* gene (Watanabe et al., 1997; Yamashita et al., 1998). The Mei2 dot is associated with *sme2+* locus on Chromosome II (Yamashita et al., 1998). Formation of the dot indicates that the cell is able to enter meiosis I. The  $\Delta$ *sme2+* cells are incapable of initiating meiosis (Yamashita et al., 1998). The role of Mei2 dot is yet to be fully elucidated to understand the underlying regulations controlling mitosis to meiosis switch in fission yeast.



Fig.5: Schematic representation of domain organization of Mei2

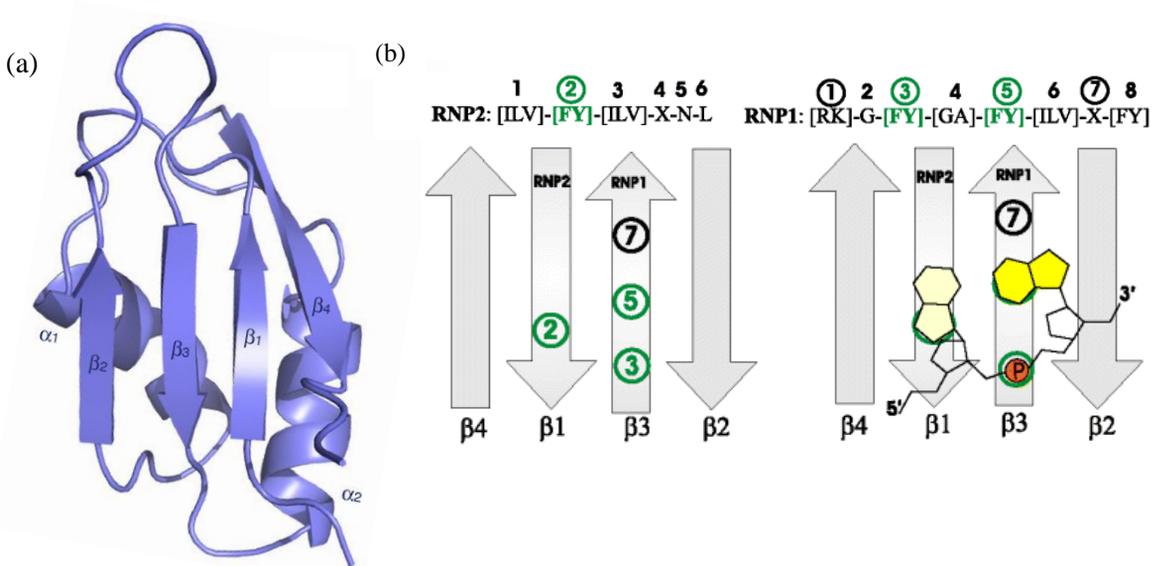


Fig.6: RRM domain. (a) Structure of a classical RRM domain, comprising of four antiparallel  $\beta$  strands stacked against two  $\alpha$  helices (adapted from PDB 2CQI) (b) Schematic representation of RRM domains of hnRNPA1, the conserved aromatic residues from RNP1 and RNP2 are highlighted in green (left). Scheme of interaction between nucleic acid and aromatic residues from RNP (right) (Clery et al., 2008)

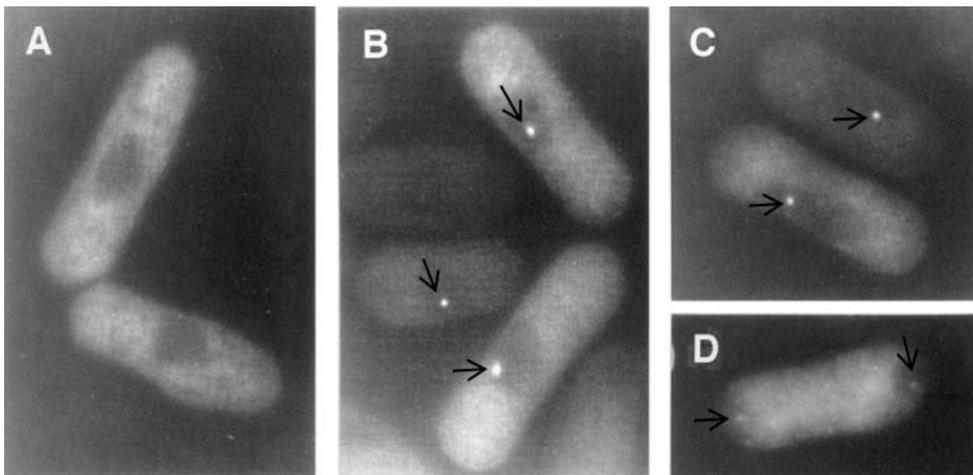


Fig.7: GFP-tagged Mei2 in fission yeast cells, grown in (A) rich medium or (B) depleted of nitrogen for 3,5 hs, (C) 4 hs and (D) 6 hs. Cells in (A) are vegetatively growing, cells in (B) and (C) are undergoing meiotic prophase and cells in (D) are going through the first meiotic division. Black arrowheads indicate Mei2 dot (adapted from Yamashita et al., 1998).

### ***1.3.3 Substrates of Mei2***

As mentioned earlier, another component of this Mei2 dot is the meiRNA transcribed from *sme2+* locus. It has been proposed that meiRNA sequesters Mei2 in the nucleus preventing its export to cytoplasm (Sato et al., 2001). However, there are experimental evidences suggesting that it is the RNA binding property of Mei2 rather than its specific binding to meiRNA that regulates mitosis to meiosis switch. Indeed, when Mei2 is transported to nucleus with SV40 nuclear localisation motif in the absence of meiRNA, meiosis can be initiated but a Mei2 mutant (Mei2p-644A) incapable of RNA binding when transported to nucleus by SV40 nuclear localisation motif, it cannot initiate meiosis (Yamashita et al., 1998). This indicates that there are possibly other RNA species that can interact with Mei2 and play a role in meiosis initiation.

### ***1.3.4 Function of Mei2***

Mei2 is indispensable to start meiosis in fission yeast. The function of this protein is not fully characterized but it has been well established that Mei2 plays a role in pre-meiotic DNA synthesis and initiation of meiosis I (Watanabe and Yamamoto, 1994).

Mei2 is essential for pre-meiotic DNA synthesis and its RNA binding ability is required for this function. However, meiRNA is not essential. This suggests that there might be at least another RNA species that can initiate pre-meiotic DNA synthesis in complex with Mei2 (Watanabe and Yamamoto, 1994).

The other function of Mei2 is to initiate meiosis I. Mei2 carries out this function by sequestering another protein, Mmi1 (meiotic mRNA interception 1), in the nucleus, with the help of meiRNA. Mmi1 is responsible for degradation of meiosis specific transcripts during vegetative cell growth (described in details in section 2.4). During meiosis, Mei2 inhibits the function of Mmi1 by sequestering it with the help of meiRNA, preventing degradation of meiotic transcripts and ensuring smooth progression through meiosis (Harigaya et al., 2006; Yamashita et al., 1998).

### ***1.3.5 Regulation of Mei2***

#### ***Pat1-Mei2 system***

Pat1 is Ser/Thr kinase that prevents *S. pombe* to enter meiosis under conditions suitable for vegetative growth (Fig. 8). In nutrient rich conditions, Pat1 kinase phosphorylates both Mei2

and Ste11 leading to their degradation. Pat1 phosphorylates Mei2 on two residues S438 and T527 located in the linker region connecting RRM2 to RRM3 (Watanabe et al., 1997). Pat1 regulates Mei2 in two ways. First, phosphorylated Mei2 is susceptible to ubiquitination and degradation by the proteasome. Secondly, phosphorylated Mei2 has increased affinity for 14-3-3 protein Rad24, which inhibits its RNA binding property (Kitamura et al., 2001). In presence of both mating type genes, *mat1-M* and *mat1-P*, in zygotes or diploids, under nutrient starvation, Mei3 protein is expressed (McLeod, and Beach, 1988; Hoffmann et al., 1995; Li and McLeod, 1996; van Heeckeren et al., 1998). Mei3 is a pseudo-substrate of Pat1. Expression of Mei3 allows Mei2 to avoid phosphorylation and therefore the stability of Mei2 is increased.

### ***Tor2-Mei2 system***

The level of Mei2 is also controlled by Tor2 signalling. TORC1 phosphorylates Mei2 at 9 positions. Mip1 (Mei2 interacting protein 1), a component of TORC1, as well as Tor2 can physically contact Mei2. Under conditions suitable for vegetative growth, Tor2 phosphorylates Mei2 leading to its ubiquitin-proteasome mediated degradation (Otsubo et al., 2014). The Tor2 phosphorylation sites are different from Pat1 phosphorylation sites. The current model proposes that nutrient starvation leads to down-regulation of Tor2 allowing a population of unphosphorylated Mei2 to accumulate. This, accompanied by cell cycle arrest in G1 and mating leads to transcription of *ste11+*. In addition to expression of Mei3, nutrient starvation increases level of unphosphorylated Mei2 and favours initiation of meiosis (Fig. 8).

### ***1.4 Selective elimination of meiosis specific mRNAs***

The Yamamoto group has observed a curious phenomenon, that during vegetative growth of cell, the meiosis specific transcripts did not accumulate even when transcribed artificially by constitutive promoter. The study was conducted on the following four genes:

- i. *mei4*, a transcription factor required for meiosis I,
- ii. *rec8*, which encodes for a subunit of meiosis-specific cohesion complex,
- iii. *ssm4*, encoding a homologue of the dynactin component,
- iv. *spo5*, an RRM type RNA binding protein required for meiosis II

In 2006, they showed that these meiosis specific transcripts harbour a conserved hexanucleotide motif U(U/C/G)AAAC. They named it determinant of selective removal (DSR)

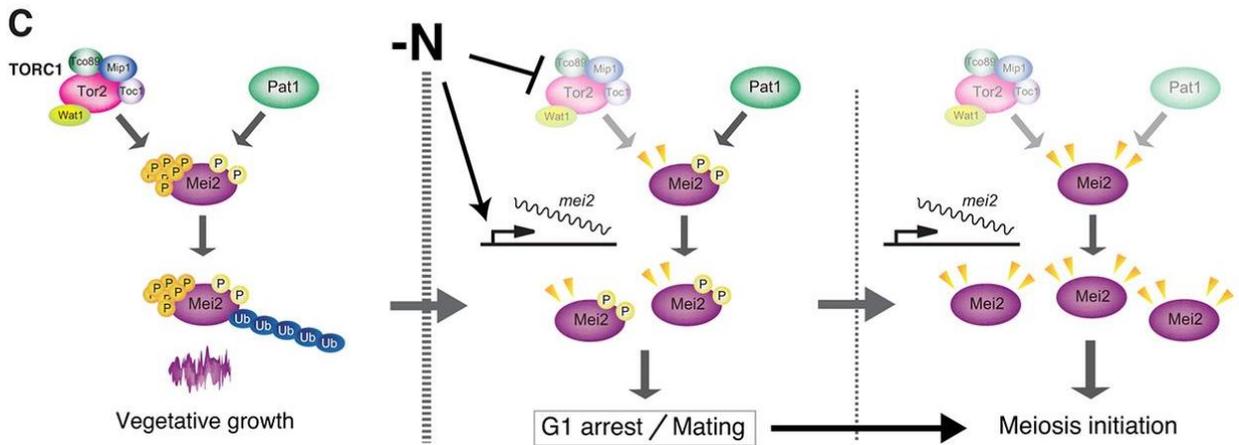


Fig.8: A proposed model to describe the regulatory role of Mei2 in switch from vegetatively growing cells to mating and meiosis. Mei2 is under the regulation of two kinases: Pat1 and Tor2. In Nitrogen starved condition Mei2 proteins experience reduced phosphorylation by TORC1 but that are still phosphorylated by Pat1 kinase and can stimulate the mating process. When Mei2 is no longer phosphorylated by Pat1 it promotes meiosis without mating. P: phosphorylation; Ub: ubiquitylation. (adapted from Otsubo et al., 2014)

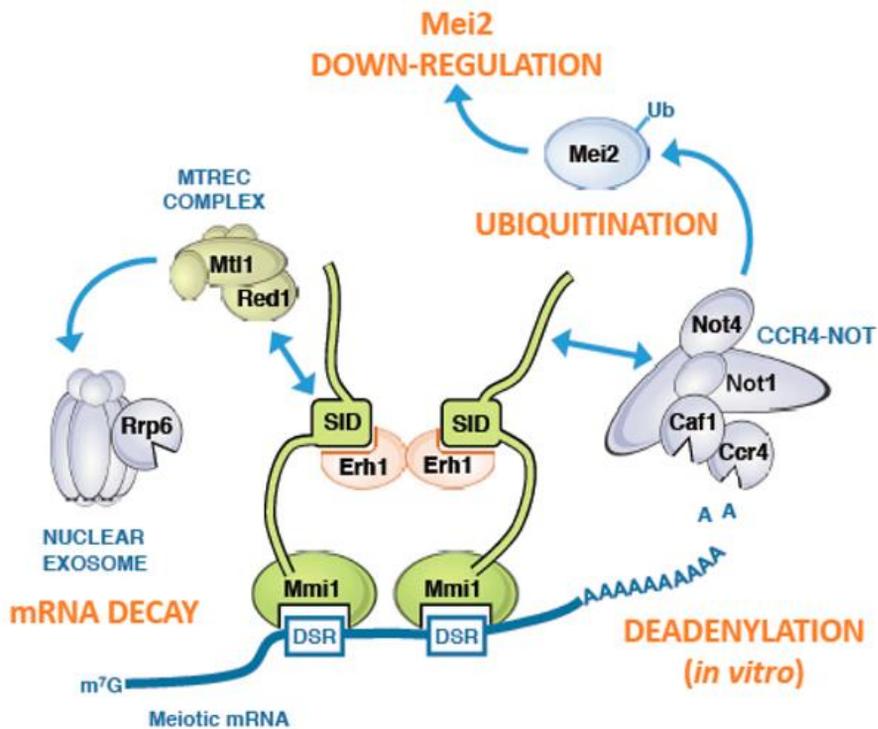


Fig.9: Schematic representation of Mmi1 mediated degradation of meiotic transcripts and Mei2 protein in vegetative cells (Hazra et al., 2019)

(Harigaya et al., 2006) and proposed that Mmi1 degrades meiotic transcripts in a DSR dependent manner. A screen for mutants incapable of DSR mediated degradation led to the discovery of Mmi1. Further genetic analysis revealed that the target genes for DSR mediated removal contain UUAAAC or UCAAAC sequences, termed as “core DSR motif” (Yamashita et al., 2012). The product of *sme2+* gene, *meiRNA* also contains multiple repeats of these core DSR motifs.

### ***1.5 DSR mediated RNA degradation by Mmi1***

Mmi1 is a protein of 488 residues, harbouring a C-terminal RNA binding YTH domain. Mmi1 can specifically recognize a consensus U(U/C/G)AAAC motif, present in multiple copies in an RNA called Determinant of Selective Removal (DSR). YTH domain of Mmi1, is responsible for the binding to the DSR and is assisted in this function by a low complexity region located upstream this YTH domain (Stowell et al., 2018). Except the YTH domain, Mmi1 does not have any other structured domain, which could be directly involved in RNA decay, indicating that Mmi1 recruits other effectors such as the exosome to degrade the meiotic transcripts. Indeed, the necessity of nuclear exosome in DSR dependent transcript degradation has been well established (Yamanaka et al., 2010). It has also been shown that Mmi1 co-localizes and interacts with Erh1, a small protein with still unclear function, to form the Erh1-Mmi1 Complex (EMC) (Sugiyama et al., 2016). EMC leads to degradation of meiotic transcripts or formation of heterochromatin island on chromosome loci of the meiotic genes (Fig. 9).

#### ***1.5.1 The MTREC complex***

MTREC, a multi-subunit protein complex, is one of the main effectors involved in DSR dependent degradation of meiotic RNAs by Mmi1. The core component of this complex contain, (i) Mtl1, an RNA helicase, with similarity to Mtr4, an RNA helicase from the TRAMP complex that recruits nuclear exosome to its RNA targets, and (ii) Red1, a zinc-finger containing protein (Lee et al., 2013; Eagan et al., 2014). Further analysis revealed that the MTREC complex is comprised of other co-factors: Poly(A) binding protein Pab2, the Pro/Ser rich factor Iss10/Pir1, zinc finger-containing proteins Red5 and Ars2, the RRM-containing protein Rmn1, two predicted nuclear cap binding proteins (Cbc1, Cbc2) and a loosely associated poly(A) polymerase Pla1 (Zhou et al., 2015) (Fig. 10).

Red1, a protein required for vegetative growth and sporulation, is essential for DSR mediated RNA decay in vegetatively growing cells. Red1 interacts with Mmi1 and during mitosis, it

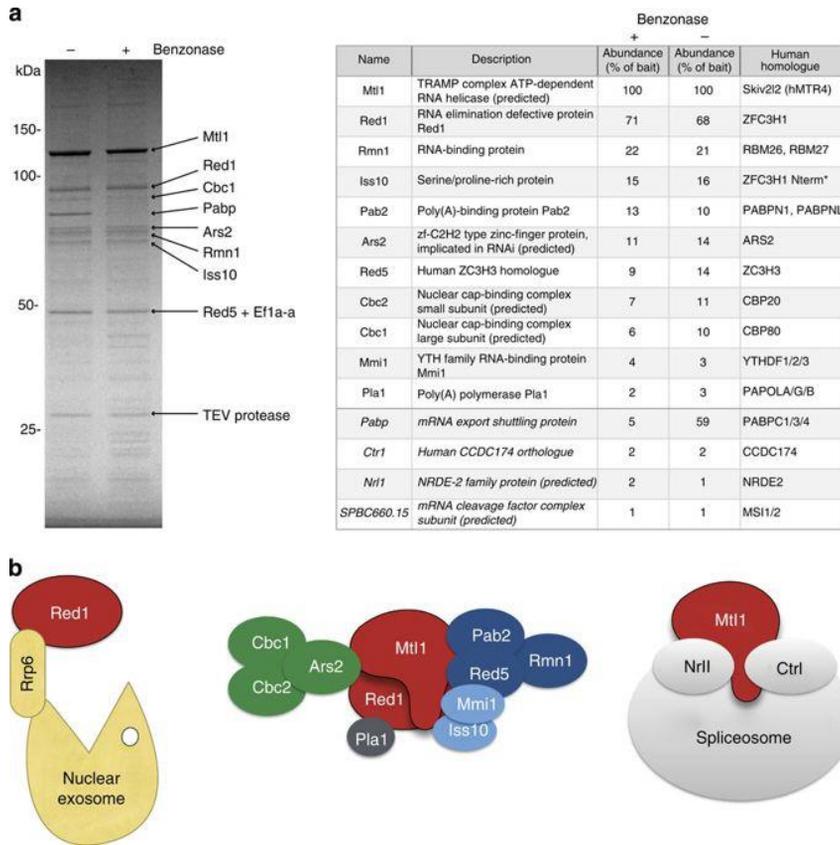


Fig.10: Mmi1 interacts with the MTREC complex (a) SDS-PAGE analysis of tandem-affinity purified MTREC complex using Mtl1 as a bait in presence and absence of Benzonase. The LC-MS/MS analysis of purified total protein is presented in tabulated form. (b) Schematic representation of the Mtl1 and Red1 containing complexes. Different colours represent submodule organisation (adapted from Zhou et al., 2015)

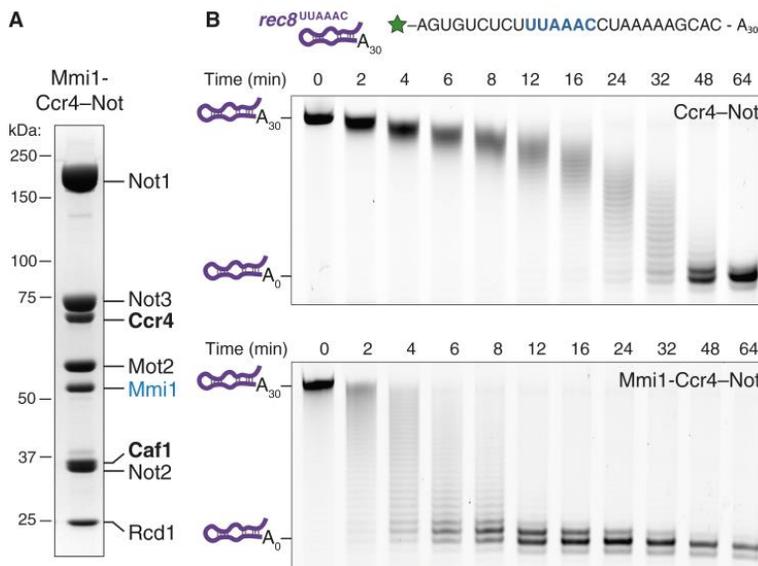


Fig.11: Mmi1 also interacts with CCR4-Not complex (A) SDS-PAGE analysis of Mmi1 co-purified with Ccr4-Not complex. (B) SDS-PAGE analysis of deadenylation of a target RNA by Ccr4-Not complex in absence (upper gel) and presence (lower gel) of Mmi1 (adapted from Stowell et al., 2016)

co-localizes with Mmi1 and the exosome, together with other protein partners essential for DSR mediated removal: Pla1, a classical polyadenylation protein, responsible for polyadenylation of both meiotic and non-meiotic transcripts and Rrp6, a subunit of the exosome, responsible for degradation of polyadenylated transcripts. Probably, RNA bound Mmi1, recruits MTREC leading to degradation of the transcripts by exosome. It has been observed that deletion or mutation of the MTREC subunits lead to accumulation of meiotic transcripts during mitosis but the exact mechanism by which Mmi1 recruits MTREC for DSR-mediated degradation is yet to be explored. It has been proposed that MTREC subunits target different classes of transcripts (DSR containing transcripts, cryptic unstable transcripts, unspliced pre-mRNA etc.) and feed them to a large machinery like nuclear exosome or splicing machinery leading to their degradation. This is supported by strong interaction between different subunits of the MTREC complex: Pla1, Red1, Mlt1 and Rrp6. The importance of polyadenylation of meiotic transcripts for degradation has been well established. A working model proposes that the transcripts are polyadenylated by Pla1 leading to binding of a poly(A) binding protein Pab2 which recruits nuclear exosome by Rrp6. This theory is supported by the experimental evidence that mutations in Pab2 or Rrp6 results in accumulation of meiosis specific transcripts having unusually long poly(A) tails (Yamanaka et al., 2010).

### ***1.5.2 Ccr4-Not complex***

Carbon catabolite repression 4-negative on TATA-less (Ccr4-Not) complex is a highly conserved eukaryotic large multi-subunit complex, which regulates gene expression in multiple levels. In yeast (*S. cerevisiae*), Ccr4-Not complex is made up of 9 subunits, Ccr4, Caf1, Caf40, Caf130, and Not1-5 (Collart and Panasenko, 2012). Ccr4-Not complex has two enzymatic functions: ubiquitination and deadenylation. Ccr4-Not has exonuclease activity that shortens poly(A) tail at the 3' end of RNA leading to its degradation. Ccr4 and Caf1 are the subunits endowed with the exonuclease activity and these two proteins physically interact with each other (Basquin et al., 2012). Not1 is the scaffolding protein which contacts other subunits of Ccr4-Not complex. Not2, Not3 and Not5 contain a Not-box motif, which facilitates interaction among the subunits (Bhaskar et al., 2013, Boland et al., 2013). These subunits also participate in RNA degradation by promoting decapping. Not4/Mot2 has a E3 ubiquitin ligase subunit that has been shown to be associated with Mmi1 to promote degradation of meiotic transcripts in vegetatively growing cells (Bhaskar et al., 2015; Simonetti et al., 2017). Recent works have also reported a tight association between Mmi1

and the Ccr4-Not complex, resulting in its recruitment to meiosis specific transcripts *in vivo* as well as in the stimulation of deadenylation activity *in vitro* (Stowell et al., 2016; Ukleja et al., 2016) (Fig. 11). However, this function of the Ccr4-Not complex is not mandatory for the degradation of meiotic mRNAs (Cotobal et al., 2015). Instead, Ccr4-Not is required to maintain the integrity of heterochromatin formation at genomic islands and sub-telomeres (Sugiyama et al., 2016). Mmi1 has also been reported to maintain heterochromatin signatures in genomic islands belonging to its targets, indicating Mmi1 recruits this complex. Polyadenylated tails are a prerequisite of Mmi1 mediated degradation (Yamanaka et al., 2010) and Ccr4-Not shortens poly(A) tails in cytoplasm. The working theory proposes that Ccr4-Not complex probably counteracts polyadenylation thereby increasing stability of the meiotic transcripts and leading to their accumulation in heterochromatin island, before being targeted for DSR mediated decay (Cotobal et al., 2015).

### ***1.6 Erh1***

Erh1 was identified as a suppressor of *sme2+* (the gene encoding meiRNA) phenotype (Yamashita et al., 2013). Erh1 belongs to the ERH family of proteins that is involved in regulation of nuclear processes. Further study showed that Erh1 and Mmi1 form a stoichiometric complex, termed EMC (Erh1-Mmi1-Complex). Erh1 was found to associate with Mmi1 in vegetative and meiotic cells (Sugiyama et al., 2016) and their localization is mutually dependent, Mmi1 foci was lost in *erh1* deleted cells and Erh1 foci was lost in *mmi1* deleted cells (Fig. 12). The EMC was found to be associated with MTREC complex (Egan et al., 2014; Zhou et al., 2015) as well as Ccr4-Not complex (Sugiyama et al., 2016), involved in RNA degradation and heterochromatin formation (Fig. 13a). Both Erh1 and Mmi1 are required for heterochromatin formation at chromosome loci of meiotic genes (Fig. 13b). Recent crystal structure of a truncated EMC complex shows that Erh1 and Mmi1 form a 2:2 heterotetramer. The Erh1 homodimer interacts with Mmi1 by a conserved interface with an unstructured amino terminal region of Mmi1 (Xie et al., 2019). They have proposed that Mmi1 binds to the DSR motif containing transcripts by its C-terminal YTH domain and recruits Erh1 and other RNA processing machinery by its N-terminal region. They have also found that the interface of Erh1 dimer, which is involved in interaction with Mmi1, is highly conserved. Although human ERH and Erh1 do not share high similarity on sequence level, the dimer interface is conserved in these two proteins.

#### ***1.6.1 Role of EMC in Heterochromatin island formation***

In addition to DSR-dependent mRNA degradation, Mmi1 suppresses transcription of meiotic genes by forming heterochromatin. Heterochromatin is a condensed state of nuclear DNA, which is transcriptionally inactive. Heterochromatin is present only in eukaryotes where it plays a role in chromatin regulation and gene expression. A characteristic of heterochromatin is hypoacetylation of histones and presence of methylated histones H3K9, H3K27 and H4K20. Depending on the factors involved in formation of heterochromatin, fission yeast has three types of heterochromatin. The first type includes heterochromatin formed at centromeres, telomeres and mating type region (Reyes-Turcu and Grewal, 2012). Formation of heterochromatin is mediated by RNAi. RNA interference (RNAi) machinery is comprised of Ago1 (argonaute), Dcr1 (Dicer) and Rdp1 (RNA-dependent RNA polymerase). From these regions, transcription of repeated elements (dg/dh repeats) produce dsRNA which is then chopped up by Dcr1 to produce siRNA (small interfering RNA). These siRNAs are fed to RNA-induced transcriptional silencing complex (RITS). RITS complex is then guided to these RNA repeats by the siRNA and recruits the methyltransferase Clr4 that methylates H3 in lysine 9 (H3K9) position, which is a hallmark of heterochromatin (Reyes-Turcu and Grewal, 2012).

The second type of heterochromatin includes small blocks of heterochromatin islands, including meiotic genes. These islands are dynamically regulated depending on environmental conditions (Zofall et al., 2012) (Fig 13a). RNAi is dispensable for formation of heterochromatin islands on meiotic genes but requires Red1 and Rrp6 (Hiriart et al., 2012; Tashiro et al., 2013; Zofall et al., 2012).

The third type, is known as HOODs (heterochromatin domains) that are dynamically regulated and appear when nuclear exosome is inactive. HOOD formation can be RNAi dependent or independent. Red1 along with Pla1 and Pab2 are involved in degradation of transcripts at these regions.

Regulation of meiotic transcripts by heterochromatin formation can be exerted by heterochromatin island formation at meiotic genes (*e.g.* *mei4+*, *ssm4+*) or by RNAi mediated heterochromatin domain formation at meiotic genes (*e.g.* *mug5+*, *mcp3+*) and genes regulated by environmental condition (*e.g.* *pho1+*). Involvement of Mmi1 in heterochromatin formation was suggested as deletion of DSR abolished HOOD formation in those loci and loss of Mmi1, Red1 and Rrp6, disrupted H3K9 formation at meiotic loci (Zofall et al., 2012).

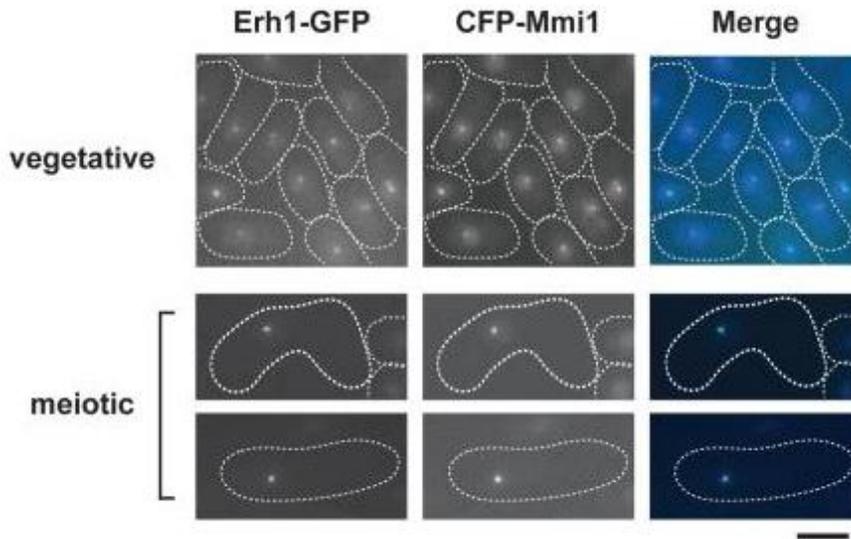


Fig.12: Erh1 co-localizes with Mmi1 in vegetatively growing cells as well as meiotic cells as shown by fluorescence microscopy (adapted from Sugiyama et al., 2016)

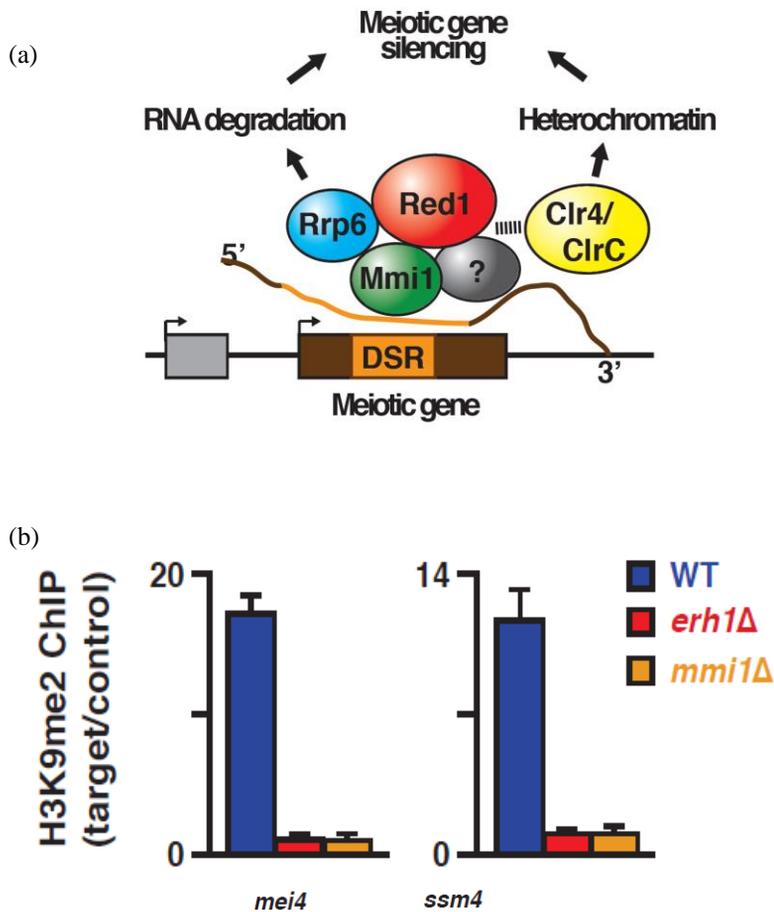


Fig.13: Mmi1 contributes to heterochromatin island formation and RNA decay (a) Model showing role of Mmi1 in facultative heterochromatin island formation at meiotic gene loci (adapted from Zofall et al., 2012). (b) ChIP-chip study indicates that Erh1-Mmi1-Complex is essential for heterochromatin formation at *mei4* and *ssm4* loci (adapted from Sugiyama et al., 2016)

H3K9 marks have been located on many meiotic genes in vegetatively growing cells which disappear with initiation of meiosis (Zofall et al., 2012).

In addition to the degradation of meiotic transcripts, a subset of these transcripts were found to be under an epigenetic regulation. For example, *mei4+* and *ssm4+* genes have low levels of Histone H3K9 methylation. Such heterochromatic islands formed at meiotic genes are mediated by Mmi1 and Red1 (Egan et al., 2014; Lee et al., 2013; Tashiro et al., 2013; Zofall et al., 2012). H3K9 methylation was lost in cells where either Erh1 or Mmi1 was deleted, suggesting EMC is essential for heterochromatin island formation (Sugiyama et al., 2016) (Fig 13b). The exact mechanism by which Mmi1 facilitates heterochromatin island formation at meiotic gene loci is still elusive. It has been proposed that Red1, one of the core component of MTREC complex recruits the only H3K9 methyltransferase of fission yeast, Clr4 which promotes heterochromatin formation at those regions. EMC associates with two complexes involved in degradation/repression of meiotic transcripts, MTREC and Ccr4-Not. EMC can assemble heterochromatin island at meiotic genes in a RNAi independent manner but Ccr4-Not is dispensable. Loss of Ccr4 has little effect on H3K9 methylation, contrary to the fact that loss of either Mmi1, Erh1 or MTREC completely abolishes H3K9 methylation at meiotic islands (Hiriart et al., 2012; Tashiro et al., 2013; Zofall et al., 2012; Sugiyama et al., 2016).

Further studies have shown that Ccr4-Not has a role in HOOD formation. HOOD is another kind of heterochromatin assembly that is triggered by environmental stimuli. Some HOOD formations are Mmi1-dependent and some are Erh1-dependent but loss of Ccr4 abolished H3K9 methylation from both EMC dependent and independent HOODs (Sugiyama et al., 2016). The study from Grewal's lab has discovered an evolutionarily conserved protein, Pir2/Ars2 which affects heterochromatin assembly at both EMC dependent and independent HOODs, in a similar way as Ccr4-Not. Pir2 associates with both Ccr4-Not and with MTREC. It has been proposed that Ccr4 and Pir2 work in co-operation with RNAi to form heterochromatin domains.

Therefore, Mmi1 silences the expression of meiotic genes in at least two different ways, through RNA degradation and nuclear retention. EMC and MTREC associate and cooperate with the Rrp6 subunit of the nuclear exosome for the selective elimination of meiotic DSR-containing transcripts (Harigaya et al., 2006; Houseley et al., 2006; Yamanaka et al., 2010). In addition, EMC sequesters meiotic mRNAs in nuclear foci, preventing their export to the

cytoplasm and their translation. Functional studies show that disruption of Mmi1-Erh1 complex causes defects in heterochromatin formation while the degradation of meiotic transcripts remains intact (Xie et al., 2019). EMC and MTREC are the key players of heterochromatin island formation and Ccr4-Not along with EMC coats meiotic loci by heterochromatin domain assembly.

### ***1.7 Regulation of Mmi1 in meiotic cells***

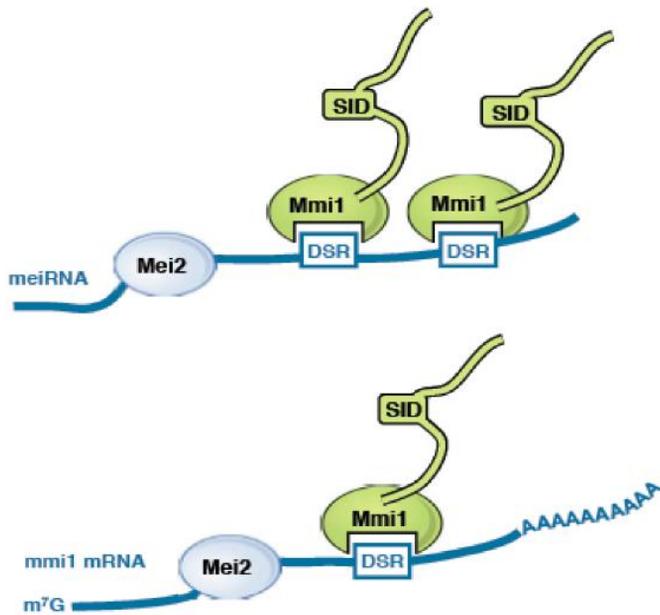
In vegetatively growing cells Mmi1 is scattered in one or multiple nuclear foci but upon starvation, these foci converge on one single dot. This dot overlaps the Mei2 dot. Further studies revealed that upon entry into meiosis, Mmi1 is sequestered in an RNP (ribonucleoprotein) complex formed by the Mei2 protein and the long noncoding meiRNA, thereby ensuring smooth progression of meiosis (Harigaya et al., 2006) (Fig. 14) and loss of either *mei2+* or *sme2+* disrupts Mmi1 foci formation (Fig. 15). This was the only known way of Mmi1-Mei2 dependent regulation of meiosis, until it was reported that Mmi1 recruits Ccr4-Not complex to promote ubiquitination and down-regulation of its own inhibitor, the meiosis inducer Mei2, via the Not4/Mot2 E3 ubiquitin ligase subunit (Fig. 16). This regulatory circuit preserves the activity of Mmi1, ensuring efficient meiotic mRNA degradation in mitotic cells.

Further investigation showed that Mei2 not only sequesters Mmi1 during meiosis but also inactivates Mmi1 at the mRNA level as it binds to its transcripts during early meiosis (Mukherjee et al., 2018). This interaction prevents transport of *mmi1+* transcripts to the cytoplasm for translation. This study has identified another target of Mei2, *rep2+* transcript, which encodes a protein, Rep2, a subunit of a transcription factor. Further studies are needed to clarify the relationships between these two main effectors of the mitosis-meiosis switch in *S. pombe*.

Erh1 was also found to co-localize with Mei2 dot in meiotic cells and lack of Erh1 affected Mei2 localization (Sugiyama et al., 2016). The exact nature of the relationship between Erh1 and Mei2 has to be explored (Fig.17; Fig. 18).

### ***1.8 Mmi1-the exceptional YTH domain***

The functional role of Mmi1 in regulating meiotic transcripts in vegetatively growing cells has been discussed in details. Mmi1 is comprised of a C-terminal YTH domain and N-terminal low complexity region (Fig. 19a). Structurally Mmi1-YTH domain has the



**REPRESSION OF Mmi1 SYNTHESIS AND FUNCTION**

Fig.14: Schematic representation of the regulation of Mmi1 level by Mei2 during meiosis (adapted from Hazra et al., 2019)

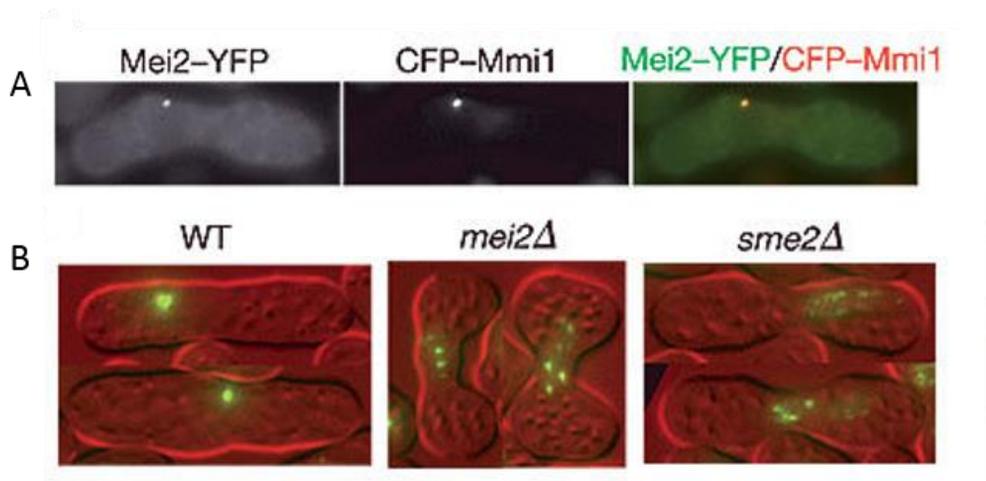


Fig.15: Mmi1 and Mei2 co-localize during meiosis (A) live cell microscopy of a cell in first meiotic prophase shows that Mmi1 co-localizes with Mei2 (B) Mmi1 foci is scattered in absence of Mei2 or meiRNA (adapted from Harigaya et al., 2006)

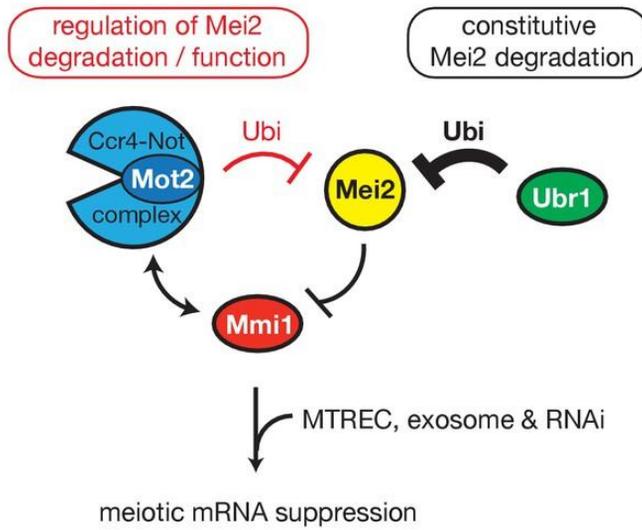


Fig.16: Mmi1 mediated downregulation of Mei2 (adapted from Simonetti et al., 2017)

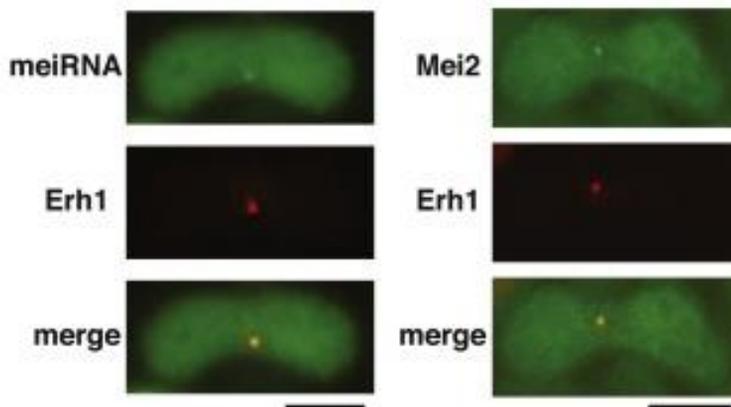


Fig.17: Erh1 co-localizes with Mei2 and meiRNA during meiosis (adapted from Sugiyama et al., 2016)

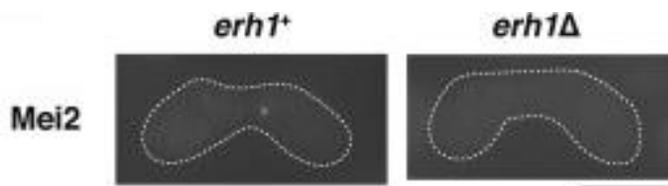


Fig.18: Mei2 focus disappears in absence of Erh1 (adapted from Sugiyama et al., 2016)

conserved YTH fold but its RNA binding mode is different from other members of the evolutionarily conserved YTH domain family. Mmi1 is an RNA binding protein that specifically recognizes a consensus U(U/C/G)AAAC motif, present in multiple copies within larger regions called Determinant of Selective Removal (DSR). This motif radically differs from the m<sup>6</sup>A consensus binding site recognized by canonical YTH domains. Although this domain is structurally similar to YTH domains from other proteins such as YTHDC1, YTHDF2, and Pho92, it is using a different surface to interact with its RNA consensus motif (Fig. 19b). This binding surface is located on the opposite face of the Mmi1 YTH domain compared to the region involved in m<sup>6</sup>A binding by classical YTH domains (Wang et al., 2015), (Wu et al., 2017). This region is conserved in fission yeasts (*S. pombe*, *Schizosaccharomyces japonicus*, *Schizosaccharomyces octosporus*, and *Schizosaccharomyces cryophilus*) but not in other YTH containing proteins. Mmi1 YTH domain is incapable of binding a GGm<sup>6</sup>AC containing RNA, and N6-methylation at any position of the DSR motif weakens binding to the RNA. Similarly, the YTH domains from YTHDC1, YTHDF2 and Pho92 do not bind DSR consensus motif. Detailed comparison of the Mmi1 pocket corresponding to m<sup>6</sup>A binding site in other YTH domain proteins shows that two Trp residues are conserved while the position corresponding to Trp, Leu, or Tyr in canonical YTH domains is occupied by a His in Mmi1. However, two main differences that could explain the Mmi1 YTH domain inability to bind m<sup>6</sup>A containing RNAs have been observed. First, the amino acid corresponding to the Asn, Asp, or His residues forming a hydrogen bond with N1 atom of the adenosine ring is substituted by Ala in Mmi1. Second, while the surface surrounding the m<sup>6</sup>A binding pocket is positively charged in canonical YTH domains, the corresponding region in Mmi1 is negatively charged, which is not favourable for RNA binding (Wang et al., 2015) (Fig. 20). Hence, Mmi1 is an exception among YTH domains as it cannot recognize m<sup>6</sup>A modification and interacts with RNA in a completely different manner.

## ***1.9 Comparative study of human YTH domains***

This section is strongly inspired by the review that I have published together with Dr Clément Chapat and Dr Marc Graille in the journal *Genes* (Hazra et al., 2019). The original version of this review is presented as Annex 1.

### ***1.9.1 Discovery of m<sup>6</sup>A modification and YTH domain***

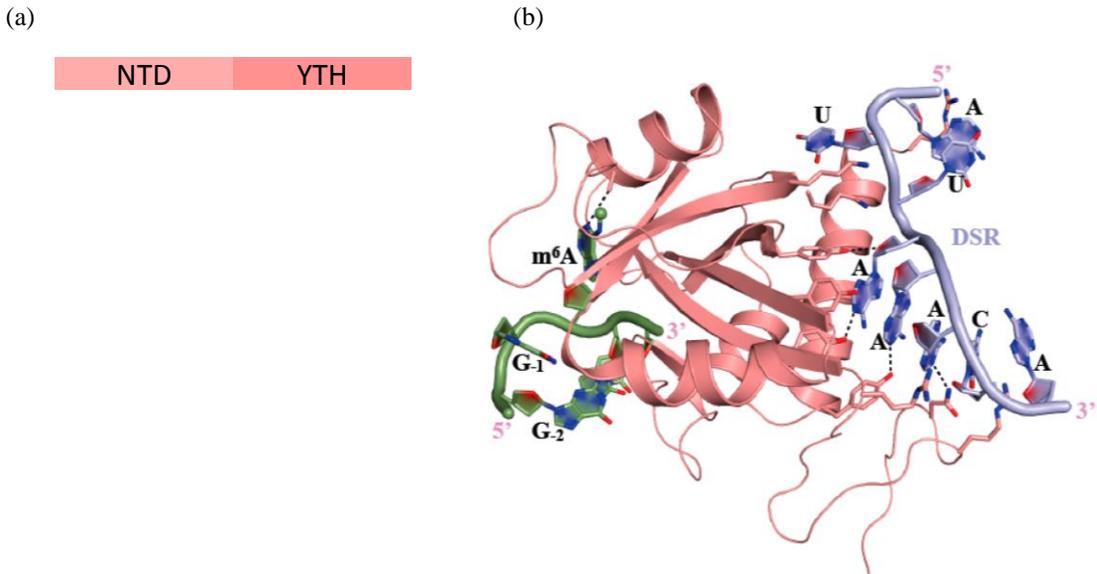


Fig.19: Mmi1-YTH domain interacts with RNA in an atypical manner (a) schematic representation of domain organization of Mmi1 (b)Mmi1 contains an atypical YTH domain. Ribbon representation of the complex between *S. pombe* Mmi1 YTH domain (pink) and a Determinant of Selective Removal (DSR) RNA sequence (adapted from Hazra et al., 2019). An m6A containing RNA fragment (green) has been modeled by superimposing the crystal structure of RNA-bound YTH domain from YTHDC1 onto the structure of *S. pombe* Mmi1. Some residues and hydrogen bonds important for the interaction between *S. pombe* Mmi1 and the RNA DSR sequence are shown as sticks or black dashed lines, respectively (adapted from Hazra et al., 2019)

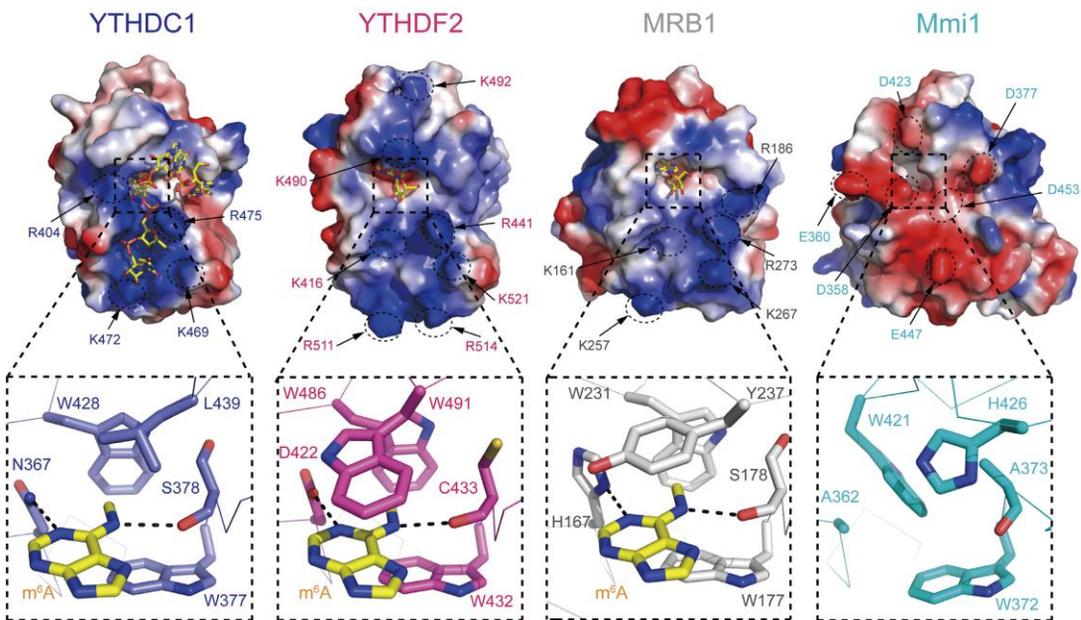


Fig.20: Comparative study of Mmi1 with other YTH domain proteins. The upper panel shows surface electrostatic potential and in in-set the position of the aromatic residues surrounding grooves of YTH domain are shown (adapted from Wang et al., 2015)

The regulation of gene expression plays a central role during development and upon cell response to stress exposure. Hence, living organisms have developed highly complex mechanisms at different steps of gene expression to tune various cellular pathways. Many of those events occur at the post-transcriptional level through the formation of protein-RNA complexes that will influence various aspects of messenger RNA (mRNA) maturation such as alternative splicing, editing, export, and polyadenylation. Therefore, RNA binding domains (RBDs) are key actors in these regulatory mechanisms through their recognition of specific RNA sequences or structures. The most common RBDs are the RNA recognition motif (RRM) (corresponding to about 2% of all RNA binding domains), the hnRNP K Homology domains (KH), Piwi, Argonaute and Zwillig domains (PAZ), and double stranded RNA-binding domains (dsRBD) (Lingel et al., 2005; Daubner et al., 2013; Masliah et al., 2013; Nicastro et al., 2013). With recent technological developments combining UV cross-linking together with oligo(dT) purification of mRNAs, we now have comprehensive lists of RNA-binding proteins in various eukaryotic organisms (Baltz et al., 2012; Castello et al., 2012; Mitchell et al., 2013; Castello et al., 2013), indicating that many other RBDs are yet to be uncovered.

The recent identification of several internal post-transcriptional modifications such as m<sup>6</sup>A (N<sup>6</sup>-methyladenosine), m<sup>1</sup>A (N<sup>1</sup>-methyladenosine), pseudouridine ( $\Psi$ ), m<sup>5</sup>C (5-methylcytosine), and ac<sup>4</sup>C (N<sup>4</sup>-acetylcytosine) within mRNAs has shed light onto an additional layer of regulation now known as epitranscriptomics (Desrosiers et al., 1974; Dominissini et al., 2012; Meyer et al., 2012; Squires et al., 2012; Schwartz et al., 2013; Carlile et al., 2014; Schwartz et al., 2014; Dominissini et al., 2014; Safra et al., 2017; Arango et al., 2018). Indeed, similar to the dynamic modifications known to occur in DNA and proteins, epitranscriptomics modifications widely contribute to the regulation of biological pathways (Gilbert et al., 2016; Peer et al., 2017). At present, the most studied mRNA modification is m<sup>6</sup>A. This modification was initially identified in mRNAs four decades ago (Desrosiers et al., 1974). However, because m<sup>6</sup>A does not alter base pairing (Roost et al., 2015), and hence does not result in the introduction of a stall or a mutation during reverse-transcription, the development of advanced techniques to precisely map m<sup>6</sup>A sites has been a main obstacle for studying its biological significance. This field was reignited in 2011, following the discovery of the m<sup>6</sup>A demethylase FTO (fat mass and obesity-associated protein), a protein involved in human obesity, as a so-called ‘eraser’ enzyme that removes m<sup>6</sup>As present on mRNAs (Jia et al., 2011). More recently, several laboratories have finally

been successful in mapping m<sup>6</sup>A at individual-nucleotide resolution using cross-linking and immuno-precipitation with m<sup>6</sup>A-specific antibodies. High-throughput sequencing of the immuno-precipitated RNA fragments revealed the presence of more than 10,000 m<sup>6</sup>A sites in human cells, affecting more than 25% of the transcriptome. Their detailed mapping showed an enrichment of m<sup>6</sup>A near the stop codon and in the 3' untranslated region (3' UTR) of the target mRNAs. In human cells, the main 'writer' methyltransferase is a multi-protein complex composed of at least METTL3, METTL14, WTAP, and KIAA1429, which is responsible for m<sup>6</sup>A deposition on the consensus motif DRA\*CH (where D is A, G or U; R is A, or G; A\* is the methylated A and H is A, C, or U; (Dominissini et al., 2012; Schwartz et al., 2014; Liu et al., 2014; Ping et al., 2014). The m<sup>6</sup>A marks can be deleted by 'erasers' such as FTO and ALKBH5 (Jia et al., 2011; Zheng et al., 2013). Furthermore, this modification can attract m<sup>6</sup>A-binding proteins known as 'readers', as well as repel various proteins regulating mRNA functions (Dominissini et al., 2012; Wang et al., 2014; Liu et al., 2015; Edupuganti et al., 2017). This differential recruitment of regulatory proteins on m<sup>6</sup>A marks subsequently determines the fate of m<sup>6</sup>A-containing mRNAs, such as splicing, translation, degradation, or cellular localization. The most studied RNA binding module known to directly recognize m<sup>6</sup>A marks is the YTH domain. Indeed, many studies are scrutinizing eukaryotic YTH-containing proteins in order to clarify their roles in the regulation of mRNA fates. The present section aims at summarizing our current knowledge on this class of m<sup>6</sup>A readers.

### **1.9.2 The YTH Domain, an m<sup>6</sup>A RNA Grip**

The foremost member of this m<sup>6</sup>A reader protein family is human YT521-B (hereafter termed YTHDC1), which was initially identified as a factor interacting with Tra2 $\beta$ , SC35, SF2, hnRNP G, and SAM68 splicing factors (Imai et al., 1998; Hartmann et al., 1999). Despite this clear interaction of YTHDC1 with splicing machinery, no known RNA binding domain was detected and the only common feature with splicing factors was the presence of repeats of charged amino acids (Hartmann et al., 1999). Further bioinformatics analyses led to the identification of an additional conserved region, named YTH (for YT521-B Homology) domain. This domain was found exclusively in eukaryotic proteins from fungi (one member in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* yeasts) through plants (13 members in *Arabidopsis thaliana*) to higher eukaryotes (five members in human, namely YTHDC 1–2 and YTHDF 1–3) (Stoilov et al., 2002; Bhat et al., 2018). Although this domain was predicted to be a putative RNA binding domain since its identification, its RNA binding

property was only demonstrated in 2010 when Zhang et al. showed that YTHDC1 can bind degenerate RNA sequences (Zhang et al., 2010). YTH-containing proteins became a topic of strong interest when YTHDF1, 2, and 3, three paralogous members of the YTH-containing protein family, were found to be the most enriched human proteins specifically retained by an RNA fragment containing an m<sup>6</sup>A modification at the heart of the DRA\*CH consensus sequence (Dominissini et al., 2012; Wang et al., 2014).

The YTH domain is made of 150 to 200 residues and adopts an  $\alpha/\beta$  fold, with four to five  $\alpha$ -helices surrounding a curved six stranded  $\beta$ -sheet as revealed by the NMR structure of YTHDC1 YTH domain determined by the RIKEN Structural Genomics and Proteomics Initiative in 2007 (PDB code: 2YUD). At the centre of the  $\beta$ -sheet lies a cavity delineated by conserved hydrophobic residues including three tryptophan amino acid side chains (in some YTH domains, one Trp residue is substituted by either Leu or Tyr) forming a so-called aromatic cage (Fig. 21). Several 3D-structures of different YTH-m<sup>6</sup>A containing RNA complexes have revealed that this aromatic cage is responsible for the specific recognition of the m<sup>6</sup>A mark by YTH domains (Li et al., 2014; Luo and Tong, 2014; Theler et al., 2014; Xu et al., 2014; Xu et al., 2015). Indeed, in the case of human YTHDC1 bound to an m<sup>6</sup>A containing RNA, the aromatic cage specifically accommodates m<sup>6</sup>A via the side chains from W377, W428, and L439 amino acids (Xu et al., 2014). The m<sup>6</sup>A methyl group forms methyl- $\pi$  interaction with W428 side chain while the purine base is sandwiched between the W377 and L439 side chains (Fig. 21). This interaction of the m<sup>6</sup>A methyl group with W428 most probably explains the higher affinity (20 to 50-fold difference) of YTH domains for an m<sup>6</sup>A containing RNA oligonucleotide compared to the same unmodified oligonucleotide (Theler et al., 2014). Sequence alignment revealed that the W377 and W428 residues are strictly conserved among the YTH-containing proteins while the position corresponding to L439 can be occupied by either Leu, Tyr, or Trp. The integrity of this aromatic cage is essential for the recognition of m<sup>6</sup>A since mutation of any of these three residues to alanine disrupts binding (Li et al., 2014; Xu et al., 2014; Xu et al., 2015). Polar amino acids (S378, N363, and N367 in human YTHDC1) also contribute to m<sup>6</sup>A recognition by forming specific hydrogen bonds with nitrogen atoms from the adenine base and then participating together with the aromatic cage to the selective recognition of m<sup>6</sup>A (Fig. 21). Residues surrounding this aromatic cage are also important for RNA binding as they form a large positively-charged surface interacting with RNA phosphate groups from nucleotides surrounding m<sup>6</sup>A and also provide some specific interactions with bases. A large majority of m<sup>6</sup>A marks generated by the main

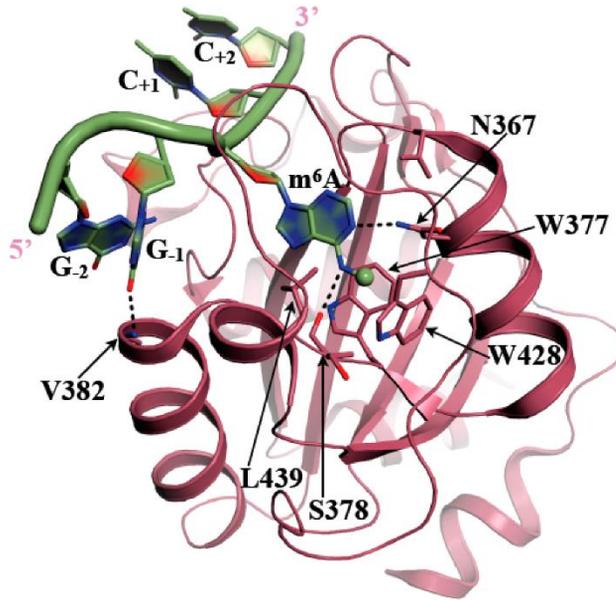


Fig.21: Ribbon representation of a GGm<sup>6</sup>ACC RNA oligonucleotide (green) bound to human YTHDC1 (PDB code: 4R3I). The methyl group grafted on N6-adenosine is shown as a sphere. The side chains from residues involved in the formation of the m<sup>6</sup>A aromatic cage and the hydrogen bonds responsible for specificity of m<sup>6</sup>A as well as for increased affinity of YTHDC1 for RNAs harboring a G just upstream of the m<sup>6</sup>A mark are shown as sticks. Hydrogen bonds are depicted by black dashed lines (adapted from Hazra et al., 2019).

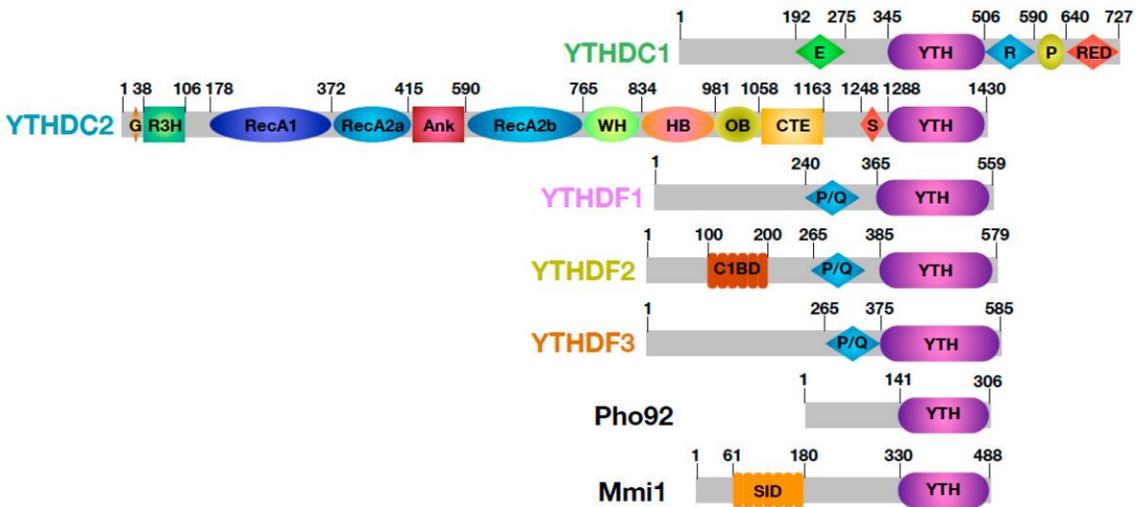


Fig.22: Schematic representation and domain composition of human, *Saccharomyces cerevisiae* (Pho92) and *Schizosaccharomyces pombe* (Mmi1) YTH-domain containing proteins. The predicted or experimentally determined limits of domains are indicated. E: Glu-rich domain. R: Arg-rich domain. P: Pro-rich domain. RED: Arg/Glu/Asp-rich domain. G: Gly-rich domain. R3H: small domain containing an invariant Arg sported from a highly conserved His by three residues. RecA1 and RecA2: RecA domains found in helicases. Ank: Ankyrin repeats. WH: Winged-helix domain. HB: Helical bundle. OB: OB-fold. CTE: C-terminal extension. S: Ser-rich domain. P/Q: Pro and Gln-rich domain. C1BD: CNOT1 binding domain. SID: Self-Interacting domain. (adapted from Hazra et al., 2019)

m<sup>6</sup>A methyltransferase ‘writer’ is found within Gm<sup>6</sup>AC (70%) or Am<sup>6</sup>AC (30%) motifs (Xu et al., 2014). Interestingly, structural analyses coupled to isothermal titration calorimetry (ITC) measurements using various RNA sequences have shown that the affinity of YTH domain from YTHDC1 for RNA fragments is higher (five- to six-fold difference) when the nucleotide immediately upstream of m<sup>6</sup>A is a G compared to an A (Xu et al., 2014; Xu et al., 2015), suggesting the co-evolution of both the active site of METTL3-METTL14 methyltransferase holoenzyme and of the RNA binding site of human YTHDC1. Indeed, in the structure of YTHDC1 bound to a GGm<sup>6</sup>ACU RNA fragment (Xu et al., 2014), the carbonyl group at position six of the G base preceding m<sup>6</sup>A forms a hydrogen bond with the main chain nitrogen group from V382 (or equivalent position in other YTH domains; Fig. 21). The amine group, present at the same position in A, is less prone to form such hydrogen bond, which could rationalize the differences in measured affinities. However, this seems to be specific of YTHDC1 as all other tested YTH-containing proteins (YTHDF1/2, YTHDC2, and *S. cerevisiae* Pho92) exhibit similar affinities for RNAs containing any of the four nucleotides at the position immediately upstream m<sup>6</sup>A (Xu et al., 2015). This probably results from the fact that this nucleotide binds into different pockets with no obvious base specificity on YTHDF1 or the *Zygosaccharomyces rouxii* Mrb1/Pho92 fungal orthologue compared to YTHDC1 (Luo and Tong, 2014; Xu et al., 2015). So far, structural studies have been performed only with 5- or 7-mers RNAs, which are known to bind more weakly (K<sub>d</sub> values in the 2 to 30 μM range) than longer RNAs (9 or 16-mers for instance; K<sub>d</sub> values ranging from 0.2 to 0.3 μM for YTHDC1 up to 7.5 μM for YTHDC2 (Xu et al., 2015). Therefore, additional structural studies with longer RNA fragments might be of interest to bring more information on RNA recognition by these YTH motifs.

In summary, YTH-domain containing proteins interact with single-stranded RNAs and selectively identify the presence of a modified m<sup>6</sup>A nucleotide at the centre of a consensus signature motif matching that of the major m<sup>6</sup>A methyltransferase machinery identified to date.

### ***1.9.3 YTH, a Building Block Governing the Fates of m<sup>6</sup>A Containing mRNAs***

Most RNA binding modules are embedded within larger proteins and are surrounded by either structured domains or low complexity regions, all with various functions (Castello et al., 2016; Lunde et al., 2007; Achsel and Bagni, 2016). The YTH domain is no exception to this rule and clearly serves as a building block mostly flanked by regions predicted to be

disordered (Fig. 22;(Wang et al., 2015; Du et al., 2016; Stowell et al., 2016)). By influencing both the sub-cellular localization of these YTH proteins and their partners, these flanking regions are thereby important for the function of YTH proteins in the regulation of m<sup>6</sup>A-containing RNA fates, i.e., splicing, mRNA nucleocytoplasmic export, translation and mRNA decay (Wang et al., 2014; Luo and Tong 2014; Xiao et al., 2016; Li et al., 2017; Roundtree et al., 2017).

As mentioned above, bioinformatics analyses have identified 13 YTH-containing proteins in *A. thaliana*, which roles are being clarified. On the basis of amino acid sequences, YTH domain proteins from different vertebrates can be divided into three families: YTHDC1, YTHDC2, and YTHDF1–3. There is no significant sequence similarity between members from those different families outside of the YTH domain, hence YTHDC1, YTHDC2, and YTHDF1–3 members cannot be considered as paralogues.

### ***1.9.3.1 YTHDC1***

YTHDC1 (previously known as YT521B) is the founding member of YTH family of proteins. In this protein, the YTH domain is the only region predicted to be folded and is surrounded by regions rich in charged residues (Glu-rich, Arg-rich and Arg-Asp-Glu-rich, or RED segments; Figure 22) or in proline (P-rich). The human protein contains nuclear localization elements and is found to localize in distinct sub-nuclear bodies, so-called YT bodies, adjacent to nuclear splicing speckles (Nayler et al., 2000). This nuclear localization is in agreement with its initially described interactions with various splicing factors (i.e., Tra2 $\beta$ , SC35, SF2, hnRNP G, and SAM68 (Imai et al., 1998; Hartmann et al., 1999). Recent studies have clearly established a role of YTHDC1 in splicing both in human cells and in *Drosophila melanogaster* (Xiao et al., 2016; Lence et al., 2016). Indeed, this protein contributes to alternative splicing by binding to the pre-mRNAs and by influencing the splice site selection (Zhang et al., 2010). Mechanistically, YTHDC1 directly interacts with SRSF3 and SRSF10, two serine/arginine-rich splicing factors, in a competitive manner (Fig. 23). In doing so, it enhances the binding of SRSF3 to targeted pre-mRNAs resulting in exon inclusion while precluding the binding of SRSF10, which is involved in exon skipping. This YTHDC1-mediated recruitment of SRSF3 is clearly dependent on m<sup>6</sup>A as either METTL3 silencing or a YTHDC1 double mutant (two Trp residues from the aromatic cage of the YTH domain are substituted by Ala) strongly reduce binding of SRSF3 to RNAs in cellulo (Xiao et al., 2016). As YTHDC1 has a 30-fold higher affinity for SRSF3 over SRSF10, it interacts

predominantly with SRSF3 and hence may favour alternative splicing of m<sup>6</sup>A-containing pre-mRNAs. This interaction between YTHDC1 and SRSF3 is not only involved in alternative splicing but also in the polyadenylation process of the pre-mRNAs through their association with the pre-mRNA 3' end processing factors CPSF6 (Kasowitz, et al., 2018). Moreover, YTHDF1 and SRSF3 collaborate with NXF1 to drive efficient export of transcripts subject to m<sup>6</sup>A control from the nucleus to the cytoplasm (Fig. 23; (Roundtree et al., 2017)).

YTHDC1 has also been shown to participate in the regulation of the abundance of MAT2A mRNA, which encodes a subunit of one of the methionine adenosyltransferases responsible for the synthesis of S-adenosyl-L-methionine (SAM) cofactor from methionine and ATP (Shima et al., 2017). The MAT2A mRNA is one of the well-characterized targets of the recently identified METTL16 m<sup>6</sup>A RNA methyltransferase, which introduces m<sup>6</sup>A marks in the 3' UTR of this mRNA (Pendleton et al., 2017), a modification crucial for mouse embryonic development (Mendel et al., 2018). Depletion of SAM was shown to enhance the removal of a retained intron in MAT2A pre-mRNA leading to induced expression of this mRNA (Pendleton et al., 2017) and to reduce m<sup>6</sup>A levels in MAT2A 3' UTR (Shima et al., 2017). Whether the role of YTHDC1 in this pathway is to favour splicing, nucleocytoplasmic transfer, or recruitment of mRNA decay enzymes remains to be clarified.

In fruit-fly, the YTHDC1 orthologue is also involved together with m<sup>6</sup>A methylation in the alternative splicing of Sex lethal (Sxl), which encodes a master regulator of sex determination and dosage compensation (Lence et al., 2016). Although Sxl is expressed in males and females, the presence of an additional internal exon in males introduces a premature stop codon that results in both the production of a truncated and non-functional Sxl protein and the rapid elimination of the transcript most probably by the nonsense-mediated mRNA decay pathway (Salz et al., 1989; Moschall et al., 2017). Inactivation of either the METTL3 subunit of the m<sup>6</sup>A mRNA methyltransferase holoenzyme or the YTHDC1 protein in the female but not the male fly results in retention of the male-specific exon concomitant with the decrease of the female-specific isoform, clearly indicating a female-specific splicing defect linked to altered m<sup>6</sup>A deposition and recognition.

Beyond its role in splicing, human YTHDC1 also helps in transcriptional repression of X chromosome genes by X-inactive specific transcript (XIST), a long non-coding RNA that plays a critical role in inactivation of one X chromosome in female cells (Patil et al., 2016).

XIST is a heavily methylated RNA with at least 78 m<sup>6</sup>A sites and the preferential binding of YTHDC1 to m<sup>6</sup>A marks is necessary for XIST-mediated transcriptional silencing. The depletion of m<sup>6</sup>A ‘writer’ causes inhibition of XIST function and this defect can be restored by artificially tethering YTHDC1 to XIST in cells lacking m<sup>6</sup>A methylation machinery. A comprehensive understanding of the mode of action of YTHDC1 on XIST will necessitate further studies.

### **1.9.3.2 YTHDC2**

Compared to the other YTH domain-containing proteins, where the YTH domain is embedded within low complexity regions, members of the YTHDC2 family are multi-domain proteins (Figure 22). Apart from the C-terminal YTH domain, there is a N-terminal R3H (arginine and histidine-rich) domain with RNA-binding property (Kretschmer et al., 2018) preceded by a Gly-rich patch, a central DEAH-box helicase domain (where an ankyrin repeat domain is inserted in the middle of the second RecA domain), an helicase associated 2 domain (HA2), an OB-fold (oligonucleotide / oligosaccharide-binding fold) and a C-terminal extension (CTE) also found in human DHX36, a DNA/RNA DEAH-box helicase involved in G-quadruplex unwinding (Figure 22; (Chen et al., 2018)). In agreement with its domain composition, human YTHDC2 has RNA dependent ATPase and 3'→5' RNA helicase activities (Morohashi et al., 2011; Wojtas et al., 2011; Jain et al., 2018). This protein is mainly a diffuse cytoplasmic protein, but it is also enriched in peri-nuclear regions (Kretschmer et al., 2018). As expected from the presence of several RNA binding domains, YTHDC2 interacts with mRNAs and in particular with m<sup>6</sup>A-rich mRNAs through its YTH domain (Wojtas et al., 2017; Hsu et al., 2017).

Initially, YTHDC2 was shown to associate with hepatitis C virus protein NS5B to facilitate viral DNA replication (Morohashiet al., 2011) and to play an important role in the proliferation of cancer cells by enhancing the translation of metastasis-related genes (Tanabe et al., 2014; Tanabe et al., 2016). More recent studies have converged towards an important role of YTHDC2 in the progression of meiotic prophase I, which is a critical and long meiosis stage characterized by many chromosomal events that will ultimately lead to severing of the genome into two halves (Wojtas et al., 2017; Soh et al., 2017). Consequently, the inactivation of YTHDC2 gene in mice results in gametogenesis defects and infertility (Wojtas et al., 2017; Jain et al., 2018; Hsuet al., 2017; Bailey et al., 2017).

In human cells, YTHDC2 interacts in an RNA-independent manner with the meiosis-specific MEIOC protein as well as with the 5'→3' exonuclease XRN1 (Kretschmer et al., 2018; Wojtas et al., 2017; Hsu et al., 2017; Soh et al., 2017; Abby et al., 2016). Whether YTHDC2 can interact simultaneously with both MEIOC and XRN1 is unclear but it is tempting to speculate that the MEIOC–YTHDC2 complex interacts with m<sup>6</sup>A-enriched mRNAs to address them to degradation by the XRN1 exonuclease. However, this model could be restricted to a subclass of mRNAs as various studies have observed different effects on mRNA translation and stability upon inactivation of YTHDC2 gene. Indeed, YTHDC2–MEIOC complex could stabilize meiosis-specific transcripts (Abby et al., 2016) while destabilizing mitotic mRNAs (Tanabe et al., 2016; Wojtas et al., 2017; Hsu et al., 2017).

Conversely, YTHDC2 has also been shown to enhance translation efficiency of mRNAs concomitantly to decrease their stability (Fig. 23; Hsu et al., 2017). This could result from the association of YTHDC2 with the head of the 40S ribosomal subunit both at the level of the 40S but also of the 80S (Kretschmer et al., 2018). More precisely, YTHDC2 binding site on the 40S subunit maps in the vicinity of the mRNA entry and exit sites, which could rationalize the dual role of YTHDC2 in enhancing translation efficiency by recruiting m<sup>6</sup>A-containing mRNAs to the ribosome but also decreasing mRNA stability by recruiting XRN1 to those mRNAs (Fig. 23).

### ***1.9.3.3 YTHDF Family***

Human YTHDF1–3 are cytoplasmic proteins made of a single C-terminal YTH-domain that binds to m<sup>6</sup>A marks separated from a N-terminal low-complexity domain by segment rich in Pro, Gln, and Asn amino acids (Fig. 22; Wang et al., 2014; Li et al., 2014; Xu et al., 2015). Those three proteins are highly homologous with 65 to 68% and about 85% of sequence identity and similarity, respectively.

YTHDF2 was the first protein to be functionally characterized, especially regarding the repertoire of mRNAs that it binds to, as well as its mode of action. Photo-activable ribonucleoside crosslinking and immunoprecipitation (PAR-CLIP) combined with RNA-immunoprecipitation coupled to sequencing (RIP-Seq) experiments showed that human YTHDF2 selectively targets more than 3000 different transcripts and binds predominantly to their 3' UTR and around the stop codon (Wang et al., 2014). Furthermore, YTHDF2 knock-down results in the accumulation of its targets in translatable or actively translating polysome pools, pointing to a crucial role of YTHDF2 in the translation repression of its targets. This

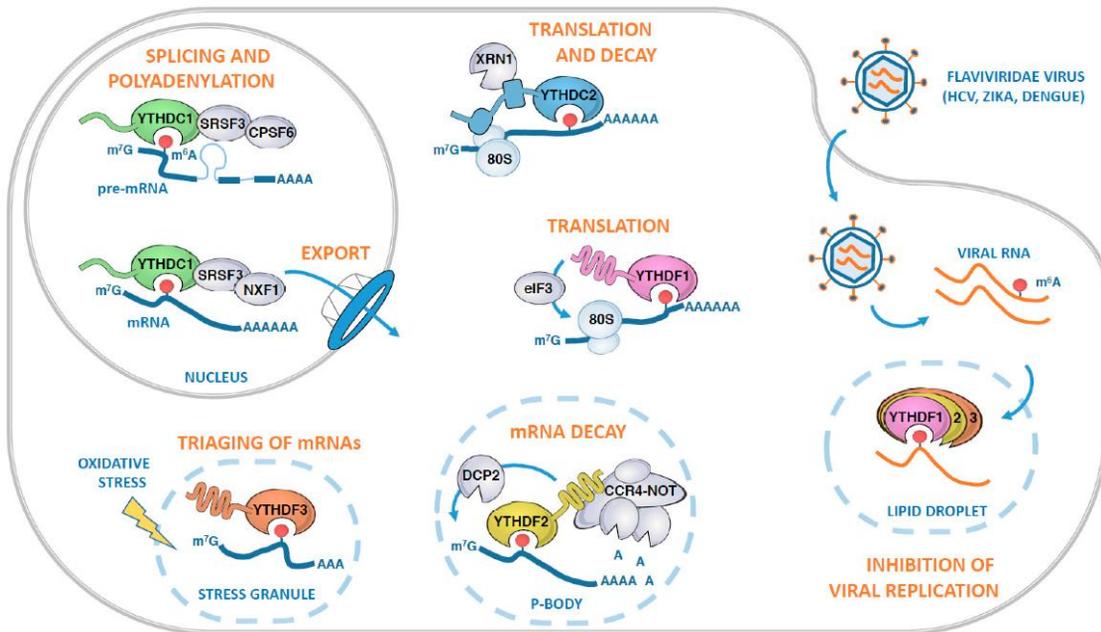


Fig.23: Roles of human YTH-containing proteins in various aspects of messenger RNA (mRNA) fates. In the nucleus, YTHDC1 recognizes m<sup>6</sup>A-modified pre-mRNAs and orchestrates their splicing, polyadenylation, and nuclear export through its association with SRSF3, CPSF6, and NXF1. Once in the cytosol, the modified mRNAs can be bound by YTHDC2, which in turn recruits both the ribosome and the XRN1 exoribonuclease. Alternatively, the mRNA can be targeted by the YTHDF proteins, either to be actively translated in an YTHDF1-dependent manner, or subjected to mRNA decay through YTHDF2 and its ability to recruit the CCR4-NOT deadenylase complex. YTHDF2-mediated mRNA decay is likely to occur in Processing bodies (P-bodies) where it co-localizes with the decapping enzyme DCP2. Following oxidative stress, m<sup>6</sup>A-modified mRNAs can also be recognized by YTHDF3, which facilitates the triaging of mRNAs into the stress granules. The YTHDF proteins can also target viral m<sup>6</sup>A-modified RNAs during infection, as illustrated by the case of hepatitis C virus (HCV) infection. During infection, the YTHDF proteins relocate to lipid droplets, sites of viral assembly, and sequester m<sup>6</sup>A-modified HCV RNAs, preventing their interaction with HCV core protein and subsequent virus particle production. (adapted from Hazra et al., 2019)

defect in translation is also accompanied by an increase in the global abundance of m<sup>6</sup>A-modified mRNAs, confirming the intimate link existing between the number of m<sup>6</sup>A sites and the instability of the targeted mRNA (Wang et al., 2014). This activity of YTHDF2 in mRNA destabilization requires both its N- and C-terminal regions as over-expression of full-length YTHDF2 leads to decay of m<sup>6</sup>A containing mRNAs, while expression of only the N-terminal or the C-terminal region does not have the same effect (Wang et al., 2014). This role of YTHDF2 in the degradation of m<sup>6</sup>A-containing mRNAs is further supported by its localization in Processing bodies (P-bodies) in which YTHDF2 co-localizes with DCP1a and DDX6 proteins known to be involved in mRNA decapping (Wang et al., 2014). YTHDF2 also directly interacts with CNOT1, the scaffolding subunit of the Ccr4-NOT mRNA deadenylase (Du et al., 2016). This interaction relies on the SH domain from CNOT1 and the YTHDF2 N-terminal domain, which is also responsible for the localization of at least YTHDF2 to P-bodies (Fig.23; Du et al., 2016). Interestingly, Pho92, the only YTH-containing protein from *S. cerevisiae*, also interacts with Pop2, another component of the Ccr4-NOT complex (Kang et al., 2014). The similarities between human YTHDF2 and *S. cerevisiae* Pho92 are emphasized by (1) the ability of human YTHDF2 gene (but not YTHDC1) to complement for the deletion of PHO92 gene in *S. cerevisiae* and (2) the role of Pho92 as an enhancer of mRNA decay (Kang et al., 2014). Altogether, this suggests that the main role of YTHDF2 is in the regulation of m<sup>6</sup>A-containing mRNA decay and that this role has been conserved throughout evolution. Interestingly, upon heat shock, YTHDF2 relocates to the nucleus and this is accompanied by a specific increase of m<sup>6</sup>A in the 5' UTR of stress-inducible mRNAs and an increased ribosome occupancy in their coding region (Zhou et al., 2015). This could then contribute to the stimulation of translation by the direct recruitment of the translation initiation factor 3 complex (eIF3) to m<sup>6</sup>A sites located within mRNA 5' UTRs (Meyer et al., 2015).

Despite the strong sequence similarity with YTHDF2, YTHDF1 knock-down does not affect the m<sup>6</sup>A/A ratio, indicating that this protein is unlikely to be involved in m<sup>6</sup>A-containing mRNA decay (Wang et al., 2015). On the contrary, YTHDF1 seems to enhance the translation efficiency of a population of transcripts encoded by around 1200 genes, to which it associates in an m<sup>6</sup>A-dependent manner. This mechanism is likely to occur through the recognition of m<sup>6</sup>A sites by YTHDF1 within the 3' UTR of mRNAs on the one hand, as well as with the 40S subunit and components of the eIF3 complex bound in the vicinity of the start codon on the other (Fig.23; Wang et al., 2015). This mechanism differs from the one

described above for 5' UTR m<sup>6</sup>As that directly recruit eIF3 under stress conditions and that is independent of YTHDF1 (Meyer et al., 2015). Hence, the effect of YTHDF1 on translation of m<sup>6</sup>A-mRNAs may be limited to a small subset of RNAs depending on various physiological situations. In the nervous system for example, transcriptome-wide mapping of YTHDF1-binding sites, combined with nascent protein labelling, revealed that YTHDF1 enhances translation of key hippocampal m<sup>6</sup>A-methylated mRNAs in response to neuronal stimulation, thus contributing to learning and memory (Shi et al., 2017).

Several observations indicate that YTHDF3 interacts with both YTHDF1 and YTHDF2 proteins and thereby works together with those factors to up-regulate translation (YTHDF1) or enhance degradation of mRNAs (YTHDF2), respectively (Li et al., 2017; Shiet al., 2017). Indeed, PAR-CLIP coupled to RIP-seq showed that YTHDF3 shares more than half of its targets with YTHDF1 but also YTHDF2, which is not surprising considering the high degree of sequence identity (86% to 89%) between the YTH domains from these three proteins (Li et al., 2017; Shi et al., 2017). However, human YTHDF3 might be also mobilized independently of YTHDF1 and YTHDF2 under specific conditions. Notably YTHDF3 co-localizes exclusively with the stress granules under oxidative stress, whereas YTHDF1 and YTHDF2 retain their cytoplasmic localization with only marginal presence in stress granules. In this specific context, YTHDF3 selectively recognizes a pool of oxidative stress-induced methylated mRNAs in order to mediate triaging of mRNAs from the translatable pool to stress granules (Fig. 23; Anders et al., 2018). Overall, this indicates that all three YTHDF family proteins may participate in a complex regulatory mechanism that results first in a higher translational efficiency of m<sup>6</sup>A-mRNAs followed by their rapid degradation. This complex interplay between these three YTHDF proteins and interacting proteins involved in these different cellular processes will need to be clarified in the future.

Interestingly, m<sup>6</sup>A marks are not restricted to cellular RNAs but are also found in viral mRNAs, where they are recognized by the YTHDF1–3 proteins (Krug et al., 1976; Gokhale et al., 2016; Kennedy et al., 2016; Lichinchi et al., 2016; Tirumuru et al., 2016). Several recent studies have focused on the roles of m<sup>6</sup>A and YTHDFs on the regulation of viral infection leading to the description of various mechanisms. For instance, the recognition of m<sup>6</sup>A marks on viral mRNAs by YTHDFs has been shown to block reverse transcription of Zika virus genome (Lichinchi et al., 2016). In the case of Hepatitis C virus, YTHDFs inhibit HCV infection without affecting RNA replication by a mechanism that could be common to most Flaviviridae (Fig.23;Gokhale et al., 2016). Finally, the role of YTHDFs on HIV-1

infection is not clear, as Tirumuru et al. have shown that YTHDFs inhibit infection by decreasing reverse transcription of the viral genome (Tirumuru et al., 2016) while Kennedy et al. presented data supporting a role of YTHDFs as enhancers of the expression of both viral RNA and proteins and of viral replication (Kennedy et al., 2016). Further studies are then clearly needed to clarify the role of m<sup>6</sup>A marks on the infection of human cells by various viruses.

### ***1.10 YTH domain of Plants***

YTH domain proteins from plants deserve a special mention. Plants are highly enriched in YTH domain proteins. *Arabidopsis thaliana* has 13 YTH domain containing proteins and 11 of these proteins were found as Evolutionarily Conserved C-Terminus(ECT) 1-11. Further search showed existence of ECT12 and CPSF30-L. ECT1-11 belongs to the DF family of proteins and ECT12 and CPSF30-L has DC type domain. ECT1 and ECT2 were found to be associated with CIPK1 (Calcineurin B-like-Interacting Protein Kinase1), a factor involved in calcium metabolism (Ok et al., 2005). Among all plant YTH domains, ECT2 has highest abundance and it is known to be involved in trichome branching. Trichomes are elongated branched single cells found at the surface of leaf epidermis. They are differentiated protodermal cells that have stopped mitosis and underwent replication cycles without cell division (a process known as endoreduplication). Further studies showed that ECT2 has a tryptophan pocket to bind m<sup>6</sup>A in a classical way and that the C-terminal domain is required for nuclear localization. The inability of ECT2 to bind m<sup>6</sup>A leads to abnormal trichome branching pattern or an increase in cell ploidy (Scutenaire et al., 2018). Another study showed that instead of conserved DRACH motif, ECT2 binds to a different m<sup>6</sup>A motif, URUAY(R=G>A, Y=U>A). This study proposes that the URUAY motif is plant specific and plant m<sup>6</sup>A methyltransferases have an affinity for this motif (Wei et al., 2018). In addition, in absence of ECT2, m<sup>6</sup>A modified transcripts involved in trichome morphogenesis are destabilized. In fact, ECT2,3,4 were shown to be involved in leaf morphogenesis and trichome branching in an m<sup>6</sup>A-dependent manner (Arribas-Hernández et al., 2018).

CPSF30-L is the only protein in *A. thaliana* that has conserved domains outside the C-terminal YTH domain. CPSF30-L is an orthologue of a 30 kDa subunit of human and yeast cleavage and polyadenylation specific factor. The gene codes for two isoforms of CPSF30, CPSF30-S, containing three CCCH type zinc finger motif and CPSF-L, which has an additional C-terminal YTH domain. This protein is involved the cell signalling pathway by

controlling target of polyadenylation but this function can be performed by the short isoform itself. Although the exclusive role of the long isoform CPSF30-L is to be elucidated, it has been shown that this isoform is capable of localizing and retaining itself in the nucleus whereas the short isoform, CPSF30-S, is dependent on other proteins for its nuclear localization (Li et al., 2017; Rao et al., 2009). Despite the high abundance of YTH domain proteins in plant, deciphering the functions of these proteins remain an active field of investigation. As proteins orthologous to Mei2 RRM3 domain are also found in plants, it will be very interesting in the future to investigate whether a Mei2-Mmi1-like system also exists in plants.

### ***1.11 Role of YTH domain protein in RNA degradation in *S. cerevisiae****

#### ***1.11.1 Structural comparison of Pho92 to human YTH domains***

Pho92 is the only YTH domain containing protein in *S. cerevisiae* having highest sequence identity with YTHDF2 (44%) and YTHDC1 (41%) (Kang et al, 2014) (Fig. 24a) and Pho92-YTH domain is structurally similar to YTHDF2 and YTHDC1 (Xu et al., 2015) (Fig. 24b). The structure of Pho92 bound to UGm<sup>6</sup>ACU provided structural insight into the budding yeast YTH domain (Xu et al., 2015). Pho92 contains the classical YTH fold, consisting of three  $\alpha$  helices and six  $\beta$  sheets. Structural comparison of YTHDF1-GGm<sup>6</sup>ACU with Pho92-YTH reveals m<sup>6</sup>A moiety is accommodated in an aromatic cage where the methyl group is specifically identified by W177, W231 and Y237 (Fig. 24c) and mutation of the tryptophan residues abolishes methylated RNA binding capacity of the protein (Xu et al., 2015). Like human YTH domains, Pho92 has affinity only for m<sup>6</sup>A methylated RNA and not for unmodified RNA (Xu et al., 2015).

#### ***1.11.2 Function of Pho92 in phosphate metabolism***

Due to lack of similarity with other YTH domain proteins, outside the conserved domain, it was difficult to predict the function of Pho92. A microarray analysis in wild-type and  $\Delta$ pho92 strain revealed that several genes involved in phosphate metabolism are specifically affected. Up-regulated genes in  $\Delta$ pho92 included repressible acid phosphatases (PHO5, PHO11, PHO12), a high affinity phosphate transporter (PHO84), an inhibitor of low affinity phosphate transport (SPL2) and among the down-regulated genes were involved in low affinity phosphate transport (PHO87, PHO88, PHO85) indicating Pho92 plays a role in phosphate metabolism (Kang et al, 2014) (Fig. 25a). Northern blot analysis indicated that Pho92 deletion resulted in up-regulation of PHO4 (Fig. 25b), an important transcription

(a)

```
Pho92 SRFFVIKSSSLKHVKRSFYNGI-WSSTHFGNKRLSEAYKKLNSGA---KVFLFFSINTSG 210
Ythdf2 GRVFI I KSYSEDDIHRSIKYNI-WCSTEHGKRLDAAAYRSMNGKG---PVYLLFSVNGSG 465
Ythdc1 ARFFLIKSNHENVSLAKAKGV-WSTLPVNEKKLNLAFRSAR-----SVILIFSVRESG 389
Ect1 AKFFVIKSYSEDDVHNCIKYGA-WSSTPTGNKKLNAAAYEAKENSQCEPVYLLFSVNASG 299
...*:*** . . . . *.: .:*. *: . . * *:*:. **
Pho92 RFCGVAEMVSDLKMDLDTSIWEDEQ-----KYGKAFKVRWVIVRDINNRSCLKRFLIPSNE 266
Ythdf2 HFCGVAEMKSAVDYNTCAGVWSQD-----KWKGRFDVRWIFVKDVPNSQLRHIRLENNE 519
Ythdc1 KFQGFARLSSESHHGGSP IHWVLPAGMSAKMLGGVFKIDWICRREL PFTKSAHLTNPWNE 449
Ect1 QFVGLAEMVGPVDFNKTMEYWQQD-----KWIGCFPVKWHI IKDIPNSLLRHI TLANNE 354
:* *.*.: . . . * * : * :.: :. :. **
Pho92 MKPITHSRDTQEIPYSIGISINLF 280
Ythdf2 NKPVTNSRDTQEVPLEKAKQVLKII 533
Ythdc1 HKPVKIGRDGQEIIELECGTQLCLLF 463
Ect1 NKPVTNSRDTQEVNLEHGTKI IKIF 368
**:. ** **:. . . :. :.
```

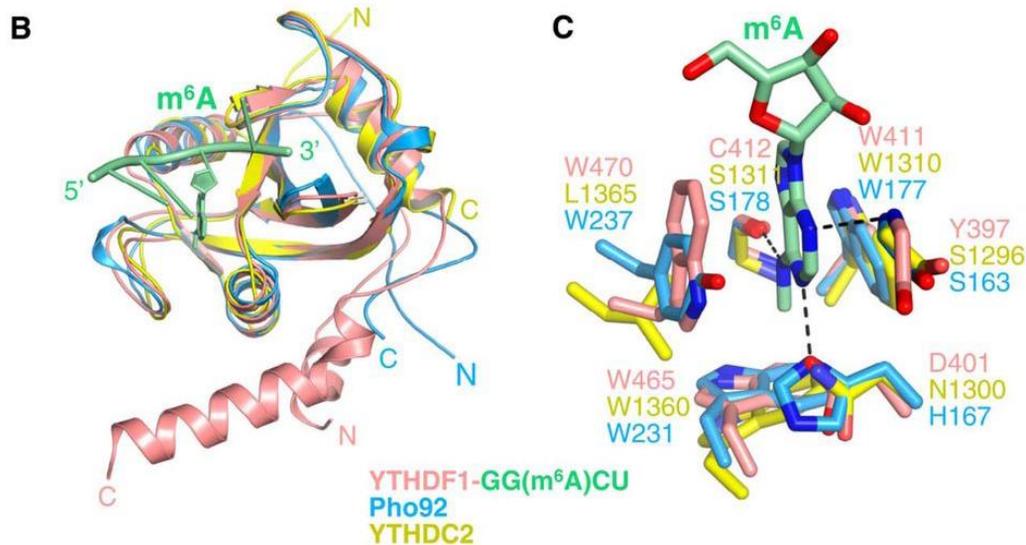


Fig.24: Pho92 YTH domain is similar to classical YTH domain (a) Multiple sequence alignment of Pho92, YTHDF2, YTHDC1 and Ect1. Pho92 has highest homology with YTHDF2 (adapted from, Kang et al., 2014). (b) Superposition of crystal structures Pho92 (blue), YTHDF1 (pink) and YTHDC2 (yellow) (c) Superposition of the m<sup>6</sup>A binding pocket of the three proteins (m<sup>6</sup>A: green stick) (adapted from Xu et al., 2015) (same colour code as panel B).

factor of phosphate metabolism pathway. This suggested that Pho92 plays a crucial role in high affinity phosphate transportation and thus in regulating the cellular phosphate concentration by degrading transcripts of Pho4. Phosphate homeostasis is important regulator of viability and the pathway that controls phosphate metabolism in budding yeast is termed the PHO pathway. There are important regulators of this pathway, transcriptional activators Pho2 and Pho4 as well as the cyclin dependent protein kinases Pho80-85. Pho4 is an important transcription factor of PHO homeostasis pathway that can shuttle between nucleus and cytoplasm. Pho4 is phosphorylated by Pho80-85 cyclin dependent kinases. Under phosphate starved condition, the activity of this complex is inhibited and Pho4 remains non-phosphorylated. This non-phosphorylated form of Pho4 remains in the nucleus and induces transcription of genes encoding high affinity transporters and genes encoding secreted acid phosphatases. In presence of inorganic phosphate, Pho4 is phosphorylated by Pho80-85 cyclin dependent kinase leading to its localization to cytoplasm (Fig. 26).

### ***1.11.3 Pho4 degradation by Pho92***

In a co-IP, Pho92 was found to interact with Pop2 (also known as Caf1) subunit of the Ccr4-Not complex (Fig. 27). Further studies show that 3'UTR of PHO4 transcript is the binding site of Pho92 (Kang et al, 2014). So, the proposed model is that Pho92 binds the PHO4 RNA with its YTH domain and with Pop2 with its N-terminal domain thereby recruiting the Ccr4-Not complex for PHO4 transcript degradation. The human YTH domain having highest homology with Pho92 is YTHDF2. Studies have shown that when YTHDF2 is introduced in Pho92 deleted cells, stability of Pho4 transcripts is reduced and is comparable to the level of Pho4 transcripts in wild type cells. However, similar effect was not observed when Pho92 deleted cells were replaced with YTHDC1 (Kang et al., 2014).

Unpublished yeast two-hybrid results from our collaborator's lab revealed that Pho92 interacts with Not1 of the Ccr4-Not complex (Fig. 28) which is in agreement with the fact that the human YTHDF2 and *S. pombe* Mmi1 also interact with Not1 of Ccr4-Not complex (Cotobal et al., 2015; Stowell et al., 2016). YTHDF2 was found to accelerate deadenylation and subsequent degradation of the target transcripts in a Ccr4-Not dependent fashion (Du et al., 2016). Therefore, it could be proposed that Pho92 can contribute to the degradation of its target transcripts by recruiting Ccr4-Not complex via Not1 (Fig. 29). The interaction between Not1 and Pho92 has been characterized in this thesis.

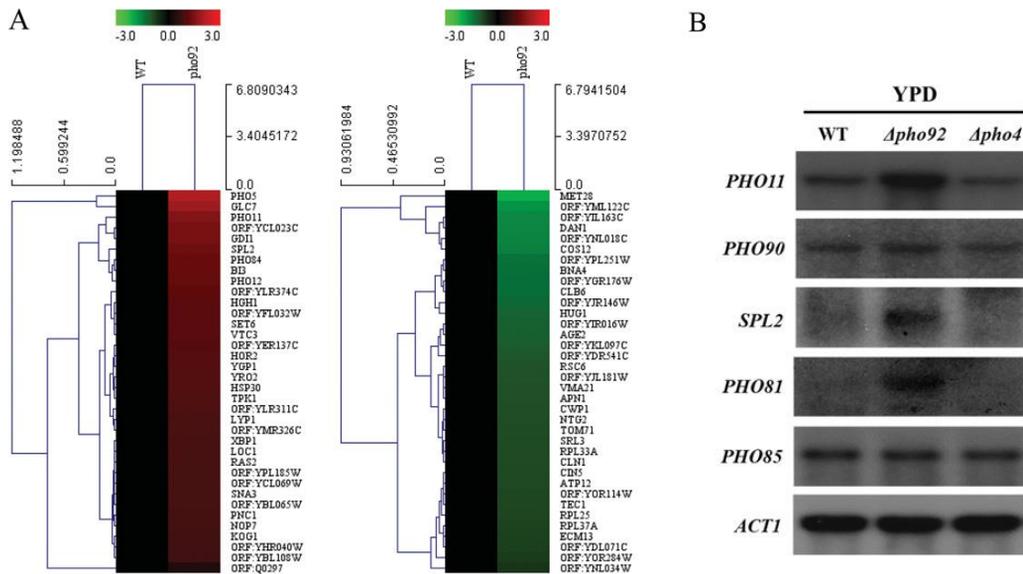


Fig.25: Pho92 is involved in phosphate metabolism (a) microarray analysis result in WT and  $\Delta$ pho92, upregulated and down-regulated genes are shown as heat-map (b) Northern blot analysis to confirm the microarray analysis (adapted from Kang et al., 2014).

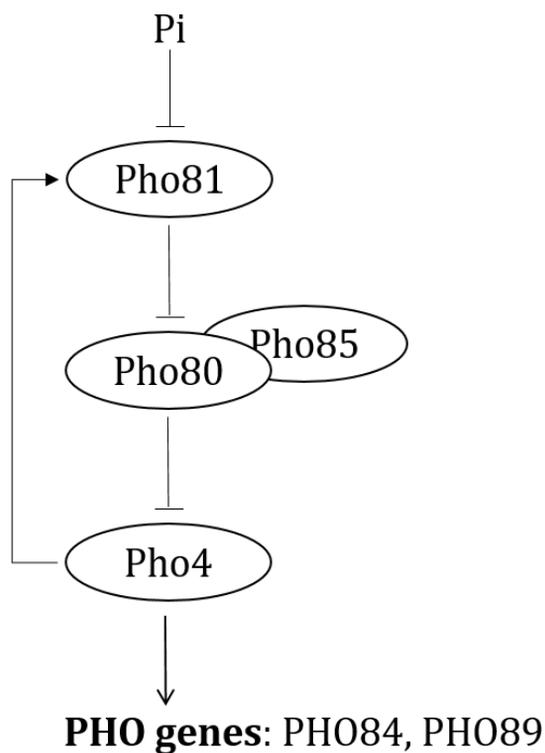


Fig.26: Under Phosphate starved condition, the PHO pathway activates the Pho4 transcription factor through Pho81-mediated inhibition of the Pho85–Pho80 kinase complex. This leads to induced expression of classic PHO genes. Importantly, the PHO81 gene is also induced, creating a positive feedback loop within the pathway (adapted from Swinnen, et al., 2005).

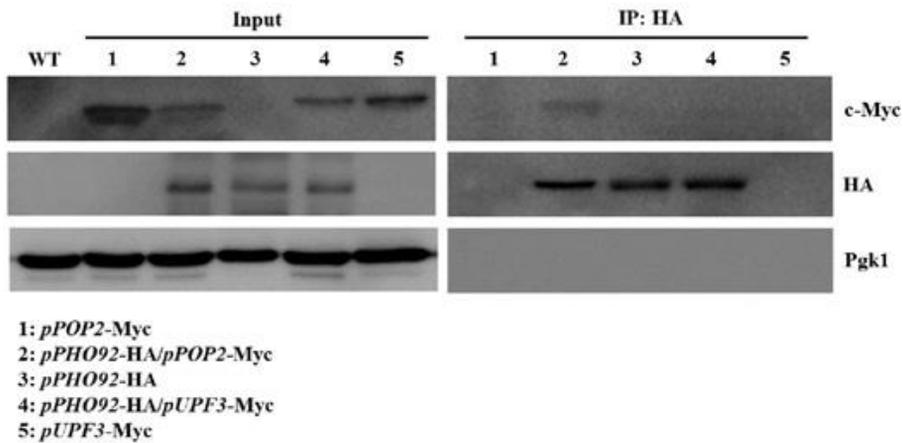


Fig.27: Pho92 interacts with Pop2, a Ccr4-Not subunit. Co-IP with Pho92-HA, Pop2-Myc, and Upf3-Myc. Left panel shows the input and right panel, the immunoprecipitation result. Pop2 immunoprecipitated with Pho92 (adapted from Kang et al., 2014)

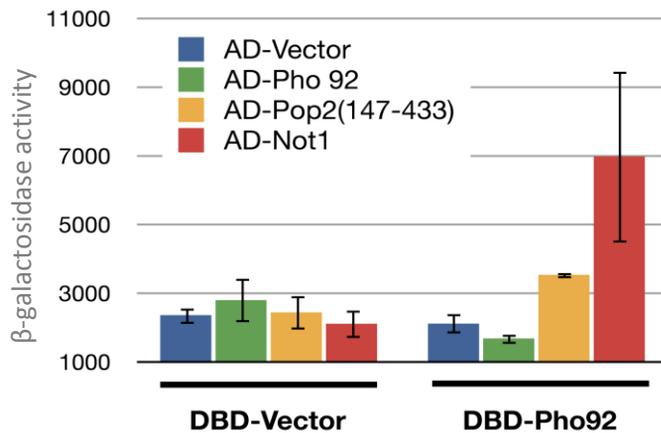


Fig.28: Pho92 interacts with Not1. Unpublished Yeast two hybrid result from our collaborators (Seraphin lab, IGBMC, Strasbourg)

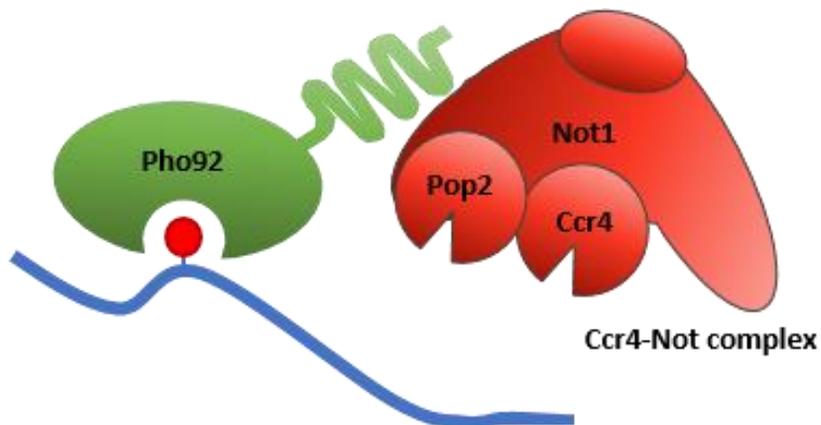


Fig.29: Proposed model showing YTH domain of Pho92 binds to  $m^6A$  modification and recruits Ccr4-Not complex by interacting with Not1 with its NTD, for degradation of its target transcript

#### ***1.11.4 Role of m<sup>6</sup>A in meiosis of budding yeast***

During meiosis, m<sup>6</sup>A accumulates on RNAs while in vegetatively growing cells amount of m<sup>6</sup>A is negligible (Clancy et al. 2002; Bodi et al., 2010). Ime4 is the only m<sup>6</sup>A mRNA methyltransferase in budding yeast and is part of the MIS (Mum2-Ime4-Slz1) complex (Agarwala et al., 2012). For proper progression through meiosis in *S. cerevisiae*, expression of Ime4 is essential. In vegetatively growing cells, Ime4 is repressed by Rme2 (regulator of meiosis 2). Rme2 is an antisense transcript, initiated from 3' end of Ime4 locus. It regulates expression of Ime4 in haploid cells. Haploid yeasts belong to one of the two mating types: a or  $\alpha$  and only a cells can mate with  $\alpha$  cells and vice-versa. Different cell types have different pattern of gene expression. There are a specific genes (asg) expressed in a cells and  $\alpha$  specific genes ( $\alpha$ srg) produced in  $\alpha$  cells. Mating of a and  $\alpha$  type cells create a diploid with a/ $\alpha$  genotype. There are some haploid specific genes (hsg) that are only expressed in haploids and some diploid specific genes (dsg) that are only expressed in diploid cells. Ime4 is a protein that is only expressed in diploid cells because a diploid cell specific repressor, a1/ $\alpha$ 2 silences haploid specific genes. Upon meiotic entry, the a1/ $\alpha$ 2 binds to Rme2 promoter up-regulating Ime4. m<sup>6</sup>A accumulation governs the cell fate but the underlying mechanism still remains a mystery. In budding yeast, the only YTH domain protein, *reader* of m<sup>6</sup>A is Pho92. Expression of the components of MIS complex and Pho92 are up-regulated during meiosis. Deletion of Pho92 delays entry into meiosis although less severely than the deletion of MIS complex (Clancy et al. 2002). It is well established that Pho92 preferentially binds to methylated RNA over non-methylated ones (Schwartz et al., 2013; Xu et al., 2015). So far the only known target of Pho92 is PHO4, a transcription activator in phosphate signal transduction pathway. But PHO4 transcripts are excluded from the list of m<sup>6</sup>A methylated RNAs (Schwartz et al., 2013). These two contradictory results raise the question of how Pho92 degrades its target transcripts. In this thesis, an attempt was made to decipher the underlying mechanism by which Pho92 recruits Not1 to degrade mRNAs during meiosis.

# CHAPTER II

## OBJECTIVES



## 2.0 Objectives

RNA metabolism is a crucial step in any cellular process. There are different RNA binding proteins that determine the course of RNA metabolism. One of the most recently discovered RNA binding domain is YTH domain. YTH domain proteins specifically recognizes N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) modification of mRNA. This is the most abundant, dynamic modification of mRNA that was discovered in the 70s. However, further studies were not possible at that time as this modification does not alter base pairing and the technology available did not allow precise mapping of m<sup>6</sup>A transcriptome. In 2011, it was discovered that m<sup>6</sup>A is the substrate of a human obesity related protein, FTO. This discovery resumed the study of m<sup>6</sup>A modification, giving rise to the field of Epitranscriptomics. The YTH domain proteins that were discovered in late 90s, constitute the majority of reader proteins that can specifically identify m<sup>6</sup>A modification. Apart from the YTH domain, there is no sequence homology between these proteins but their cellular function is diverse. Indicating that the YTH domain proteins recruit different effectors to direct the cell fate. YTH domain proteins are abundant in eukaryotes and absent in prokaryotes. Human beings have five YTH domain proteins that are functionally different, YTHDF1-3, YTHDC1-2. Although it is evident that these proteins are controlling cellular fate, the function of each protein and their network is yet to be clarified.

The budding yeast *Saccharomyces cerevisiae* and fission yeast *Schizosaccharomyces pombe* are the two most common model organisms for studying eukaryotic biological processes. Both of these yeasts have only one YTH domain containing protein, Pho92 in budding yeast and Mmi1 in fission yeast. Pho92 is a classical YTH domain whereas Mmi1 is an exceptional YTH domain that binds to a unmethylated conserved RNA motif instead of m<sup>6</sup>A-containing motif. Outside the YTH domain, there is no similarity between these proteins. But surprisingly, both YTH domain proteins are involved in the regulation of mitosis-meiosis switch by degrading target transcripts. In this thesis, the objective was to characterize the function and network of these two YTH domain proteins: Mmi1 and Pho92.

### 2.1 Characterization of Mmi1 regulation:

Mmi1 is the only YTH domain protein of *Schizosaccharomyces pombe*. Mmi1 has been studied for long time for its role in regulation of mitosis. Mmi1 degrades meiotic transcripts in vegetatively growing cells. Mmi1 forms a complex with a small protein, Erh1, called EMC (Erh1-Mmi1-Complex)(Sugiyama et al., 2016). In a mitotically growing cell, EMC down-

regulates the meiotic genes either by degrading the transcripts or by forming heterochromatin at their genetic loci. To achieve this function, EMC recruits two multi-subunit complexes, MTREC or Ccr4-Not. When cell switches to meiosis, a RNA binding protein, Mei2, along with a long non-coding RNA, meiRNA sequesters Mmi1 to a nuclear foci allowing smooth progression of meiosis (Harigaya et al., 2006). In a recent study, it has been discovered that in vegetatively growing cells, Mmi1 degrades Mei2 in a ubiquitination-dependent manner, by recruiting Ccr4-Not complex (Simonetti et al., 2017). The objective of this thesis was to characterize the interaction between Mmi1-Erh1 and Mei2. In a previous study, interaction has been reported between Mmi1 and Mei2 (Harigaya et al., 2006). The following goals were set to elucidate interaction between Mmi1 and Mei2:

- i. Cloning, expression and crystallization of Mei2-RRM1-2 and Mei2-RRM3
- ii. RNA substrate determination of Mei2-RRM3
- iii. To determine if the interaction between Mmi1 and Mei2 is bridged by RNA
- iv. Co-crystallization trial of Mmi1-YTH, Mei2-RRM3 and RNA carrying the binding site for both proteins

The second part was to find out what is the role of Erh1 in Mmi1's function and if there is any interaction between EMC and Mei2. To obtain this information, subsequent steps were set:

- i. Identification of Mmi1 domain that interacts with Erh1
- ii. Reconstitution EMC
- iii. Crystallization of Erh1-Mmi1 complex/ Erh1 alone
- iv. Structural and functional characterization of Erh1 and Erh1-Mmi1 complex
- v. To investigate if there is any interaction between EMC and Mei2

## ***2.2 Characterization of Pho92 function:***

In *S. cerevisiae*, there is only one YTH domain protein, Pho92. The Pho92 YTH domain is structurally similar to other classical YTH domain proteins such as YTHDC1 or YTHDF2. Sequence analysis indicates that Pho92 has highest sequence homology (44%) with YTHDF2 (Kang et al., 2014). The function of Pho92 has not been fully characterized yet. The only

available information is that Pho92 degrades transcripts of a transcriptional activator, *pho4*, Pho92 is up-regulated during meiosis, it interacts with Pop2 of Ccr4-Not complex (Kang et al., 2014) and *in vitro* Pho92 prefers m<sup>6</sup>A methylated RNA over unmethylated substrates. From our collaborators unpublished result of yeast two hybrid assay, interaction was detected between Pho92 and Not1, the scaffolding subunit of Ccr4-Not complex. While this work was ongoing, it was published that YTHDF2, the human homologue of Pho92, degrades m<sup>6</sup>A containing transcripts by recruiting Ccr4-Not via Not1 (Du et al., 2016). Based on these results, our hypothesis is that *S. cerevisiae* Pho92 binds to m<sup>6</sup>A methylated transcripts and leads to degradation of the target transcripts by recruiting Ccr4-Not complex. In order to delineate the interaction between Not1 and Pho92, the following objectives were set:

- i. To determine whether Not1 and Pho92 interact physically or if their interaction is bridged by RNA.
- ii. Determination of the domain boundary for Not1 and Pho92
- iii. Crystallization and structure solution of Not1-Pho92 complex
- iv. Structural and functional characterization of the Not1-Pho92 complex.



CHAPTER III  
GENERAL MATERIALS  
AND  
METHODS



### 3.1 Cloning:

The coding sequences of desired genes were amplified by PCR from *S. pombe* cDNA library (for Mmi1, Mei2 and Erh1 genes) or *S. cerevisiae* genomic DNA (Not1 and Pho92), using respective primers (refer to Table S2). The composition of the reaction mixture is detailed in Table 3.1.

Table 3.1: Composition of the reaction mixture for PCR amplification:

Component stock	Final Conc <sup>n</sup>	Volume ( $\mu$ l)
DNase free Water		75.0
PCR Buffer (5x HF buffer)	1x	20.0
Template DNA	50ng/ $\mu$ l	1.0
FP (100pmol/ $\mu$ l)	0.1pmol/ $\mu$ l	1.0
RP (100pmol/ $\mu$ l)	0.1pmol/ $\mu$ l	1.0
20mM dNTPs	0.2mM	1.0
Phusion pol (2U/ $\mu$ l)	2.0U	1.0
Total		100.0

The PCR reactions were performed using the protocol described in Table 3.2.

Table 3.2: Steps for PCR amplification

No. of Steps	Steps	Temperature ( $^{\circ}$ C)	Duration (s)
1.	Initial Denaturation	95	300
2.	Denaturation	95	10
3.	Annealing	$(T_m-5)^{\circ}$	15
4.	Extension	72	X
Steps 2 to 4 were repeated for 30 cycles			
5	Final extension	72	600
6	Final storage	4	$\infty$

Where X is 30s/kb for plasmid and 45s/kb for genomic DNA.

The PCR products were purified by NucleoSpin Gel and PCR clean up kit from Macherey-Nagel according to the manufacturer's instructions.

Restriction digestions were performed on the purified PCR products and the desired vectors (Refer to Table S2) at 37°C using FastDigest (Thermo Fisher Scientific) enzymes. The digested products were separated by gel electrophoresis (1% agarose gel) with 1X TBE (Tris-Borate-EDTA) buffer. Well separated products were purified from the agarose gel using Macherey-Nagel NucleoSpin Gel and PCR clean up kit.

For ligation, the reaction mixture contained purified digested insert and vector in 3:1 molar ratio in total reaction volume of 20µl. Ligation was carried out at 22°C for 15 minutes using 1U of T4 ligase (Thermo Fisher Scientific). Counter-selection was performed by adding 1µl of a restriction enzyme cleavage site of which is located in the middle of the two restriction digestion enzyme sites on the MCS.

XL1-blue *E. coli* bacteria were transformed with the ligation products. For transformation 7-10µl of ligation sample was mixed with 50µl of XL1-Blue and incubated in ice for 15 minutes, followed by a heat shock at 42°C for 50 seconds. After the heat shock, the tube is immediately transferred to ice and incubated in ice for 5 minutes. 500µl of 2YT media is added and incubated at 37°C for one hour before plating on Petri-dishes with LB-agar medium supplemented with corresponding antibiotic. After overnight incubation at 37°C, some colonies were picked for DNA extraction by classical mini-prep protocol (NucleoSpin plasmid preparation kit from Macherey-Nagel). The sequence of obtained plasmids were checked by sequencing.

### ***3.1.1 Site directed mutagenesis:***

The plasmid encoding for the protein mutants were obtained by efficient one-step site-directed mutagenesis using partly overlapping primers (Table S4) following the protocol described by Zheng et al (Zheng et al., 2004). Details about PCR amplification reactions can be found in Table3.3.

Table 3.3: Steps for PCR amplification for Site directed mutagenesis

No. of Steps	Steps	Temperature (°C)	Duration (s)
1.	Initial Denaturation	95	300
2.	Denaturation	95	15
3.	Annealing	68	15
4.	Extension	72	X
To step 2 for 30 cycles			
5	Final extension	72	600
6	Final storage	4	∞

Where X= 30s/kb, as the template was a plasmid.

The PCR product was purified with NucleoSpin plasmid preparation kit from Macherey-Nagel according to the manufacturer's instructions and the product was digested with DpnI to degrade methylated template plasmid. The plasmid containing mutation was transformed into XL1-Blue cells following the protocol described in "cloning" section.

### ***3.2 Solubility profiling:***

To identify the experimental conditions suitable for maximum production of desired protein in soluble state, solubility profiling was done. The corresponding plasmid was used to transform BL21 (DE) Gold cells or BL21 (DE) Codon plus RIL cells. From over-night cultures, 50 µl of cells were used to inoculate 5 ml of 2YT or TBAI (Terrific Broth Auto-Inducible) media, along with required antibiotics and cells were grown at 37°C. For cells grown in 2YT media, protein expression was induced with different concentrations of IPTG (100µM, 250µM) when OD<sub>600nm</sub> reached 0.6. After induction (2YT media) or when cells OD<sub>600nm</sub> reached 0.6 (TBAI), cells were transferred either at 18°C overnight or kept at 37°C

for 4 hours. Cells were collected by centrifugation (at 2000 rcf, 4 °C, 20 mins) and the pellets were resuspended in 1 ml lysis buffer (20 mM Tris-HCl pH 7.5, 200 mM NaCl, 5mM 2-mercaptoethanol (or  $\beta$ Me)). Cells were sonicated for 75 sec pulse, 60 sec rest, for five times, and centrifuged (at 13,000 rcf, 4°C, 20 mins) to remove cell debris. Supernatant was collected and along with expressed cells and control (without IPTG), was loaded to small scale columns containing 30 $\mu$ l of Glutathione-sepharose matrix (Glutathione Sepharose 4B from GE-Healthcare) or Ni-NTA matrix (Protino Ni-NTA agarose from Macherey-Nagel). After washing with 1 mL lysis buffer, elution was performed with 60 $\mu$ L of Elution buffer (Refer to Table S5). Samples were prepared by mixing 20 $\mu$ l of elution with 5x loading dye (5 $\mu$ l). 12  $\mu$ l of sample were loaded on 15% SDS PAGE gel. After Coomassie-Blue staining of the gels, the condition yielding to the highest expression level of the protein of interest was selected.

### ***3.3 Protein over-expression:***

Depending on the result of expression test, the ideal condition was selected for expression either in YT media with IPTG induction or TBAI media, with correct antibiotic resistance. Transformed bacteria cells were incubated overnight with 50 ml of 2YT or TBAI media at 37°C and after 16 hours of incubation, 1L of media was inoculated with 10 ml of the overnight culture. After OD<sub>600</sub> reached 0.6 at 37°C, culture was transferred to 18°C. After a post-induction period of 16 hours, cells were harvested by centrifugation (4,100 rcf, 45 minutes).

#### ***3.3.1 Selenomethionine labelling:***

For Se-Met labelling, the cells were grown in a special media containing selenomethionine instead of methionine. To over-express selenomethionine labelled protein, 10 mL of overnight culture was used as inoculum for 1 L selenomethionine containing minimal media. Culture was grown at 37°C until OD<sub>600</sub> reached 0.6 and then induced with 100 $\mu$ M IPTG and transferred to 18°C for 16 h. The Se labelled protein was purified following the protocol used for purification of the native protein.

### ***3.4 Purification:***

#### ***General overview of purification steps:***

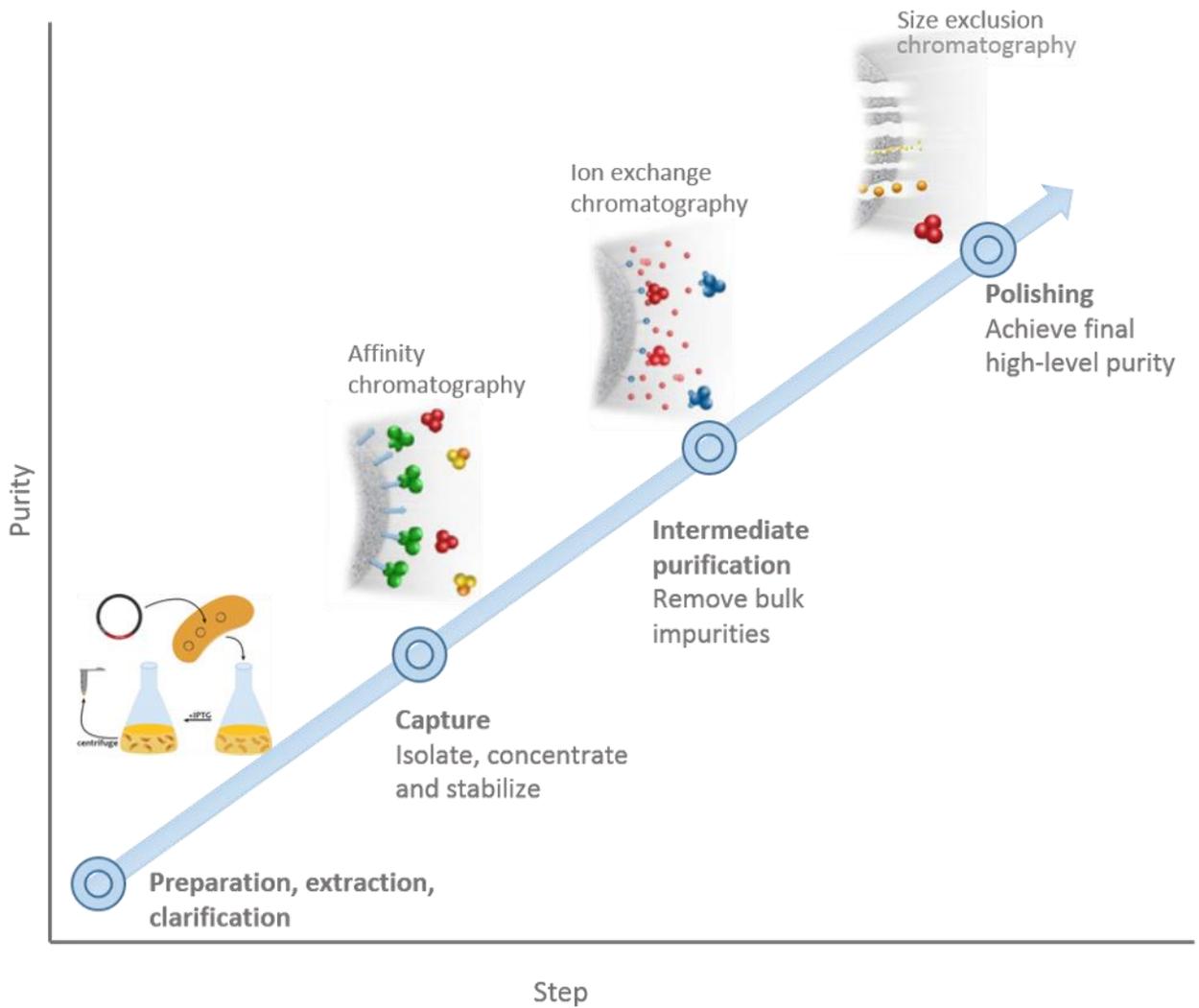


Fig.30: Schematic representation of purification steps followed to obtain a protein with sufficient purity and quantity. (i) Preparation, extraction and clarification: the first step of purification is sample preparation. The objective of this step is to obtain clarified extract of source material. (ii) Capture: in this step the target product is isolated, concentrated and stabilized. (iii) Intermediate purification: bulk impurities, like other proteins and nucleic acids are removed in this step. (iv) Polishing: all impurities are removed in this stage and the protein is transferred in a condition ideal for crystallization and storage

Purification of the protein of interest in sufficient quantity and quality from bacterial cell lysate is a challenging job. The Capture, Intermediate Purification and Polishing (CIPP) is used for simple planning of protein purification strategy (Fig. 30). The starting point of protein purification is sample preparation. This step involves obtaining a clarified extract of source material. To do this cells were harvested by centrifugation (4,100 rcf, 45 minutes), after a post-induction period of 16 hours. The supernatant was discarded. The cell pellet obtained from harvesting the cells, was resuspended in 50 ml lysis buffer (20mM TrisHCl, pH 7.5, 200mM NaCl, 5mM  $\beta$ Me) containing 100  $\mu$ M PMSF. The cells were lysed by sonication, 75 seconds pulse, 60 seconds rest for 6 times. The lysate was centrifuged at 20,000 rcf for 45 minutes. The supernatant is then loaded to affinity chromatography for a first step of purification.

### ***3.4.1 Affinity chromatography:***

Affinity chromatography separates protein on the basis of a reversible interaction between protein of interest and a ligand, immobilized on an insoluble support and packed in a column. The sample is applied to column under conditions favourable for interaction between protein and ligand, after successive rounds of washing to remove non-specific proteins, the conditions are changed so that the target protein can be eluted from the column. All the affinity purifications described in this thesis were either GST or Ni-NTA affinity chromatography as the recombinant proteins contain either GST or His<sub>6</sub> tags. When protein mixture is applied to the column only the protein containing the tag will bind and it can be eluted from the column in highly pure form.

#### ***3.4.1.1 GST-purification:***

For GST-fused proteins, the supernatant was loaded in a gravity column containing Glutathione-sepharose matrix (GE Healthcare Biosciences) pre-equilibrated with lysis buffer (2 ml of matrix was used for 1 L culture). The matrix was mixed with the supernatant on a rotating wheel for one hour. The column was extensively washed with first lysis buffer to remove any non-specifically bound proteins, and then a high salt buffer wash (20mM Tris-HCl, pH 7.5, 2M NaCl, 5mM  $\beta$ Me) to remove any nucleic acids contamination. The GST-tagged protein was eluted with elution buffer (20mM Tris-HCl, pH 7.5, 200mM NaCl, 20mM GSH, 5mM  $\beta$ Me). After elution, the GST-tag was cleaved by 3C protease (10U for 1mg of protein) with overnight dialysis against lysis buffer. After cleavage, the tag was removed either by passing through an ion-exchange column or Glutathione-sepharose matrix.

#### **3.4.1.2 Ni-NTA purification:**

The basic technique is similar to GST purification. Ni-NTA agarose resin (Proteino, Macherey Nagel) was equilibrated with the lysis buffer (2ml matrix was used for 1l culture). After mixing the matrix with the supernatant for one hour on a rotating wheel, the flow-through was collected. Extensive washing steps were performed with lysis buffer, followed by high salt buffer to remove nucleic acid contamination and by wash buffer (20mM Tris-HCl, pH7.5, 200mM NaCl, 20mM Imidazole, 5mM  $\beta$ Me). Finally, the protein was eluted by an elution buffer containing 400mM imidazole (20mM Tris-HCl, pH7.5, 200mM NaCl, 400mM imidazole, 5mM  $\beta$ Me).

For proteins expressed as fusion with His<sub>6</sub>-tag followed by the repeat of Z domains (ZZ), an immunoglobulin-binding domain engineered from *S. aureus* protein A (Samuelsson et al., 1994), the His<sub>6</sub>-ZZ tag was removed by 3C protease, by overnight dialysis as described above.

The protein eluted from affinity chromatography was concentrated using Vivaspin concentrator for further purification steps.

#### **3.4.2 Ion exchange chromatography:**

In ion-exchange chromatography, proteins are purified on the basis of their global charge. In anion-exchange chromatography, positively charged beads are used to purify proteins that have a net negative charge on their surface while in cation-exchange chromatography, negatively charged beads are used to purify positively charged proteins. The interaction between protein and the charged beads can be weakened with increasing ionic strength. Hence, the protein is eluted from ion-exchange chromatography column by a continuous salt gradient (50mM-1M NaCl). Ion exchange was performed on ÄKTA Purifier system (GE Healthcare Biosciences) with 5 ml HiTrap Q-FF (anion exchanger). Another method used to purify DNA/RNA binding proteins, was Heparin column. Heparin is a highly sulphated glycosaminoglycan that mimics the negatively charged phosphate backbone of nucleic acids. The interaction between Heparin and protein can be weakened by high salt concentration. In this case, 5ml HiTrap Heparin HP column (GE) was used for purification using the same protocol as for ion-exchange chromatography.

The column was equilibrated with Buffer A (50mM Tris-HCl, 50mM NaCl, 5mM $\beta$ Me) prior to injection of protein sample. 5 ml protein sample in Buffer A was injected. Elution was

performed using a linear gradient ranging from 100% Buffer A to 100% Buffer B(50mM Tris-HCl, 1M NaCl, 5mM  $\beta$ Me).

All the buffers mentioned in this section are standard buffers. However, depending on the proteins, buffer compositions were optimized and are described in the results section.

### ***3.4.3 Size exclusion chromatography:***

The final step is the Size Exclusion Chromatography. As the name suggests, this chromatography separates proteins on the basis of hydrodynamic volume and size. The porous beads packed in a column allow smaller proteins to enter the pores and slowly elute from the column while the larger proteins are incapable of entering the column and then elute earlier from the column. The elution from affinity or ion-exchange column, containing the protein of interest was concentrated by Vivaspin concentrator of the correct molecular weight cut-off (Sartorius) and subjected to size exclusion chromatography using HiLoad 16/600 Superdex 75 or Superdex 200 prep grade columns (GE Healthcare Biosciences) on an ÄKTA Purifier system (GE Healthcare Biosciences). Gel filtration buffer was used for size exclusion chromatography. The standard gel filtration buffer has the following composition: 20mM Tris-HCl, pH7.5, 200mM NaCl, 5mM 2-mercaptoethanol, however depending on the protein, the gel filtration buffer varied. Buffer composition for each step of purification for all proteins are tabulated in Table S5. The purified protein was concentrated using Vivaspin concentrator. The concentration of the protein was determined by measuring the absorbance at 280 nm with a nanodrop spectrophotometer and purity was verified by 15% SDS-PAGE.

### ***3.5 Fluorescence anisotropy:***

Fluorescence anisotropy or fluorescence polarization is a standard method to study interactions between two molecules. It measures the change in orientation of the molecule in space with respect to time. A polarization filter allows light wave of a single orientation to pass. This light wave excites a fluorophore attached to a biomolecule. If a vertical polarizer is used for excitation only vertically polarized light will be transmitted. When this vertically polarized light excites a bio-molecule, vertically polarized light is emitted. The small molecules in the solution are constantly tumbling, therefore the orientation of the emitted light changes with the orientation of the molecule. If the fluorophore tagged bio-molecule is polarized with vertically polarized light the emission will retain vertical polarization. Using another polarization filter the strength of the emitted light can be measured. The emitted light which has the same orientation as the polarizer is allowed to pass but emitted light with an

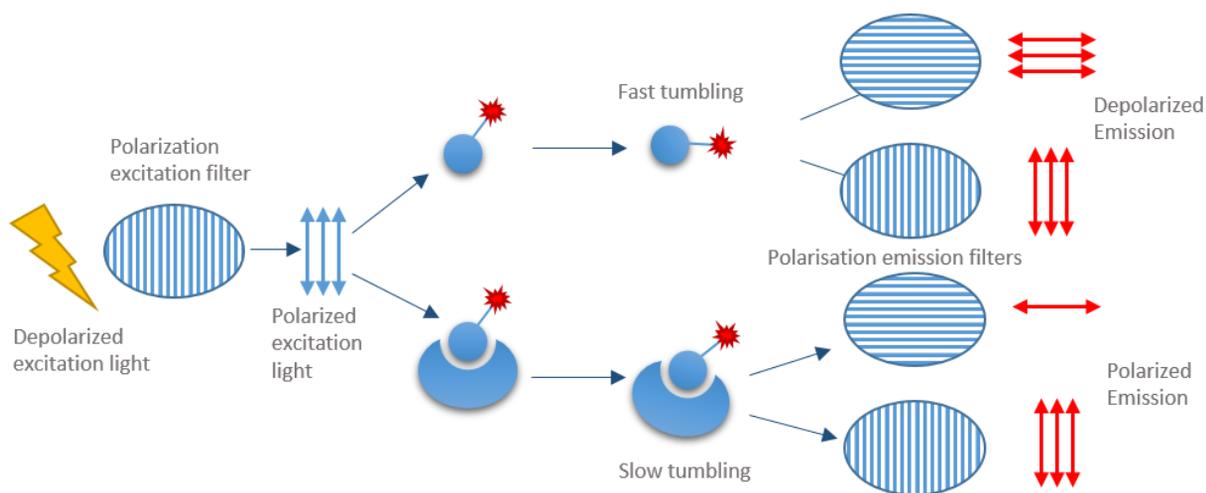


Fig.31: Principle of fluorescence polarization: when a fluorophore is excited with polarized light the emission is also polarized. A freely tumbling fluorophore in the solution scrambles the polarization by radiating at a different direction than incident light. When a fluorophore tagged ligand interacts with a larger molecule, it will slow down the speed of tumbling which will increase polarization of the emitted light reduce the scrambling (adapted from Tgk Scientific)

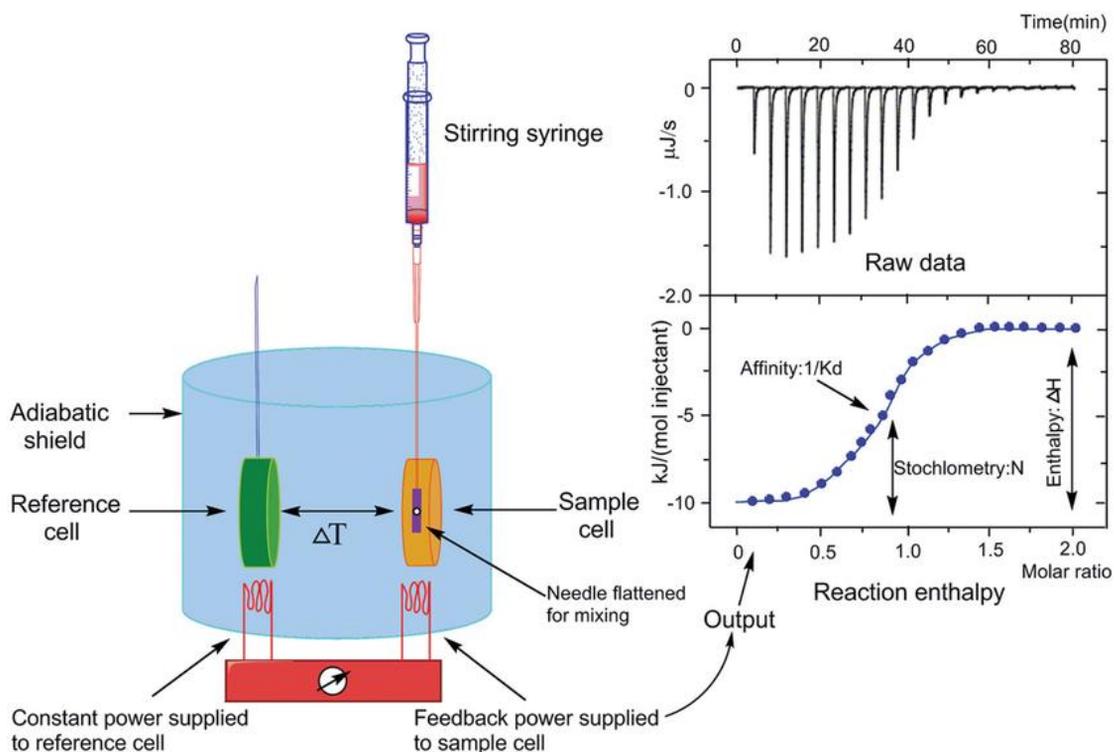


Fig.32: Basic principle of isothermal titration calorimetry. Schematic representation of the isothermal titration calorimeter (left) and a characteristic titration experiment (upper right panel) with its evaluation (lower right panel). In upper right panel, the titration isotherm is represented as heat release per unit of time after each injection of the ligand into the protein solution (sample cell). The lower right panel represents, the relation of released heat after each injection with the molar ratio of total ligand concentration and total protein concentration. Experimental data points are represented as blue dots and the line corresponds to the best fit. (Freyer and Lewis, 2008; Martinez et al., 2013).

orientation perpendicular to the polarized is completely blocked and this signal is recorded. The faster the bio-molecule tumbles the more depolarized the emitted light will be. When a ligand binds to the bio-molecule, the speed of tumbling decreases, this results in retention of vertically polarized light for a longer time. Increase of molecular weight is positively correlated with polarization signal (Fig. 31).

To use this information, polarizers are used in excitation and emission pathways of a fluorometer. Using a vertical direction for the polarized excitation source, the steady-state fluorescence anisotropy ( $r$ ) was calculated according to:

$$r = \frac{I_V - I_H}{I_V + 2I_H}$$

where  $I_V$  and  $I_H$  correspond to the parallel (vertical) and perpendicular (horizontal) fluorescence emission intensity components, respectively.

Interaction between Not1 (1343-1565) and FITC tagged Pho92 (21-44) peptide (PM194) was studied by steady state fluorescence anisotropy parameter using FP-8300 Spectrofluorometer (JASCO) at 20°C, with excitation and emission slits 2.5 and 5 nm respectively. The excitation and emission wavelengths were centred at 495 and 520, respectively. Titration experiments were performed in 20mM Tris-HCl at pH 8.0, 100mM NaCl, 5mM 2-mercaptoethanol in total reaction volume of 1ml by maintaining a constant concentration of PM194 (8 nM) while adding increasing concentrations of Not1 (1343-1565) (0-8.0  $\mu$ M). The affinity  $K_d$  and  $dR_{max}$  (maximum value of anisotropy) were calculated using the following formula:

$$dR = (dR_{max} * [Not1]) / (K_d + [Not1])$$

where  $dR$  is the anisotropy difference for a given Not1 concentration ( $[Not1]$ ). This relationship takes into account the change in the fluorescence intensity of PM194 observed along the titration. The  $K_d$  characterizing the Not1/PM194 complex was calculated by fitting the plot of  $dR$  versus Not1 concentration ( $\mu$ M) using the OriginPro 9.0 software.

### **3.6 ITC:**

Isothermal Titration Calorimetry is a sensitive technique to determine thermodynamic parameters of a binding reaction. ITC can accurately measure the heat change when two

molecules interact. The ITC machine has an adiabatic chamber containing a sample cell and a reference cell. The reference cell contains buffer and the sample cell contains one of the binding partners (receptor) and a paddle shaped needle containing the other binding partner (ligand). The needle injects the ligand in the sample cell in small aliquots (2 $\mu$ l) at regular intervals under constant stirring to mix the binding partners (700rpm). The interaction between two molecules is either exothermic or endothermic. The heat required to keep zero temperature difference between the sample cell and the reference cell is measured. The heat signature produced for each injection is plotted as a function of time to generate an isotherm. This curve is then normalized for concentration. The thermodynamic parameters, affinity ( $K_d$ ), stoichiometry (n) and the enthalpy of interaction ( $\Delta H$ ) can be calculated from this titration curve (Fig. 32).

Isothermal titration calorimetry (ITC) was used to quantify thermodynamic parameters of Mei2-RRM3-RNA binding event. Prior to ITC experiments, protein and RNA samples were dialyzed against 20mM Tris-HCl pH7.5, 200mM NaCl, 5mM  $\beta$ Me. The exact concentration was measured by absorption spectroscopy, from known extinction coefficient at OD<sub>280</sub> for protein and OD<sub>260</sub> for RNA. The interaction between RRM3 and different RNA constructs (Dharmacon) was studied by ITC with an ITC<sub>200</sub> machine at 20°C. For all ITC experiments, 200  $\mu$ L of protein at 10  $\mu$ M concentration were titrated by several injections of 2  $\mu$ L of RNA (100  $\mu$ M) at intervals of 180 s. A theoretical curve assuming a one-binding site model calculated using Origin Software (MicroCal Inc.) gave the best fit to the experimental data.

### ***3.7 Limited proteolysis:***

Limited proteolysis is a widely used technique to trim the flexible parts of proteins, that can hinder crystallization. For this technique, the protein is incubated with very low concentration of different proteases (trypsin, chymotrypsin, papain, ...). The well folded parts of the protein are difficult to be accessed by the proteases while the unfolded parts are more prone to be cleaved by the enzymes. After limited proteolysis, the results are analyzed by SDS-PAGE. Appearance of bands of lower molecular weight than the full-length intact protein indicates cleavage of the protein. The parameters that can be obtained from the limited proteolysis trials are, type of protease, concentration of the protease and time of digestion.

Multiple crystallization trials failed to crystallize co-purified Not1 and Pho92 protein complex. The N-terminal of Pho92, which interacts with Not1 is a low complexity region lacking secondary structure elements according to predictions. To overcome the

disadvantage, limited proteolysis was performed with Not1-Pho92 complex of different constructs. To optimize the enzyme and concentration, 20 µg of protein was incubated with varying concentrations of proteases. The ratio of protein: trypsin was 1:100, 1:200, 1:1000 (w/w) and protein: chymotrypsin 1:100 and 1:200 (w/w). The reaction was carried out at 20°C for 20 minutes. The result was analysed by running the samples on SDS PAGE with undigested protein as control.

### **3.8 Pull down:**

Pull down experiments were performed to determine the interaction between two proteins *in-vitro*. Pull-down technique is a modified form of affinity purification, where instead of a column packed with immobilized ligands (GST/Ni-NTA), magnetic beads are used. In pull-down experiment, a tagged-protein (bait) is applied on an affinity ligand specific for its tag. If other proteins interact with the bait protein, they will also be retained on the immobilized affinity beads and hence will be purified.

The experiments were performed by mixing 50 µg of a tagged protein (bait), with 1.2 molar excess of its putative partner (prey). In case of GST-tagged proteins, the prey was also mixed with GST alone as a negative control. Binding buffer (20 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM β-mercaptoethanol) was added to a final volume of 60 µL. The reaction mixtures were incubated on ice for 1 hour. 10 µL were withdrawn as “input” fraction for SDS-PAGE analysis. The remaining 50 µL were incubated with 500 µg of GST-agarose magnetic beads (Glutathione Magnetic Agarose, Thermo Scientific) equilibrated in binding buffer to a final volume of 200 µL at 4°C for 1 hour. Washing steps were performed with lysis buffer. Bound proteins were eluted with binding buffer complemented with 20mM GSH. Samples were initially resolved on 15% SDS-PAGE and visualized by Coomassie blue staining

Pull down experiments with NiNTA-magnetic beads (HisPur<sup>TM</sup> Ni-NTA Magnetic Beads, Thermo Scientific) were performed using the protocol described above for GST-pull-down with slight differences, *i.e.* beads were washed three times with 500 µL of Wash Buffer (Binding buffer+20mM Imidazole) and the elution buffer contained 400mM imidazole (pH 7.5) instead of 20mM GSH.

### **3.9 Thermal shift assay:**

Thermal shift assay is a standard technique to determine the effect of different buffer components on the stability of a protein. A fluorescent dye, either SYPRO Orange or 1-anilino-8-naphthalenesulfonate (ANS), is used with a Real Time-PCR machine to study thermal denaturation of proteins under varying conditions. Fluorescence of SYPRO Orange is quenched by aqueous environment but when this dye binds to a hydrophobic environment, it becomes strongly fluorescent. Based on this principle, the dye is mixed with protein in different buffer conditions, and the temperature is gradually increased. As the protein unfolds, the dye interacts with hydrophobic regions from the studied protein and the fluorescence signal can be detected. Fluorescence signal is plotted as a function of temperature to calculate the melting temperature ( $T_m$ ). Increment of melting temperature indicates increased stability of protein (Fig. 33).

Thermal shift assay, also known as the differential thermal scanning, was used to study stability of Mei2-RRM1-2 construct in different buffer composition. To perform this assay, concentrated protein (1.0 mg/ml) is mixed with solutions of different buffer conditions, in a Thin-Wall 96-well skirted PCR plates (BioRad) and the Sypro Orange dye. In a real time PCR machine (BioRad CFX96 Real-Time PCR system), the temperature is gradually increased. The recorded fluorescence signal from SyproOrange is used to calculate the melting temperature ( $T_m$ ) in different conditions. Increase in  $T_m$  indicates improved stability of protein.

### ***3.10 SEC-MALLS:***

To precisely determine the molecular weight and oligomeric state of a protein alone, a protein-RNA or a protein-protein complex in solution, experiments of Size exclusion chromatography along with a multiple-angle laser light scattering (SEC-MALLS) were performed. Normally, a SEC is coupled with a UV detector where elution volume is calibrated against known standards (Fig. 34). Often for proteins with non-globular shape and sticky in nature, using SEC alone can give erroneous estimation of molecular weight. The use of a multiple-angle laser light scattering along with SEC can overcome this shortcoming to accurately determine the MW of a macromolecule independently of its shape and directly from scattered light from three different lasers set at different angles. Intensity of the scattered light is directly proportional to average molecular weight and concentration of the solution. This is a direct method to determine the accurate molecular weight of the sample without any calibration curve.

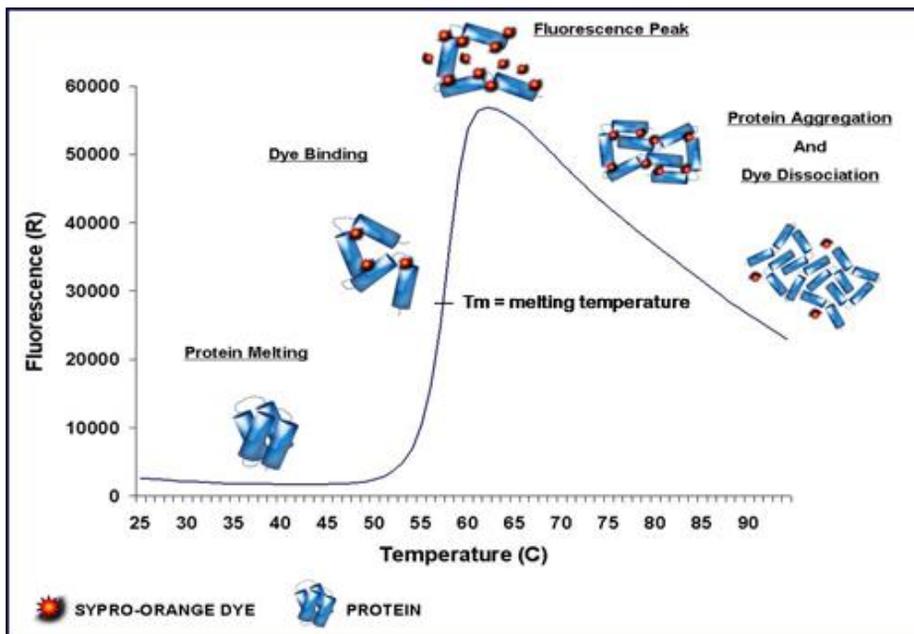


Fig.33: Schematic representation of Thermal Shift assay. Fluorescent dye is mixed with protein. With increasing temperature, protein unfolds the dye binds to hydrophobic patch of protein and fluoresces. This fluorescent signal is plotted as a function of temperature and the melting temperature can be calculated from this graph.

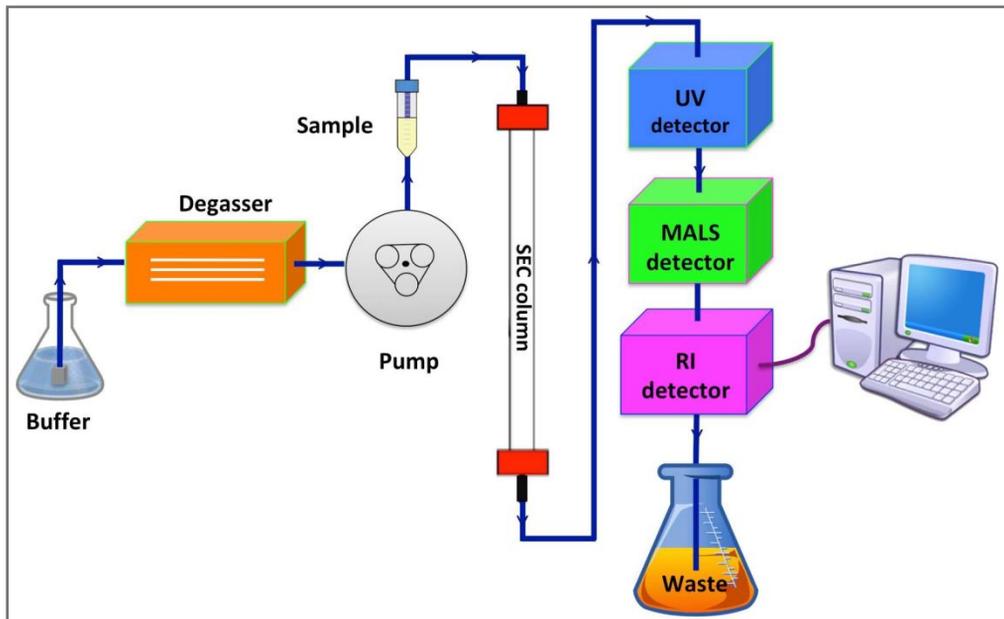


Fig.34: Schematic representation of a SEC-MALLS system. The sample passes through SEC column and is separated on the basis of their hydrodynamic volume. Then it passes through a series of three detectors, UV, MALLS and RI detector. The output is calculated by a software provided by manufacturer.

For SEC-MALLS, 100 $\mu$ l of sample (1-3mg/ml concentration) was injected at a flow rate of 0.75ml/min on a Superdex 75 10/300 column (GE healthcare) in corresponding gel filtration buffer (Refer to table Y). Elution from the column was followed by a UV-Vis Spectrophotometer, a RID-20A refractive index detector (Shimadzu), aMiniDawn TREOS detector (Wyatt Technology) and the data processing was done with ASTRA 6.1 (Wyatt Technology).

### ***3.11 Crystallization, and diffraction data collection:***

#### ***3.11.1 Crystallization:***

Protein molecules are too small to be visualized under microscope. Therefore, determination of the structure of a protein at atomic resolution requires a specialized technique. The most powerful technique for determination of a protein structure is X-ray crystallography. A crystal is a 3-dimensionally ordered periodic arrangement of molecules. Protein crystals can be grown by slowly decreasing solubility of the protein. In theory, the proteins can form crystals when the protein-dissolved solution is brought to a supersaturated state, which provides thermodynamic driving force for crystallization. Crystallization depends on protein concentration as well as type of precipitant used. Initially, protein solution is in an under saturated state where no crystals can be formed. The protein concentration is then increased over its solubility curve by different crystallization methods as seen in the phase diagram during the crystallization process, reaching the supersaturated state (Fig. 35). At this state, the system is not at equilibrium and thermodynamically driven to a new equilibrium situation with a new minimized free energy. Particular interactions can happen between individual molecules, leading to formation of aggregates, known as nucleation. Under suitable conditions, these aggregates can reach a critical size, forming stable nuclei. The next step is known as the crystal growth resulting from lowering protein concentration through nuclei formation in the nucleation to another phase shown in diagram as metastable zone. In this range, nucleation does not spontaneously occur and the stable nuclei will play a role as surface suitable for crystal growth, putatively leading to diffracting crystals in case suitable conditions are observed. The science of protein crystallization is still elusive but one of the most important factor for crystallization is considered to be the level of purity of the protein. Presence of other molecules are highly undesirable and also all protein molecules should be conformationally identical.

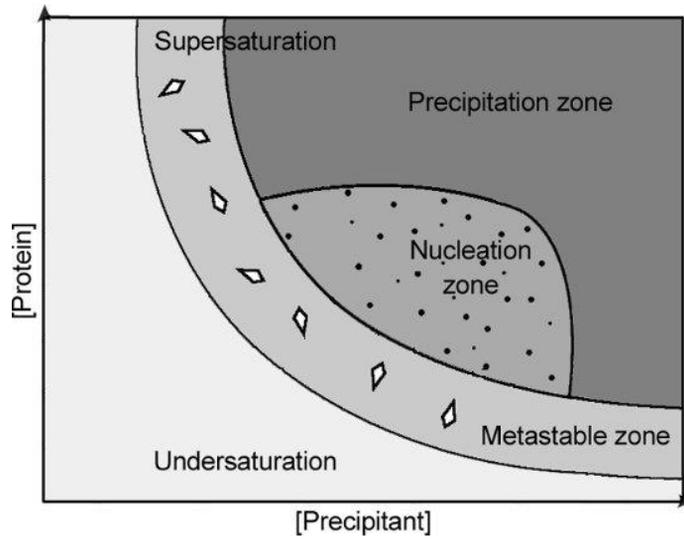


Fig.35: Phase diagram of crystallization

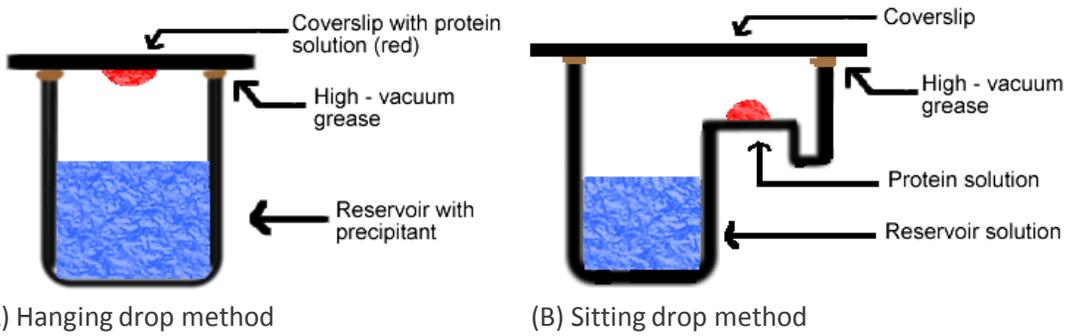


Fig.36: Vapour diffusion techniques for crystallization (A) Hanging drop method (B) Sitting drop method

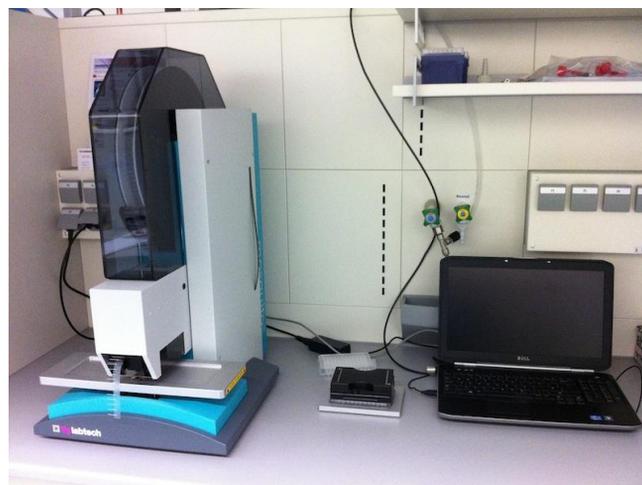


Fig.37: Mosquito crystallization robot from TTPLabtech

Obtaining diffracting protein crystals is the hardest bottleneck in protein crystallography. Protein crystallization is a trial and error based technique, where protein is slowly precipitated from its solution. There are four methods for protein crystallization:

(Giegé and Mikol, 1989, Chayenet *et al.*, 1992).

- Batch method
- Vapour diffusion method
- Dialysis
- Free interface diffusion method

Here, only vapour diffusion method was used. Two modes of vapour diffusion have been used: Sitting drop method and Hanging drop method (Fig. 36). The principle behind this method is that the protein sample is mixed with precipitant normally in the ratio of 1:1 and equilibrated against a reservoir containing precipitant. The system is then carefully sealed to make a closed environment. In the drop, the concentration of precipitant is lower than that in the reservoir, resulting in a higher water concentration in the sample drop. In the equilibrium system, the water vapour will leave the sample drop and end up in the reservoir, leading to an increase in the concentration of both protein and precipitant in the sample drop to reach a level where the crystallization process can occur provided that optimal conditions are met.

For sitting drop vapour diffusion, 150nl of protein was mixed with equal volume of precipitant and equilibrated against 80µl of precipitant solution. Sitting drop method was used for preliminary screening with commercially available crystallization kits with Mosquito crystallization robot (TTPlabtech) (Fig. 37).

When crystals appeared in primary sitting drop crystallization trials, hanging drop method was used to optimize the conditions further. For optimization of crystallization condition, a gradient of the precipitant was used in a range of pH. This method was performed by mixing 1µl of protein with 1µl of precipitant on siliconized glass coverslip. This coverslip was placed upside down on a well that is partly filled with the reservoir solution (1ml). The chamber is sealed by applying grease on the periphery before placing the cover-slip.

### ***3.11.2 Diffraction data collection:***

A collimated, monochromatic and high energy X-ray beam is used for diffraction experiment from protein crystals. X-ray sources can be broadly classified into two groups, home sources and synchrotrons. Home sources are generally equipped with a micro focus sealed tube or rotating anode generator with a high end optics for higher brilliance and a compatible

electronic 2D detector (Image plate/area detectors/CCDs and very recently the hybrid pixel detectors). Data collection is a crucial experimental step for structure determination. In X-ray diffraction, there are three methods of data collection, namely crystal rotation method, Laue method (stagnant crystal or rotating crystal in polychromatic wavelength), and Debye-Scherrer method (powder diffraction). Crystal rotation method (or oscillation) is mostly used for macromolecular crystal data collection. The data collection strategy depends on the crystal size, cell parameters, crystal stability and the nature of diffraction pattern. X-ray diffraction data were collected at Soleil Synchrotron, France. The crystals were rotated through 360° with 0.1° rotation per frame, collecting 3600 images.

Prior to data collection, the crystals were quick-soaked in cryo-protectant solutions containing 15% (v/v) and then 30% ethylene glycol or glycerol in corresponding well solutions and flash-frozen in liquid nitrogen to collect the data set.

When a monochromatic collimated beam like X-ray passes through an object, such as a macromolecule, the rays are scattered in every direction by each electron present in the object. The magnitude of this diffraction is proportional to the size of electron cloud of that atom. This is the underlying principle of a diffraction experiment. The relation between diffraction spots and the internal symmetry of crystals was described by W. L. Bragg in 1913. It is known as Bragg's law.

$$n\lambda = 2d\sin\theta$$

where  $n$  is an integer,  $\lambda$  is the X-ray wavelength,  $d$  is the interplanar distance in Miller indices ( $h, k, l$ ) and  $\theta$  is the angle of reflection. When a family of crystal planes satisfy Bragg's law, diffraction will take place. The resulting wave from that family of planes will add up (constructive interference) to produce a spot on the detector and this resulting wave is called structure factor. Structure factor is denoted by  $|F|_{hkl}$  where  $hkl$  is the family of planes. As structure factor is a wave, it has two parts, an amplitude,  $|F|$  and a phase,  $\phi$ .

Structure factor can be expressed in terms of electron density:

$$F(hkl) = V \int_{x=0}^1 \int_{y=0}^1 \int_{z=0}^1 \rho(xyz) \exp[2\pi i(hx + ky + lz)] dx dy dz$$

The goal of the diffraction experiment is to calculate the electron density  $\rho$  at every position,  $x, y, z$  of the unit cell. A crystal can be considered as a continuous electron density function

$\rho(x,y,z)$  reaching maxima at the atomic centres. The general electron density function can be expressed as a Fourier transform of  $F(hkl)$ :

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l F(hkl) \exp[-2\pi i(hx + ky + lz)]$$

As  $F = |F| \exp[i\alpha]$ , the equation can be re written as,

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)]$$

For the positive half  $0 \leq x, y, z < \infty$ , cosine part is positive and sine part is negative and for  $0 \geq x, y, z > -\infty$ , cosine part remains positive but sine part becomes positive. The summation over  $-\infty \geq x, y, z > \infty$ , the imaginary part containing sine is cancelled. Assuming Friedel's law is true the resulting electron density becomes

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \cos[2\pi(hx + ky + lz) - \alpha(hkl)]$$

So, to determine the electron density, information about two parameters are required, structure factor and phase. The structure factor can be calculated from the intensities of the reflection spots, as intensity of any given reflection ( $hkl$ ) is related to the structure factor  $F_{hkl}$

$$I(hkl) \sim |F(hkl)|^2$$

But the phase information is lost, this is known as the "Phase problem" of crystallography.

### 3.12 Structure solution:

#### 3.12.1 Data processing:

A carefully collected full data set has been processed to find out the structure factors to solve the structure. It is a sequential process and following steps are involved. An expert has to supply a few input parameters (detector position, direct beam position, detector type, goniometer type etc) to initiate the process and it requires human judgment to opt between several most possible results. The basic frame work of data processing is described below.

- Indexing (Determination of crystal symmetry or point group)
- Integration
- Scaling

Peak search in auto-indexing is conducted to determine the unit cell parameters, lattice type. Best fitted values are then refined against crystal and detector parameters. This requires fitting between the calculated and observed spot position. Integration of images is implemented in two ways, either by 2D profile fitting (Diamond, 1969, Rossmann, 1979) or 3D profile fitting (Kabsch, 1988). Different correction factors are applied during scaling regarding radiation damage, adsorption, polarization and Lorentz factor etc. Scaling also considers averaging the symmetry related reflection.

In this thesis, XDS (Kabsch, 1993) was used solely as a data processing software. During data processing, it goes through several steps like

**XYCORR:** for determination of correction table, beam centre and positions

**INIT:** for detector gain and background detection

**COLSPOT:** for identification of strong spots

**IDXREF:** for determination of cell parameters and lattice type from spots found in the COLSPOT step

**DEFPIX** and **INTEGRATE:** for detector information to apply in image integration

**CORRECT:** for scaling symmetry related reflection and apply different correction parameters. Several parameters ( $R_{\text{merge}}$ , CC1/2, signal to noise ratio, redundancy) are used to check the data quality (Weiss, 2001).

The **POINTLESS** program from the CCP4 suite can be used to determine the presence of screw axis.

A statistical method is used to measure the quality of diffraction (an agreement between multiple symmetry related reflection) which is known as  $R_{\text{merge}}$ . The quantity can be mathematically expressed for N number of observations:

$$R_{\text{merge}} = \frac{\sum_{hkl} \sum_{j=1}^N ||F_{hkl}| - |F_{hkl}(j)||}{\sum_{hkl} \sum_{j=1}^N |F_{hkl}(j)|}$$

where,  $\overline{|F_{hkl}|}$  is the mean value of the structure factor amplitude, and  $|F_{hkl}(j)|$  are the individual measurement of each symmetrically related structure factor.  $R_{\text{merge}}$  is dependent on the

redundancy (Weiss and Hilgenfeld, 1997; Diederichs and Karplus, 1997) thus can be replaced by a true data quality indicator  $R_{\text{meas}}$  or  $R_{\text{r.i.m.}}$ .

Signal to noise ratio as calculated by the  $I/\sigma(I)$  is another indicator of the data quality. The redundancy and completeness are two very important parameters to be considered during data processing. Recently, a new parameter to check data quality, known as CC1/2 has been introduced by Karplus&Diederichs (2012) and Evans (2011). They have suggested that the Pearson correlation coefficient of two half data sets (each derived by averaging half of the observations of a given reflection) must be a better measure of data quality over  $R_{\text{merge}}$ .

### **3.12.2 Space group determination:**

A crystal can be considered as a three dimensional lattice, made up by a basic building block, (the asymmetric unit) by applying symmetry operators (space group) and translation (unit cell). The asymmetric unit of the crystal is the fundamental unit of construction. It can be made up of one or more molecules that are not related by a symmetry operator of the crystal. By applying symmetry operators to the asymmetric unit content, a unit cell can be generated, these symmetry operators are determined by the space group of the crystal. There are 230 possible type of space group but biological samples have only one chiral form of monomer (e.g. proteins are formed of L-amino acids), thereby limiting the possible number of space group to 65. These 65 space groups can be further divided into 32 point groups, that define symmetry between molecules in a 2D-plane. A unit cell contains all the information required to generate a crystal as a crystal is a translationally periodic, finite assembly of unit cells.

During data processing by XDS, space group can be determined in the IDXREF step. This step takes into account each possible space group based on parameters calculated from reflections and displays possible space groups along with the penalty score (quality of fit). The most suitable space group is selected by highest symmetry and lowest penalty score.

### **3.12.3 Phasing:**

To determine experimental electron density map from a diffraction dataset, two information are required,

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \cos[2\pi(hx + ky + lz) - \alpha(hkl)]$$

Structure factor and phase. Although structure factors can be calculated from the intensity of reflections recorded during the diffraction experiment, the phase information is lost. There are four methods to extract the phase information:

1. ***Isomorphous replacement method (MIR)***: this method requires attachment of a heavy atom(s) to the protein molecule without distorting the unit cell parameters, *i.e.* the unit cell parameters of native protein and heavy atom derivative should be identical. In a nutshell, comparing isomorphous datasets collected from native crystals and heavy atom derivatives can allow the determination of experimental electron density maps.
2. ***Anomalous diffraction method***: this method exploits presence of strong anomalously scattering atoms in the protein. This method is commonly used by labeling the protein with selenomethionine and relies on the same principle as MIR except that a single crystal can be sufficient if data are collected at specific wavelengths.
3. ***Molecular replacement method***: this method can be used when there exists a known structure that has high sequence similarity and hence structural resemblance with the unknown protein.
4. ***Direct method***: This method is mostly used for small molecule crystallography and it is still under development for application on protein crystals.

### ***3.12.3.1 Data processing and Phase determination of Mei2-RRM3:***

As described above anomalous diffraction exploits wavelength dependence of anomalously scattering atoms already present in the protein. The electrons in an atom are considered to be free electrons, and the diffracted X-ray beam by these free electrons, differ from the incident beam by 180° in phase. As the atomic number keeps increasing, this does not remain true, the inner electrons surrounding the nucleus are more tightly bound than the outer electron cloud and the phase of diffracted beam does not differ from the incident beam by 180° anymore. The structure factor  $F_{hkl}$  can be expressed as a summation of atomic structure factors of all atoms present, and it can be calculated if positions (x,y,z) and elements of all atoms are known:

$$F(h, k, l) \propto \sum f_j \exp 2\pi i(hx + ky + lz)$$

Where  $f_j$  is the atomic scattering factor specific for each type of atom. Presence of anomalous scatterer divides the atomic scattering factor in two parts, a real one and an imaginary one. This is represented by an equation as:

$$f_{anom} = f + \Delta f + if'' = f' + if''$$

where,  $f_{anom}$  is the total atomic scattering factor which is summation of normal and anomalous components.  $f$  represents scattering of X-ray by a free electron, and  $f'$  is the real component of anomalous scattering that is in-phase with  $f$  and  $f''$  is the imaginary part of anomalous scattering that is out of phase from  $f$  by  $90^\circ$ . The two anomalous parts,  $f'$  and  $f''$  are dependent only on the wavelength used for data collection. Presence of anomalously scattering atom ( $if''$ ) renders Fridel's law false.

$$|F_{h, k, l}| \neq |F_{-h, -k, -l}|$$

The difference between the Bijvoet pairs can be exploited to locate the position of the anomalously scattering atoms by a Patterson map. Locating the anomalous scatterers will help to determine phase. The structure factor of a protein containing anomalous scatterer can be explained as:

$$F_{PH} = F_P + F_H$$

Where,  $F_{PH}$  is the structure factor of protein containing anomalously scattering atom,  $F_P$  is for native protein and  $F_H$  for the heavy atom or anomalous scatterer.

The structure of Mei2-RRM3 was solved by Se-SAD. The data processing to structure refinement strategy for Mei2-RRM3 has been described by the flowchart. Data processing for both native and Se-labelled protein, was performed by XDS. XDSCONV was used to convert the hkl file to mtz file format. The coordinate of Se atoms were located using SHELXD. SHELXD is a robust program for locating substructures in anomalous Patterson map. Substructure is the coordinates of a small set of atoms of a crystal structure. The output from SHELXD included a coordinate file of Se atoms. This information was used by Autosol, a pipeline from Phenix, to calculate experimental phases, electron density maps and then rebuild the model. Autosol is a compilation of multiple programmes. The first step is Xtriage, that checks for problems like, twinning, translational pseudo-symmetry or weak anomalous signal. Phasing is performed by Phaser from position of anomalous scatterer while density modification and model building are performed by Resolve. The final model was refined using Refmac 5 from CCP4 suite. This model was used as an input for molecular replacement

by Molrep, on the native data set collected at higher resolution. Successive rounds of model building and refinement led to the final structure described in the results section.

### ***3.12.3.2 Data processing and phasing of Erh1:***

Data processing for Erh1 was also performed by XDS and phase was determined by molecular replacement. Molecular replacement is a phasing technique which uses the structure of a known or homologous protein to determine structure of an unknown protein. The basis of this method is that two proteins highly homologous at the sequence level, should have a similar three-dimensional structure.

Molecular replacement method utilizes a reference structure and places it within the fine grid of the target structure. Then it orients the reference structure in all possible orientations within the target system. Structure factors from the model in each orientation are compared with the observed structure factors of the real model and indicated by the R factor or reliability factor. For a successful Molecular replacement, it is important to place the model molecule in correct orientation and position of the unit cell. This process can be divided in two steps: rotation and translation. Molecular replacement method uses Patterson function to calculate the atomic distances. *Self-Patterson vectors* denote the distance between the intra-molecule atoms and tend to gather around the origin, they are used for the rotation function. *Cross Patterson vectors* denote the distance between the inter-molecule atoms and found away from the origin. They are used for the translation function. Self-vectors of a Patterson function from both oriented and target system are calculated and compared to reach the best agreement.

The crystal structure of *H. sapiens* ERH (PDB entry 1W9G) was used as a template (30% sequence identity with *S. pombe* Erh1) for molecular replacement trials. Initial solution was subjected to rigid body refinement followed by several cycles of restrained refinement in Refmac5 (Murshudovet *al.*, 1997) until an interpretable electron density map was generated. COOT (Emsley & Cowtan, 2004) was used to visualize the electron density map and fitting of amino acids according to the sequence and stereochemistry, commonly known as model building. Alternate cycles of model building and restrain refinement improves the quality of model.  $R$  and  $R_{\text{free}}$  (Brünger, 1992) are the best judge for the progress of refinement. Steady decrease and convergence of  $R$  and  $R_{\text{free}}$  was achieved successfully during the refinement process.

### 3.13 Model building and Structure refinement:

Model building and refinement are the last stages of the structure determination process where an atomic model is fitted in the electron density. Successive rounds of building and fitting of model in real space following refinement in reciprocal space helps to determine the correct structure. Different methods of phasing, like molecular replacement or Se-SAD, allow to calculate structure factors. But these calculated structure factors ( $F_C$ ) are often different than the observed structure factors ( $F_O$ ). The difference between the calculated and observed structure factors is calculated by a statistical parameter:

$$R_{\text{factor}} = \frac{\sum_{hkl} ||F_o| - |F_c||}{\sum_{hkl} |F_o|} \times 100\%$$

Where  $F_O$  is the observed structure factor and  $F_C$  is the calculated structure factor. There are methods to lower the  $R_{\text{factor}}$ . Unfortunately, there are some demerits using  $R_{\text{factor}}$ , like over-refinement that can lead to artificially low  $R_{\text{factor}}$ . Therefore, a statistically independent measure,  $R_{\text{free}}$  is used to evaluate the structure (Brünger, 1992, Kleywegt and Brünger, 1996).  $R_{\text{free}}$  is calculated as  $R_{\text{factor}}$  with the only exception that it is calculated from  $\sim 5\%$  of the all unique reflections that are not used in refinement.

The refinement process tries to optimize the agreement between atomic model and observed data. After fitting the main chain backbone, individual residues are fitted in electron density by local real space refinement. Once the entire model is built, global refinement is used to correct potential problems like stereo-chemical and conformational problems encountered during real space model building. The stereo-chemical and conformational restrains can be applied by two methods:

1. Rigid body refinement: bond lengths and bond angles are considered to be rigid and only dihedral angles can be varied.
2. Restrained refinement: bond lengths and bond angles can vary around a standard value.

Structures reported in this thesis were refined using phenix.refine (Adams et al., 2010) and autoBUSTER (Bricogne et al., 2016) and model building was done with COOT.

### 3.14 MATERIALS

Table S1. Summary of different strains with their genotypes and purposes used in the thesis

Strains	Genotype	Usage	Sources
<i>E.coli</i>			
XL1 - blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacIq ZΔM15 Tn10 (Tet<sup>R</sup>)]</i>	Cloning and plasmid preparation	Stratagene
BL21 (DE) Gold	<i>B F<sup>-</sup> ompT hsdS(rB<sup>-</sup> mB<sup>-</sup>) dcm<sup>+</sup> Tet<sup>R</sup> gal λ(DE3) endA Hte</i>	Protein expression	Agilent Technologies
BL21 (DE) Codon plus RIL	<i>B F<sup>-</sup> ompT hsdS(rB<sup>-</sup> mB<sup>-</sup>) dcm<sup>+</sup> Tet<sup>R</sup> gal λ(DE3) endA Hte [argU ileY leuW Cam<sup>R</sup>]</i>	Protein expression	Agilent Technologies

Table S2. Oligonucleotides and plasmids used to over-express *S. cerevisiae* proteins in *E. coli*

<i>S.cerevisiae</i> genes (Domain boundary)	Name	Sequence	Enzyme	Plasmid & tag	Plasmid generated
Not1 1348-2093	oMG416	TATACCATGGGCCACCACCATCAT CACCATAATCCTCAAGGTGGGATT GC	NcoI	pET28b	pMG858
	oMG417	GTGCTCGAGTTATTGGTCATCTTG TTCCTGG	XhoI	6x-His	
Not1 1565-2093	oMG418	TATACCATGGGCCACCACCATCAT CACCATAAAACCACAAGAACAGA AAAGC	NcoI	pET28b	pMG859
	oMG417	GTGCTCGAGTTATTGGTCATCTTG TTCCTGG	XhoI	6x-His	
Not1 1071-1565	oMG513	TATAGAATTCAATCCTTTTA ACAACCTTAC	NcoI	pET28b HisZZ	pMG907

	oMG516	ATATGCGGCCGCTTATTTGGTACC TTTCCTCACTGGC	XhoI	6x- HisZZ	
Not1 1071-1282	oMG513	TATAGAATTCAATCCTTTTA ACAACCTTAC	NcoI	pET28b HisZZ	pMG908
	oMG515	ATATGCGGCCGCTTAAAACGGAAT AACATCCAAGTTTGG	XhoI	6x- HisZZ	
Not1 1343-1565	oMG514	TATAGAATTCGGAGTGAATGTCCC AAATCC	NcoI	pET28b HisZZ	pMG909
	oMG516	ATATGCGGCCGCTTATTTGGTACC TTTCCTCACTGGC	XhoI	6x- HisZZ	
Pho92 1-306	oMG419	CCTGGGATCCATGAATCAAATCTG GTCTACAGG	BamHI	pGex- 6P-1	pMG860
	oMG420	GCCGCTCGAGCTATTCATACGTCT CATCCAAG	XhoI	GST	
Pho92 1-142	oMG419	CCTGGGATCCATGAATCAAATCTG GTCTACAGG	BamHI	pGex- 6P-1	pMG862
	oMG422	GCCGCTCGAGCTAAGCACTTTTTC TTTTATTAATTTTCATT	XhoI	GST	
Pho92 143-306	oMG421	CCTGGGATCCATGGCAATTATTCC CCCTTGG	BamHI	pGex- 6P-1	pMG861
	oMG420	GCCGCTCGAGCTATTCATACGTCT CATCCAAG	XhoI	GST	
Pho92 1-69	oMG419	CCTGGGATCCATGAATCAAATCTG GTCTACAGG	BamHI	pGex- 6P-1	pMG910
	oMG532	TATACTCGAGCTATTCTTTCTTGTT AATATCATTTAAAGC	XhoI	GST	
Pho92 70-143	oMG533	CCTGGGATCCAATAAAGAAGAGA TCACTCATGAAA	BamHI	pGex- 6P-1	pMG911
	oMG534	TATACTCGAGCTATGCAGCACTTT TTCTTTTATTAATTTCA	XhoI	GST	
Pho92 21-44	oMG535	CCTGGGATCCAATAAGAGAA ATGACAGGAC	BamHI	pGex- 6P-1	pMG912

	oMG536	TATACTCGAGCTATCTCTCCAAGG AATGGATCAACCC	XhoI	GST	
--	--------	---	------	-----	--

**Table S3. Oligonucleotides and plasmids used to over-express *S. pombe* proteins in *E. coli***

<i>S. pombe</i> genes (Domain boundary)	Name	Sequence	Enzyme	Plasmid & tag	Plasmid generated
Mmi1 1-488	oMG425	TATACCATGGGCCACCACCATCAT CACCATTCAAACA CAACTTCTC TACTTCTAGG	NcoI	pET28b	pMG863
	oMG426	GGTGCTCGAGTCAACGGTCTCTTC CAATTCGC	XhoI	6x-His	
Mmi1 322-488	oMG427	TATACCATGGGCCACCACCATCAT CACCATAGCGAAC ATAGAAATGA GAAGGGGG	NcoI	pET28b	pMG864
	oMG426	GGTGCTCGAGTCAACGGTCTCTTC CAATTCGC	XhoI	6x-His	
Mmi1 1-350	oMG425	TATACCATGGGCCACCACCATCAT CACCATTCAAACA CAACTTCTC TACTTCTAGG	NcoI	pET28b	pMG866
	oMG451	GAGACTCGAGTCAGGAGCGGTGT GAAATTCCATTTTCG	XhoI	6x-His	
Mmi1 1-322	oMG425	TATACCATGGGCCACCACCATCAT CACCATTCAAACA CAACTTCTC TACTTCTAGG	NcoI	pET28b	pMG865
	oMG452	GGTGCTCGAGTCAGCTTGCTCTGG AAAAATTTAATGG	XhoI	6x-His	
Mmi1 1-75	oMG506	TATAGAATTCATGCACCACCATCA TCACCATTCAAACA CAACTTCTC	EcoRI	pET28b	pMG870
	oMG510	TATACTCGAGTCATGACTCTGGTG CAGGTCTTTTC	XhoI	6x-His	

Mmi1 1-178	oMG506	TATAGAATTCATGCACCACCATCA TCACCATTCAAACA CAACTTCTC	EcoRI	pET28b	pMG869
	oMG509	TATACTCGAGTCAACTAAGTGTGC GTCGCTTGGGAGG	XhoI	6x-His	
Mmi1 1-240	oMG506	TATAGAATTCATGCACCACCATCA TCACCATTCAAACA CAACTTCTC	EcoRI	pET28b	pMG868
	oMG508	TATACTCGAGTCAACGAGTGTCTT CATGAGAAAGTTGG	XhoI	6x-His	
Mmi1 1-294	oMG506	TATAGAATTCATGCACCACCATCA TCACCATTCAAACA CAACTTCTC	EcoRI	pET28b	pMG867
	oMG507	TATACTCGAGTCAATTAGAAAGCA GGTAAGAAG	XhoI	6x-His	
Mmi1 95-122	oMG605	TATAGGATCCGGTAAATATGATTT TAGCAGGC	EcoRI	pET28b	pMG915
	oMG606	TATACTCGAGTCAAGACTCACGAC GAAGG	XhoI	6x-His	
Mmi1 60-178	oMG539	TATAGAATTCATGCACCACCATCA TCACCATGGTTCAAACA ACTTTTC TTCACC	EcoRI	pET28b	pMG913
	oMG509	TATACTCGAGTCAACTAAGTGTGC GTCGCTTGGGAGG	XhoI	6x-His	
Mmi1 78-178	oMG540	TATA GAATTC ATG CACCACCATC ATCACCA T GCTCCTATCGGTAGA AGATTGATGG	EcoRI	pET28b	pMG914
	oMG509	TATACTCGAGTCAACTAAGTGTGC GTCGCTTGGGAGG	XhoI	6x-His	
Mei2 1-750	oMG428	CCTGCCCCGGGTATGATTATGGAAA CCGAATCACC	SmaI	pGex-6P-1	pMG916
	oMG429	GCCGGCGGCCGCTCAACATTTGCT TGCAGTTGG	NotI	GST	
Mei2 184-372	oMG430	CCTGGGATCCATGTCTGACGATAT AGATATATTT TCTCATGC	SmaI	pGex-6P-1	pMG917

	oMG431	GCCGGCGGCCGCTCAGCTAACTGA ATCAGCCATTGC	NotI	GST	
Mei2 579-750	oMG432	CCTGCCCCGGGTTTCAGAT AGAAATTCTG TCGATTATGC	SmaI	pGex-6P-1	pMG918
	oMG429	GCCGGCGGCCGCTCAACATTTGCT TGCAGTTGG	NotI	GST	
Mei2 1-429	oMG428	CCTGCCCCGGGTATGATTATGGAAA CCGAATCACC	SmaI	pGex-6P-1	pMG919
	oMG502	TATAGCGGCCGCTCATCCAAAGTT ATTGACATTCC	NotI	GST	
Mei2 429-750	oMG503	CTCTGGATCC <u>G</u> ATCC GTTCCCTTAGG	BamHI	pGex-6P-1	pMG920
	oMG429	GCCGGCGGCCGCTCAACATTTGCT TGCAGTTGG	NotI	GST	
Erh1 1-102	oMG511	TATAGGATCCAGCCCCCACCCGC CG	BamHI	GST	pMG921
	oMG512	GCGGCTCGAGTTACGGAATCTGAC GAGCCGC	XhoI	pGex-6P-1	

**Table S4: Oligonucleotides and plasmids used to over-express proteins in *E. coli***

Protein	Residue	Primer	Primer sequence (underlined faces are the restriction site)
Mei2- RRM3	S694E	oMG490	5'TTCCGAAACGAGTGTGTTATGGACGAGAACCCTGCTTACC3'
		oMG491	5' ATAACACACTCGTTTCGGAATTTTTCAATCAGCCTGTCTT 3'
Mei2- RRM3	Y642A	oMG488	5' GCAATGTAGGAGCTGCGTTTATAAACTTTATTGAACC 3'
		oMG489	5' GTTTATAAACGCAGCTCCTACATTGCATTTATTAACAAA 3'
Mei2- RRM3	F644A	oMG486	5' GTAGGATATGCGGCTATAAACTTTATTGAACCTCAATCTA 3'
		oMG487	5' AAGTTTATAGCCGCATATCCTACATTGCATTTATTAACAA 3'
Erh1	I11R/L1 3R	oMG629	5'CGAATCTCATATCAGGCTGAGGATTCAGCAAGGTTCTGACC CT AAAACC 3'
		oMG630	5'CCTTGCTGAATCCTCAGCCTGATATGAGATTCGGCGGGTGG GGGC 3'

## Media

**Luria-Bertani broth (LB):** Tryptone (Bacto) 10g/l, yeast extract 5g/l and NaCl 5g/l, pH 7.5 (adjusted by NaOH).

**2YT:** Tryptone (Bacto) 16g/l, yeast extract 10g/l and NaCl 5g/l, pH 7.0 (adjusted by NaOH).

**Terrific broth auto inducible (TBAD):** Tryptone (Bacto) 12g/l, yeast extract 24g/l and MgSO<sub>4</sub> 0.15g/l, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 3.3g/l, KH<sub>2</sub>PO<sub>4</sub> 6.5g/l, Na<sub>2</sub>HPO<sub>4</sub> 7.1g/l, glucose 0.5g/l, Alpha Lactose 2.0g/l, pH 7.0 (adjusted by NaOH).

## Buffers

### Different buffers used for different protein purifications from different organisms

Table S5

#### Mei2-RRM3

Step	Buffer	Composition
Cell Lysis	Lysis Buffer	20mM Tris HCl pH7.5, 200 mM NaCl, 5mM β-mercaptoethanol
GST	Wash	Lysis Buffer
	Elution	Lysis Buffer+20mM GSH
Heparin	Buffer A	50mM Tris HCl pH7.5, 50 mM NaCl, 5mM β-mercaptoethanol
	Buffer B	50mM Tris HCl pH7.5, 1 M NaCl, 5mM β-mercaptoethanol
Gel Filtration	Gel filtration Buffer	20mM Tris HCl pH7.5, 200 mM NaCl, 5mM β-mercaptoethanol

#### Erh1

Cell Lysis	Lysis Buffer	20mM Tris HCl pH7.5, 200 mM NaCl, 5mM β-mercaptoethanol
GST	Wash	Lysis Buffer
	Elution	Lysis Buffer+20mM GSH
Gel Filtration	Gel filtration Buffer	20mM Tris HCl pH7.5, 200 mM NaCl, 5mM β-mercaptoethanol

#### Mmi1

Cell Lysis	Lysis Buffer	20mM Tris HCl pH7.5, 200 mM NaCl, 5mM β-mercaptoethanol
GST	Wash	Lysis Buffer
	Elution	Lysis Buffer+20mM GSH
Heparin	Buffer A	50mM Citrate pH6.0, 50 mM NaCl, 5mM β-mercaptoethanol
	Buffer B	50mM Citrate pH6.0, 1 M NaCl, 5mM β-mercaptoethanol
Gel Filtration	Gel filtration Buffer	20mM Tris HCl pH7.5, 200 mM NaCl, 5mM β-mercaptoethanol

#### Not1

Cell Lysis	Lysis Buffer	20mM Tris HCl pH8.0, 200 mM NaCl, 5mM $\beta$ -mercaptoethanol
Ni-NTA	Wash	Lysis Buffer + 20mM Imidazole
	Elution	Lysis Buffer + 400 mM Imidazole
Gel Filtration	Gel filtration Buffer	20mM Tris HCl pH8.0, 100 mM KCl, 5mM $\beta$ -mercaptoethanol

## **Pho92**

Cell Lysis	Lysis Buffer	20mM Tris HCl pH7.5, 200 mM NaCl, 5mM $\beta$ -mercaptoethanol
GST	Wash	Lysis Buffer
	Elution	Lysis Buffer+20mM GSH
Gel Filtration	Gel filtration Buffer	20mM Tris HCl pH7.5, 200 mM NaCl, 5mM $\beta$ -mercaptoethanol

# RESULTS



## CHAPTER IV

Structural and functional  
characterization of Mmi1, Mei2 and  
Erh1



#### ***4.1 S. pombe Mei2, the Master regulator of meiosis:***

Bioinformatics analyses were performed using BLAST against non-redundant database to identify *S. pombe* Mei2 domains. This indicated that Mei2 had two closely spaced RRM domains RRM1(194-265) and RRM2(287-362) in its N-terminal half and a third RRM domain RRM3(596-692) in its C-terminal half (Fig. 38a). From Yamamoto group's work, it is known that the RRM3 is an essential domain to retain the protein's activity. The domain boundaries for Mei2 constructs were determined on the basis of multiple sequence alignment (Fig. 38b, c) and secondary structure prediction from HCA diagram (Callebaut et al; 1997) and the following constructs were designed and tested: Mei2-FL, RRM1-2 (184-372), RRM3 (579-750).

##### ***4.1.1 Cloning:***

Based on the domain predictions described above, the DNA fragments encoding for the corresponding domains have been amplified using oligonucleotides listed in Table S3 and cloned into pGEX-6P1 vector using adequate restriction enzymes as described in the material and method section. This should allow the purification of GST-tagged versions of these various Mei2 domains. All PCR amplified products were verified by running on 1% agarose gel. DNA plasmids were extracted from XL1-Blue colonies visible on Petri dishes containing LB-agar media supplemented with the ampicillin antibiotic (as the pGEX-6P1 vector contains an ampicillin resistance gene) after overnight incubation at 37 °C. These plasmids were digested with the restriction enzymes used for cloning and the products were analysed on a 1% agarose gel so as to specifically select plasmids containing an insert of the desired size for sequencing.

##### ***4.1.2 Expression assays:***

From these plasmids, expression assays were performed at 18 °C and 37 °C for the following constructs of Mei2: FL (full length), RRM1-2 (184-372), RRM3 (579-750) in TBAI media and either BL21 Codon+ or Gold *E. coli* cells. As all the constructs are GST-tagged, a Glutathione-sepharose resin was used to perform small-scale purification. All fractions collected during this procedure were analysed on 12% SDS/PAGE. This revealed that for GST-Mei2-FL protein, no band corresponding to the correct molecular weight was observed in the elution fraction from the Glutathione-sepharose, indicating that the protein is most

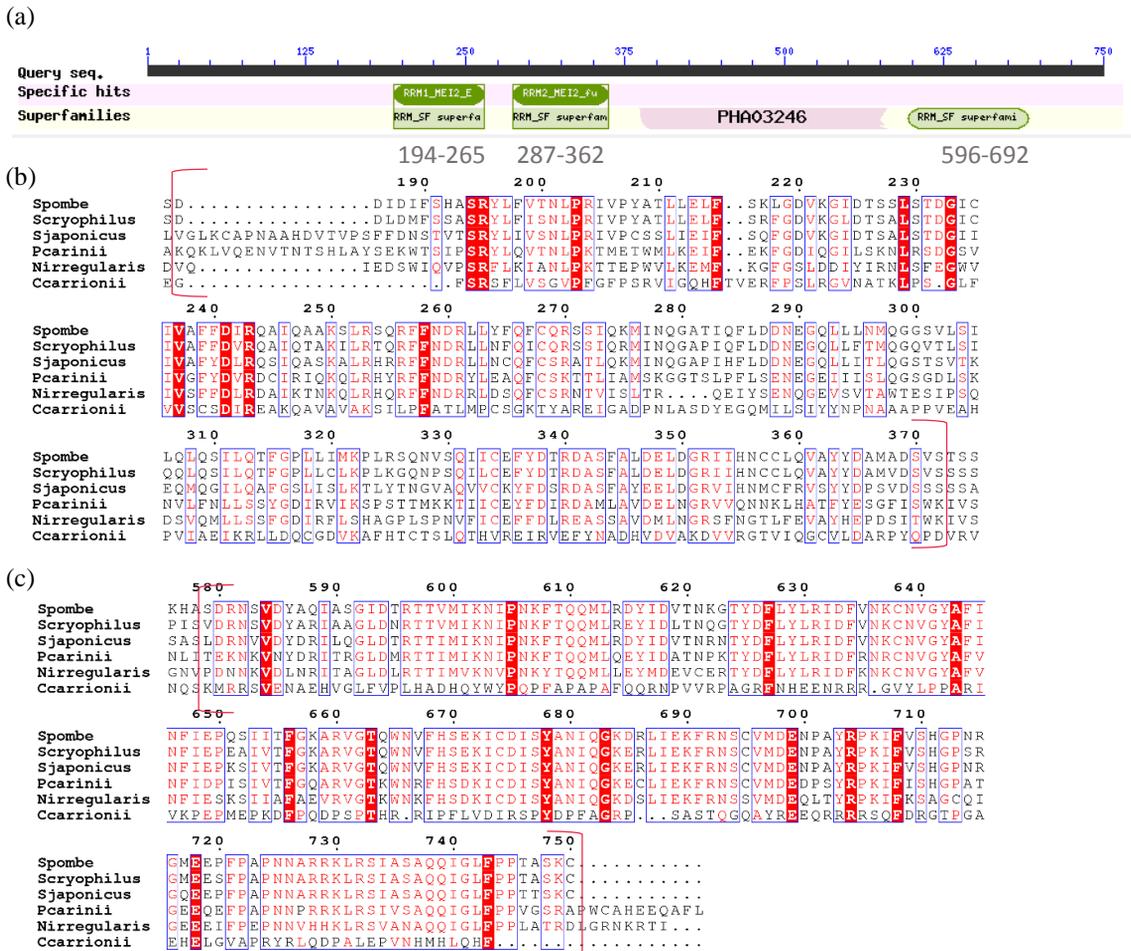


Fig. 38: Bio-informatics analysis of *S. pombe* Mei2 protein (a) Domain organization of Mei2 from CDD (Conserved domain database-NCBI). Multiple sequence alignment containing (b) RRM1-2 and (c) RRM3. The domain boundaries are highlighted with red bracket. Figure B & C generated using Esprint server

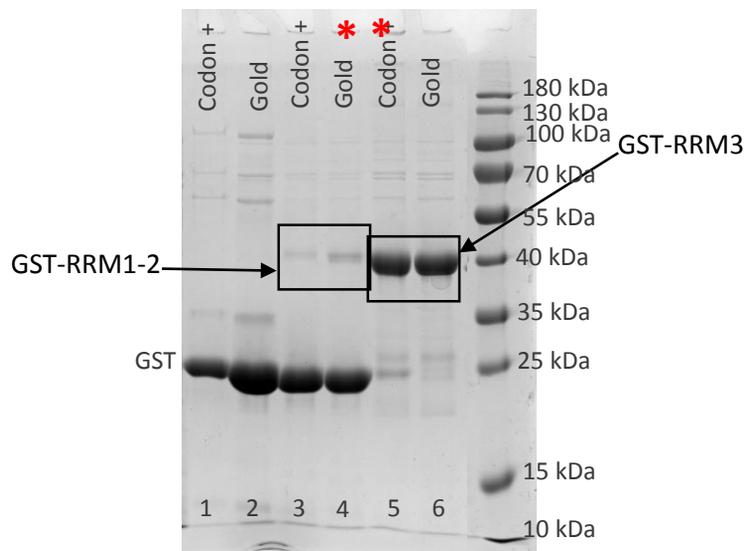


Fig. 39: 12% SDS-PAGE gel for expression check of GST-Mei2 constructs at 18°C. Lanes 1, 2 correspond to Mei2-FL (~100kDa), lanes 3, 4 to RRM1-2 (~45kDa), lanes 5, 6 to RRM3 (~43kDa). There was no detectable band corresponding to the molecular weight of Mei2-FL. Bands corresponding to the molecular weight of RRM1-2 and RRM3 are highlighted. The bands with asterisk indicate the conditions selected for large culture

probably not expressed or expressed as an insoluble protein. For GST-Mei2-RRM1-2, very faint band was observed in expression from Gold cells. Interestingly, an intense band migrating at the expected molecular weight for GST fused Mei2-RRM3 was detected in the elution fraction whether BL21 Codon+ or Gold cells were used for expression assays (Fig. 39). The probable reason behind this is the proteins were expressed but they are not stable and degraded over time, only the GST remained. From this expression assay, Gold cells and TBAI were selected to be best condition to express RRM1-2 at 18 °C and Codon+, TBAI media was selected for large scale production of RRM3.

#### **4.2 Study of Mei2-RRM1-2:**

##### **4.2.1 Protein expression and purification:**

Based on the expression assays results, Mei2-RRM1-2 was over-expressed at 18 °C using BL21 (DE) Gold *E. coli* cells and 1L of TBAI media (Terrific Broth Auto-Inducible media; ForMedium AIMTB0260) containing Ampicillin (100 µg/ml).

Mei2-RRM1-2 was purified by a two steps purification procedure consisting first of a GST affinity chromatography sets followed by a SEC. Protocol used for purification was described in section 3.4. Initial purification trial was performed following the standard buffer (20 mM Tris-HCl, pH7.5, 200 mM NaCl, 5 mM β-mercaptoethanol). During this trial protein precipitation was observed therefore, to prevent precipitation all the buffers for GST purification were supplemented with 10% Glycerol and the final gel filtration buffer with 5% Glycerol. During this purification, the protein eluted from the Glutathione-sepharose resin was incubated with 3C protease overnight and dialyses against the following buffer (20mM Tris-HCl, pH7.5, 200 mM NaCl, 10% Glycerol, 5mM β-mercaptoethanol) to remove the GST-tag from the protein of interest. SDS-PAGE analysis of the sample incubated with 3C protease reveals that the band corresponding to GST-Mei2-RRM1-2 disappears while 2 major bands migrating at the expected molecular weights of GST and Mei2-RRM1-2 appear, clearly supporting that the purified protein is indeed Mei2-RRM1-2 (Fig. 40). The cleaved GST-tag was removed upon incubation with Glutathione-sepharose resin. The Mei2-RRM1-2 cleaved from the GST-tag was not retained by the Glutathione-sepharose resin and was injected on a preparative Superdex 75 size-exclusion chromatography pre-equilibrated with gel filtration buffer (20 mM Tris-HCl, pH7.5, 200 mM NaCl, 5% Glycerol, 5 mM β- mercaptoethanol). In the resulting chromatogram, a single major peak is observed and

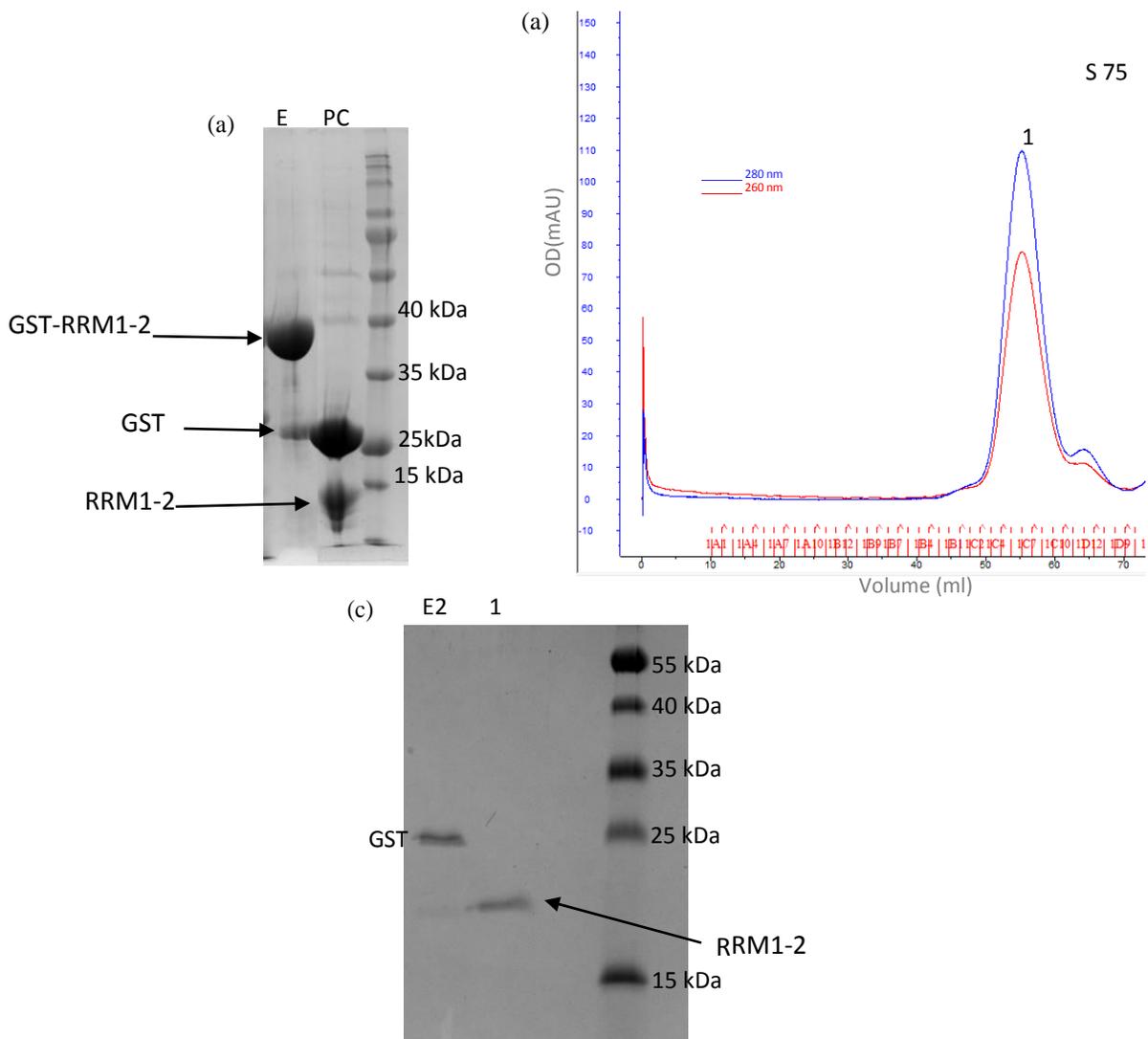


Fig. 40: Purification of Mei2-RRM1-2. (a) 12% SDS PAGE gels to check purity of protein complex after GST purification. (b) Elution profile of Mei2-RRM1-2 from S 75 gel filtration column. E: elution from Glutathione sepharose column of the GST tagged protein, PC: Post cleavage protein sample containing two bands, one corresponding to GST the other one to RRM1-2. (c) The RRM1-2 is highly pure after gel filtration. E2: GST tag eluted from second round of GST purification of the post cleavage sample, 1: refers to the peak fraction of gel filtration from (b).



Fig. 41: Crystals of Mei2 RRM1-2 grown from 1M Ammonium Sulfate, 0.1M Bis-Tris pH 5.5, 1% PEG3350, at 4 °C

analysis of its content by SDS-PAGE, reveals the presence of a protein migrating at the expected molecular weight for Mei2-RRM1-2.

#### ***4.2.2 Crystallization of Mei2-RRM1-2:***

The Mei2-RRM1-RRM2 purified fragment were subjected to crystallization trials by sitting drop vapour diffusion method, with 6.0 mg/ml protein and JCSG+ and Index screen at 4 °C. Small needle shaped crystals appeared in the following condition: 1.0M ammonium sulphate, 0.1 M Bis-Tris pH5.5, 1% PEG3350 (Fig. 41).

As these crystals were too small to be tested for diffraction at the synchrotron, thermal shift assay was performed to identify conditions resulting in higher protein thermostability. 1mM MgCl<sub>2</sub> induced a 3 °C increase in thermostability (Fig. 42). This compound was added to the gel filtration buffer during another purification performed as described above.

From the protein purified in this condition, another crystallization trial was performed with JCSG+ as well as AmSO<sub>4</sub> suite (Qiagen) as initial crystals were obtained with ammonium sulphate as precipitant. New crystals were obtained from two conditions at 4 °C: 0.1 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1 M Bis-Tris pH5.5, 17% w/v PEG 10,000 and 0.1 M MES Sodium salt pH6.5, 2.0 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 5% (w/v) PEG 400 (Fig. 43). Further optimization by the hanging drop technique and by mixing 1 µL of protein with 1 µL of crystallization condition was performed on the second condition, by making a gradient of (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (1.4 M-2.1 M) and at three different pH (100 mM MES Sodium salt, pH6.3, 6.5 & 6.8). Two crystals appeared in only one condition (1.54M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1M MES Sodium salt, pH6.3, 5% (w/v) PEG400) but these crystals did not diffract when tested on synchrotron beamlines. Further optimization of the crystallization conditions will be required in the future.

### ***4.3 Study of the RRM3 domain from Mei2:***

#### ***4.3.1 Protein expression and purification:***

Based on the results of the expression assays (4.1.2), the Mei2-RRM3 was over-expressed in BL21 (DE) Codon+ cells and in TBAI media (Terrific Broth Auto-Inducible media; ForMedium AIMTB0260), containing Ampicillin (100 µg/ml) and Chloramphenicol (25 µg/ml). After OD<sub>600</sub> reached 0.6 at 37 °C, culture was transferred to 18 °C for 16 h. To over-express selenomethionine labelled protein, 10 mL of overnight culture was used as inoculum for 1 L selenomethionine containing minimal media. Culture was grown at 37°C until OD<sub>600</sub>

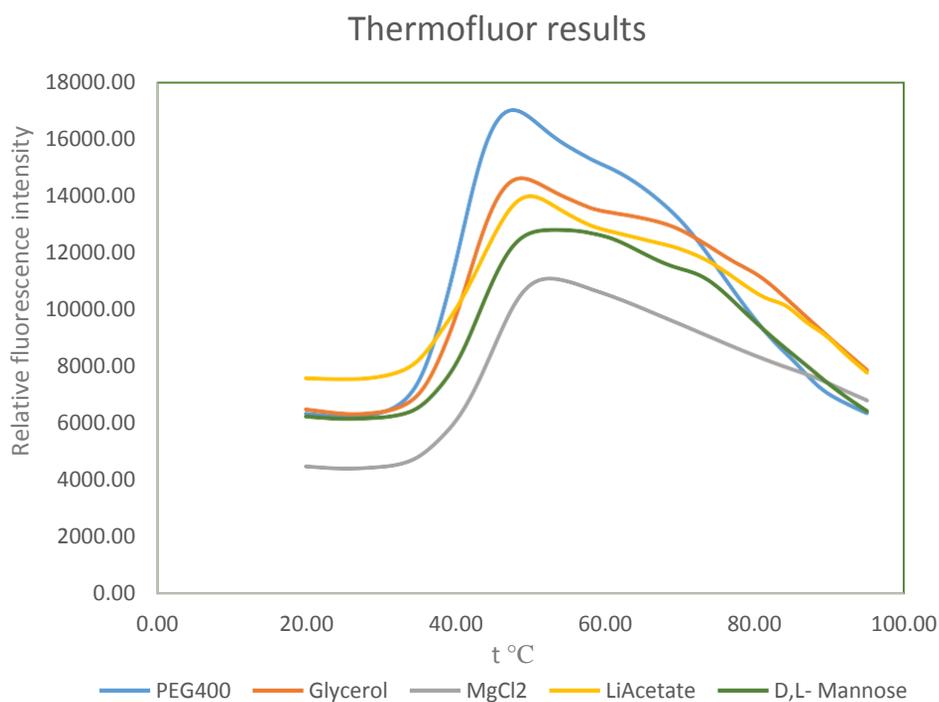


Fig. 42: Thermal denaturation assay for buffer optimization



0.1M  $(\text{NH}_4)_2\text{SO}_4$ , 0.1M Bis-Tris pH5.5, 17% w/v PEG10,000



0.1M MES Sodium salt pH6.5, 2.0M  $(\text{NH}_4)_2\text{SO}_4$ , 5% (w/v) PEG400

Fig. 43: Crystal obtained after addition of 1 mM  $\text{MgCl}_2$  in gel filtration buffer

reached 0.6 and then induced with 100  $\mu$ M IPTG and transferred to 18 °C for 16 h.

Mei2 RRM3 domain was purified in a three step purification process: affinity chromatography on Glutathione-sepharose, ion-exchange chromatography using a Heparin resin and SEC. Following elution from Glutathione-sepharose, the GST-tag was removed by 3C protease, with overnight dialysis against lysis buffer, as described for Mei2-RRM1-2 fragment. The Mei2 RRM3 domain cleaved from the GST-tag was purified on a Heparin column and eluted using a linear gradient ranging from 100% Buffer A (50 mM Tris-HCl, pH7.5, 50 mM NaCl, 5 mM 2-mercaptoethanol) to 100% Buffer B (50 mM Tris-HCl, pH7.5, 1 M NaCl, 5 mM 2-mercaptoethanol). The protein eluted from Heparin column was concentrated by Vivaspin 10 concentrator (10 kDa MWCO, Sartorius) and loaded on a HiLoad 16/60 Superdex 75 prep grade size exclusion chromatography column (GE Healthcare Biosciences) on an ÄKTA Purifier system (GE Healthcare Biosciences). Gel filtration buffer consisting of 20 mM Tris-HCl, pH7.5, 200 mM NaCl, 5 mM  $\beta$ -mercaptoethanol was used for size exclusion chromatography. The purified protein was concentrated using Vivaspin concentrator (10 kDa MWCO). The concentration of the protein was determined by measuring the absorbance at 280 nm and purity was verified by 15% SDS-PAGE (Fig. 44). The Se-Met labelled protein was purified following the protocol used for WT protein.

#### ***4.3.2 Crystallization of Mei2-RRM3:***

Mei2-RRM3 was concentrated up to 40 mg/ml. Initial crystallization trials were performed by sitting drop vapour diffusion method with three different protein concentrations (10 mg/ml, 20 mg/ml and 40 mg/ml) by mixing concentrated protein sample (150 nl) with equal volume of precipitant and equilibrating against 80  $\mu$ l of reservoir solution.

Crystals appeared in the following condition: 1 M Ammonium Sulphate, 0.1 M Bis-Tris pH 5.5, 1% PEG3350, protein concentration 40 mg/ml (Fig. 45a). Interestingly, this is the same crystallization condition as for Mei2-RRM1-2 fragment. Further optimization was performed by hanging drop method, by varying the crystallization condition in two directions, by creating an ammonium sulphate gradient (0.75 M-1.25 M) and by varying the pH (0.1 M Bis-Tris, pH:5.3, 5.5, 5.8). In this method, 1  $\mu$ l of concentrated protein (20 mg/ml) was mixed with 1  $\mu$ l of precipitant on a siliconized glass cover slip and equilibrated against 1ml reservoir solution, the cover-clips were sealed using grease.

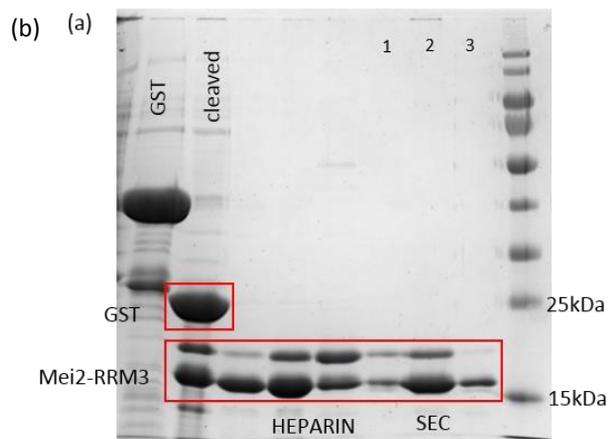
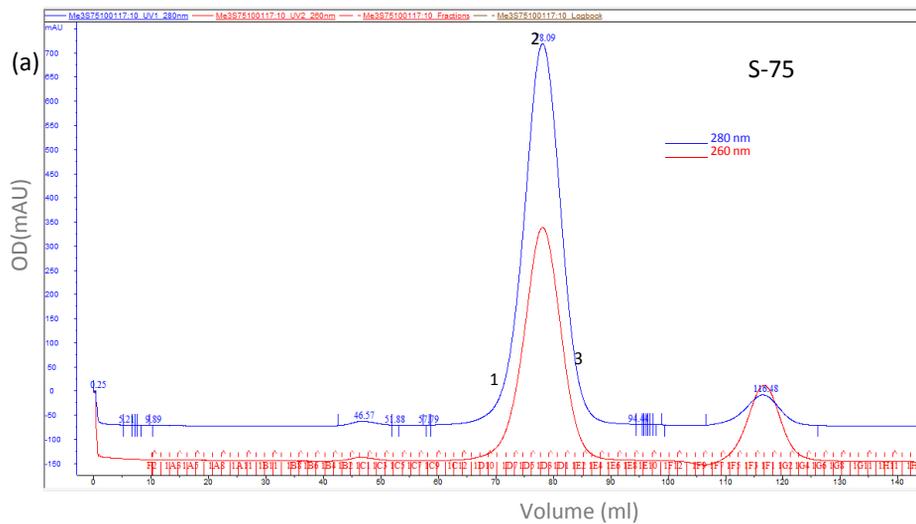


Fig. 44: Purification of Mei2-RRM3 domain (a) Elution profile of Mei2-RRM3 from gel filtration in S75 gel filtration column (b) 12% SDS-PAGE to check homogeneity of Mei2-RRM3 after GST, 3C protease cleavage, Heparin and SEC purification.

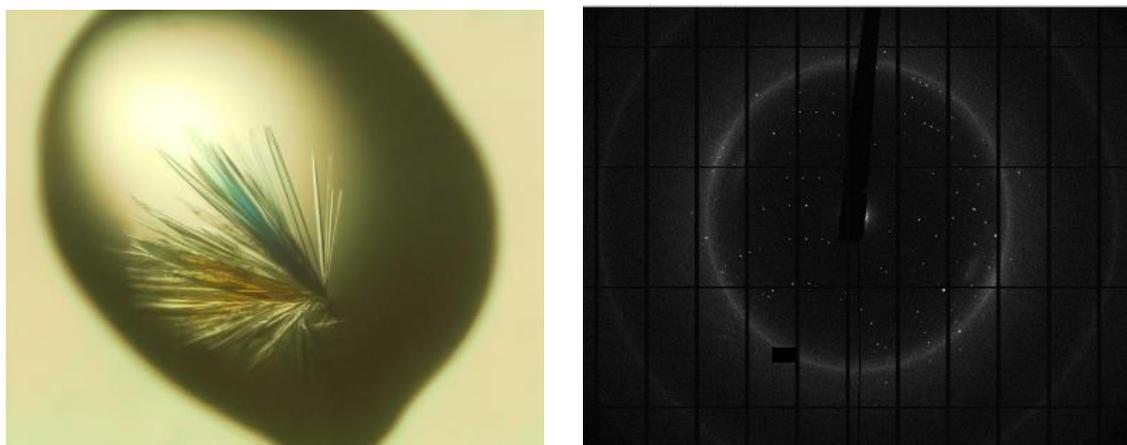


Fig. 45: Mei2-RRM3 crystals diffract (a) Mei2-RRM3 crystals grown from 1M Ammonium Sulfate, 0.1 M Bis-Tris pH 5.5, 1% PEG3350. (b) diffraction pattern of Mei2-RRM3.

As Mei2-RRM3 had very low sequence homology with existing structures, Se-Met labelled crystals were grown for experimental phasing. The SeMet-labelled protein (40 mg/ml) was crystallized by hanging drop vapour diffusion method in the same conditions as the native protein.

#### **4.3.3 Diffraction data collection and structure solution:**

Prior to data collection, the crystals were quick-soaked in cryo-protectant solutions containing 15% (v/v) and then 30% ethylene glycol in corresponding well solutions and flash-frozen in liquid nitrogen to collect the diffraction datasets. For both native and SeMet- labelled crystals, X-ray diffraction data were collected on PROXIMA 1 beamline at Soleil Synchrotron, France at 100K. The crystal was rotated through 360° with 0.1° rotation per frame, collecting 3600 images (Fig. 46b). The structure was solved with Se-SAD with a dataset collected at 0.978 Å. Finally, a 1.9 Å data was collected with the native crystal. The datasets were processed with XDS (Kabsch, 1993).

From IDXREF log file, Bravais lattice tP (P4/mmm) [primitive tetragonal cells, where a=b≠c, α=β=γ=90°] was selected, which satisfies the criteria of higher symmetry [ap>mp>op>tp>cp] and lower penalty score (Table 4.1). The point group of the crystal was determined by taking into account systematic absences at three axes of reciprocal space (h,0,0), (0, k,0) and (0,0, l). Careful observation shows that the intensity is highest for (0,0,4n) (Table 4.2) which satisfies the condition for both P4<sub>1</sub> and P4<sub>3</sub> (enantiomorphic space groups). Finally, an interpretable electron density map was obtained with space group P4<sub>1</sub>, indicating that this is the correct space group.

##### **4.3.3.1 Estimating the number of molecules in asymmetric unit:**

The number of molecule present in an asymmetric unit was calculated by Matthew's Coefficient with the following formula:

$$V_m = \frac{V}{M * n * Z}$$

V<sub>m</sub> = Matthew's coefficient (ranging from 1.66 to 4 corresponding to 30%-75% solvent content)

M = molecular weight of protein in Daltons

V = volume of unit cell

n = number of asymmetric units in unit cell

Z = number of molecules in asymmetric unit.

Table 4.1: IDXREF log file of Mei2-RRM3 crystal, indicating the possible Bravais lattice

LATTICE- CHARACTER	BRAVAIS- LATTICE	QUALITY OF FIT	UNIT CELL CONSTANTS (ANGSTROM & DEGREES)					
			a	b	c	alpha	beta	gamma
* 31	<u>aP</u>	0.0	70.8	75.5	<u>75.6</u>	<u>89.9</u>	90.0	90.0
* 44	<u>aP</u>	0.0	70.8	75.5	<u>75.6</u>	<u>90.1</u>	90.0	90.0
* 35	<u>mP</u>	0.4	75.5	70.8	<u>75.6</u>	<u>90.0</u>	90.1	90.0
* 20	<u>mC</u>	0.9	<u>106.9</u>	<u>106.7</u>	70.8	90.0	90.0	90.0
* 23	<u>oC</u>	0.9	<u>106.7</u>	<u>106.9</u>	70.8	90.0	90.0	90.0
* 25	<u>mC</u>	0.9	<u>106.7</u>	<u>106.9</u>	70.8	90.0	90.0	90.0
* 33	<u>mP</u>	1.0	70.8	75.5	<u>75.6</u>	<u>90.1</u>	90.0	90.0
* 34	<u>mP</u>	1.3	70.8	75.6	<u>75.5</u>	<u>90.1</u>	90.0	90.0
* 32	<u>oP</u>	1.3	70.8	75.5	<u>75.6</u>	<u>90.1</u>	90.0	90.0
* <b>21</b>	<b><u>tP</u></b>	<b>1.9</b>	<b><u>75.5</u></b>	<b><u>75.6</u></b>	<b><u>70.8</u></b>	<b><u>90.0</u></b>	<b>90.0</b>	<b>90.1</b>
10	<u>mC</u>	69.3	<u>103.5</u>	<u>103.5</u>	75.6	90.0	90.1	93.7
14	<u>mC</u>	69.3	<u>103.5</u>	<u>103.5</u>	75.6	90.0	90.1	86.3
13	<u>oC</u>	70.0	<u>103.5</u>	<u>103.5</u>	75.6	90.0	90.1	86.3
11	<u>tP</u>	70.0	70.8	75.5	<u>75.6</u>	<u>90.1</u>	90.0	90.0

Table 4.2: CORRECT log file of Mei2-RRM3 crystal, with statistics of reflection intensity on each axis

H	K	L	RESOLUTION	INTENSITY	SIGMA	INTENSITY/SIGMA	#OBSERVED
0	0	16	4.422	0.3020E+04	0.9715E+02	31.09	1
0	0	17	4.162	0.6504E+01	0.2418E+01	2.69	1
0	0	18	3.931	0.1509E+01	0.2914E+01	0.52	1
0	0	19	3.724	0.1044E+02	0.3050E+01	3.42	1
0	0	20	3.538	0.6231E+03	0.2103E+02	29.62	1
0	0	21	3.369	0.6553E+01	0.3119E+01	2.10	1
0	0	22	3.216	0.5130E+01	0.3251E+01	1.58	1
0	0	23	3.076	0.1294E+01	0.3375E+01	0.38	1
0	0	24	2.948	0.2995E+03	0.1103E+02	27.16	1
0	0	25	2.830	0.7007E+01	0.2551E+01	2.75	2
0	0	26	2.721	0.4497E+01	0.2458E+01	1.83	2
0	0	27	2.621	0.8423E+01	0.2366E+01	3.56	2
0	0	28	2.527	0.9770E+02	0.3719E+01	26.27	2
0	0	29	2.440	0.8045E+01	0.2888E+01	2.79	1
0	0	30	2.359	0.2548E+01	0.2379E+01	1.07	1
2	0	0	37.766	0.7375E+03	0.2390E+02	30.86	1
3	0	0	25.177	0.4347E+03	0.1025E+02	42.41	2
4	0	0	18.883	0.1622E+02	0.1212E+01	13.39	2
5	0	0	15.106	0.6104E+03	0.1453E+02	42.01	2
6	0	0	12.589	0.9401E+03	0.2209E+02	42.55	2

For Mei2-RRM3, the molecular weight M is 19638 Da and the space group is P4<sub>1</sub> meaning that Z is equal to 4,

According to the cell parameters ( $a = 75.0\text{\AA}$ ,  $b = 75.0\text{\AA}$ ,  $c = 67.9\text{\AA}$ ,  $\alpha = \beta = \gamma = 90^\circ$ ), the cell volume V is  $75.0 \times 75.0 \times 67.9\text{\AA}^3 = 381937.5\text{\AA}^3$

So,  $V_m = 381937.5 / (19638 \times n \times 4)$

If  $n = 1$ ,  $V_m = 4.86 \rightarrow$  too high

If  $n = 2$ ,  $V_m = 2.43 \rightarrow$  within permissible range

So, Matthew's coefficient strongly suggests that there are likely two molecules, in each asymmetric unit.

#### **4.3.3.2 Structure solution:**

Dataset diffracting up to  $1.9\text{\AA}$  were obtained from Mei2-RRM3 native crystals. A dataset at  $2.6\text{\AA}$  resolution was collected from a Se-Met labelled crystal at a wavelength close to the absorption edge of selenium so as to solve the structure of this domain by Se-SAD. According to the Mei2 RRM3 sequence, one monomer contains 4 methionine residues and then 4 selenium atom should be present per monomer of the SeMet-labelled protein. Based on the estimation of the asymmetric unit composition, eight Se atoms were searched and successfully located using SHEXD (Schneider and Sheldrick, 2002). Experimental phasing was performed with the PHASER\_EP program using the position of these Se atoms and density modification was performed with the RESOLVE program as implemented in the Phenix program (Terwilliger, 2004, McCoy et al., 2007, Terwilliger et al., 2008, Adams et al., 2010) (Fig. 46a; b). A first model was obtained by iterative cycles of building and refinement performed using COOT (Emsley et al., 2010) and BUSTER (Bricogne et al., 2016) programmes, respectively. The structure was further refined at high resolution using the  $1.9\text{\AA}$  resolution native dataset. The R and  $R_{\text{free}}$  of the final structure are 18.4% and 21.7%, respectively. In the final structure of the Mei2-RRM3 fragment, corresponding to residues 580-726 are visible. In addition, 7 molecules of Ethylene glycol from cryo-protectant, 3 molecules of PEG and 7 sulphate ions for crystallization buffer as well as 214 water molecules were modelled in the electron density maps (Table 4.3).

Table 4.3: Data processing, Phasing and Refinement statistics for Mei2-RRM3

<b>Data collection</b>		
	SeMet	High resolution
Space group	P4(1)	P4(1)
Unit cell parameters	75.12 Å, 75.12 Å, 67.75 Å; 90°; 90°; 90°	75.53 Å, 75.53 Å, 70.76 Å; 90°; 90°; 90°
Wavelength (Å)	0.980105	0.980105
Resolution (Å)	41.8-2.49 (2.59-2.49)	42.63-1.89 (1.94-1.89)
R <sub>merge</sub> (%)	8.1 (120.4)	5.8(84.8)
<i>I</i> / $\sigma I$	14.0 (1.3)	11.1(1.3)
Completeness (%)	99.0 (92.4)	99.0(89.6)
CC <sub>1/2</sub> (%)	99.7 (43.1)	99.9(11.1)
Redundancy	6.5	3.7
Observed reflections	86530	117825
Unique reflections	13280	31476
<b>Refinement</b>		
Resolution (Å)		42.63-1.89
R / R <sub>free</sub> (%)		18.1/21.5
<i>Number of atoms</i>		
Protein		2396
Ligands (Ethylene glycol/PEG/Sulfate)		28/21/35
Water		153
<i>R.m.s deviations</i>		
Bond lengths (Å)		0.01
Bond angles (°)		1.59

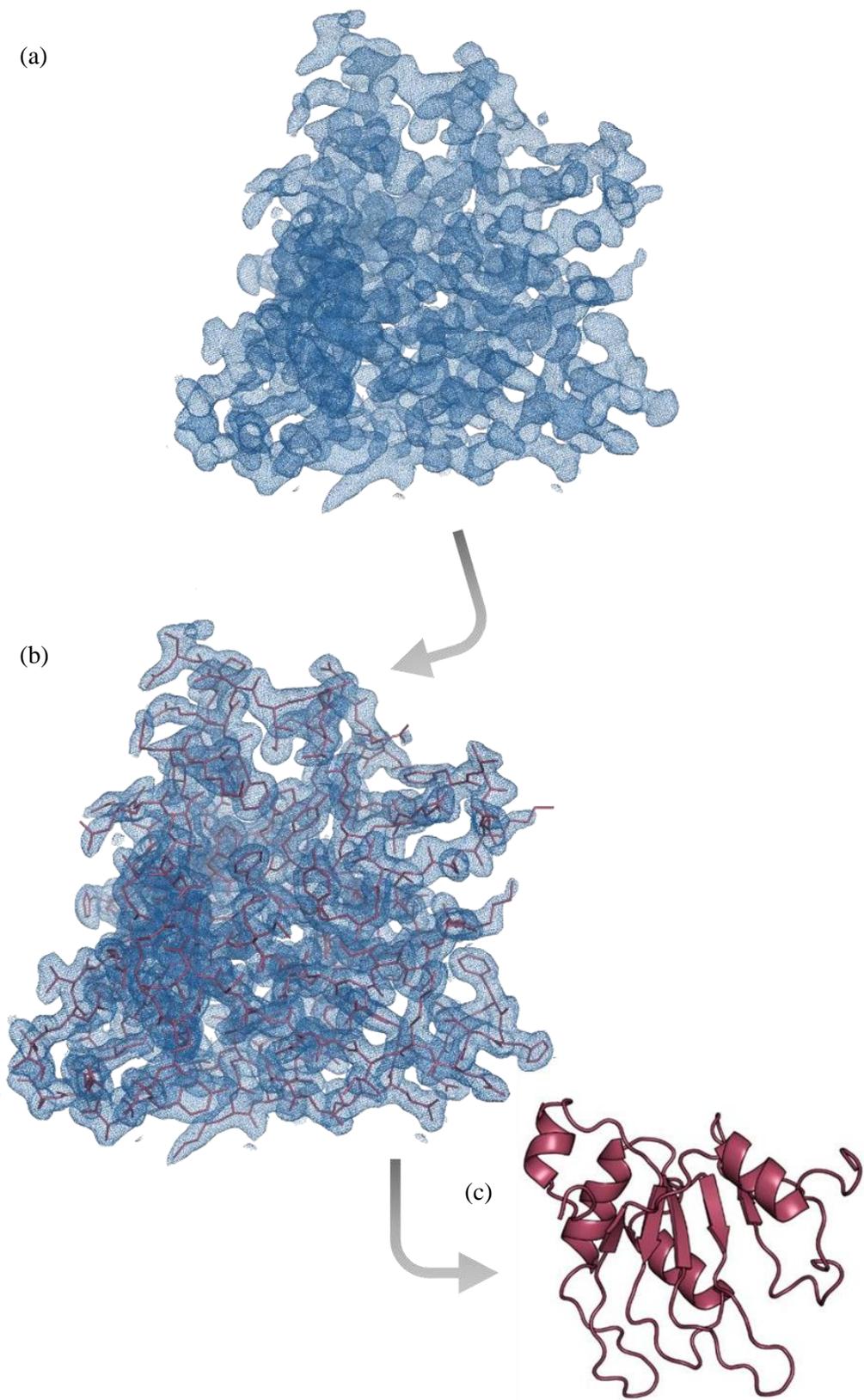


Fig. 46: Experimental electron density obtained by Se-SAD phasing (a) without the final model (b) with final model (c) Ribbon representation of final structure of Mei2-RRM3

#### 4.3.4 Structure analysis of Mei2-RRM3:

##### *Description of the structure and comparison with existing RRM domain structures:*

This domain folds as an  $\alpha/\beta$  protein with a central 6 stranded antiparallel  $\beta$ -sheet surrounded by 3  $\alpha$ -helices on one side and one helix on the other side. The Mei2 RRM3 domain is unique as the classical RRM fold ( $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$  topology) is extended by a N-terminal  $\alpha$ -helix ( $\alpha_0$ ) as well as an  $\alpha$ -helix ( $\alpha_3$ ) and two  $\beta$ -strands ( $\beta_5$  and  $\beta_6$ ) at the C-terminal extremity (residues 684 to 727) (Fig. 46c), features that have never been found simultaneously in the hundreds of RRM domain structures already solved. The two Mei2-RRM3 copies present in the asymmetric unit are virtually identical as testified by the low rmsd value of 0.433Å over 133  $C_\alpha$  atoms. The current structure confirms the bioinformatics analyses predicting that this Mei2 fragment belongs to the RRM family although our structure highlights some important differences. Indeed, classical RRM domains have a conserved  $\beta_4\alpha_2\beta_1\beta_3\alpha_1\beta_2$  topology forming a central anti-parallel  $\beta$ -sheet with 2  $\alpha$ -helices on one face while the other face of the  $\beta$  sheet surface is involved in the interaction with RNAs.

Compared to canonical RRM domain, in Mei2-RRM3, there is one additional N-terminal  $\alpha$ -helix ( $\alpha_N$ ) spanning from residues Y586 to S591. There is also a C-terminal extension containing a short  $\alpha$ -helix ( $\alpha_C$ ) ranging from residues K684 to R692 followed by a 5<sup>th</sup>  $\beta$ -strand ( $\beta_5$ ). Surface electrostatic potential calculation reveals presence of a positively charged region which largely overlaps with the classical RNA binding site from other RRMs (Fig. 47).

Mei2-RRM3 structure was next compared to existing structures of RRM domains using PDBefold (Protein structure comparison service PDBeFold at European Bioinformatics Institute (<http://www.ebi.ac.uk/msd-srv/ssm>), authored by E. Krissinel and K. Henrick) and DALI servers (Holm and Laakso, 2016) yielding rmsd values ranging from 2.0Å to 5.0 Å.

RRM domains are made of two RNP motifs, RNP1 and RNP2. The RNP motifs of Mei2-RRM3 were identified from multiple sequence alignment (Fig. 48). The consensus sequence for the two RNP motifs are as following:

RNP1 consensus	R/K	G	F/Y	G/A	F/Y	I/L/V	X	F/Y
RRM3-RNP1	V	G	Y	A	F	I	N	F
RNP2 consensus	I/L/V	F/Y	I/L/V	X	N	L		
RRM3-RNP2	V	M	I	K	N	I		

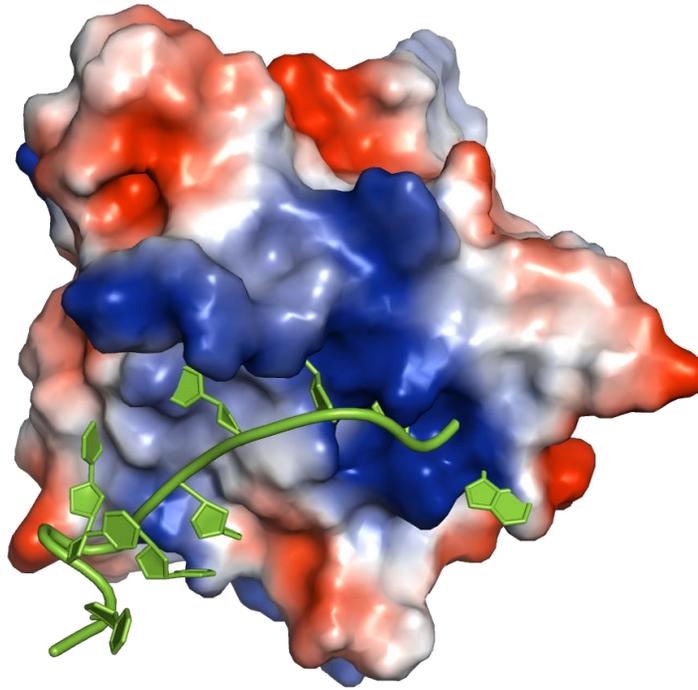


Fig. 47: Distribution of electrostatic potential on the surface of MeI2-RRM3 with Sxl- RNA superposed on it (green). Positivity ( $xk_B T/e$ ) and negativity ( $-xk_B T/e$ ) regions are coloured in red and blue respectively, neutral regions are coloured in white

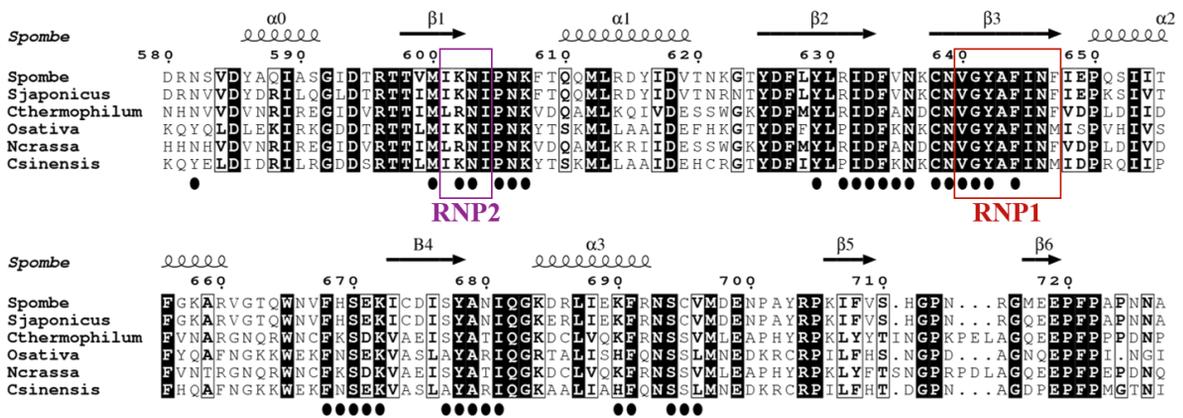


Fig. 48: Multiple sequence alignment of MeI2-RRM3 from diverse organisms. Secondary structure elements derived from the *S. pombe* MeI2-RRM3 crystal structures are shown above the sequence

Suggesting that RNP1 is highly conserved in RRM3 domain, except for the first residue (V instead of R/K).

The conserved aromatic residues (highlighted in blue) are responsible for interactions with nucleotides in classical RRM domains. Depending of conservation of the aromatic residues of RNP1, the following structures of RNA-bound RRM domains were selected to study the probable RNA binding mode of RRM3: human IMP3 RRM1-2 (PDB ID: 6fqr), human hnRNPA2B1 RRM domain (PDB ID: 5wwe), human N-terminal RRM domains of HuR (PDB ID: 4ed5) and fruit-fly Sxl-UNR translation regulatory complex RRM domain (PDB ID: 4qqb) (referred to as Sxl-UNR RRM onwards). Superposition with these structures confirms that RRM3 has the conserved RRM fold and that the aromatic residues of RNP1 are structurally conserved in all these structures suggesting RRM3 has a classical RNA binding mode. Of particular interest is the high similarity between Mei2-RRM3 and *D. melanogaster* Sxl-UNR RRM (rmsd=0.352Å); (Fig. 49a). RNP1 is identical (from position 2 to 5) among these two RRM domains, both of them have Y and F at 3<sup>rd</sup> and 5<sup>th</sup> position of RNP1. In Sxl, F170 interacts with G11 by  $\pi$ - $\pi$  stacking, whereas Y168 is inserted between U10 and G11 from the RNA. Sxl-UNR RRM Y168 and F170 match respectively with Y642 and F644 from Mei2-RRM3 (Fig. 49b). This observation strongly suggests that the hydrophobic residues of RNP1 from Mei2-RRM3 can be involved in the recognition of a G.

The Mei2 N-terminal and C-terminal extensions present in Mei2 RRM3 domain are of particular interest. So far, the presence of a fifth  $\beta$  strand has been observed in two RRM domains (RBD2, RBD3) of human Polypyrimidine tract binding protein (PTB) (Conte et al., 2000; Simpson et al., 2004; Oberstrass et al., 2005). The structure of these two domains were solved by solution NMR (2adb, 1sjr). The RNA bound structure of RBD2 (2adb) indicate that presence of the additional  $\beta$  strand helps to accommodate more residues facilitating binding of a longer RNA. The two residues, K266 and Y267 present on the loop connecting  $\beta$ 4 and  $\beta$ 5 accommodate one additional nucleotide (Fig. 50). However, the comparison with existing structures also reveals that the C-terminal extension containing helix  $\alpha$ C and strand  $\beta$ 5 might prevent an RNA to interact with Mei2-RRM3 as observed in the Sxl-UNR RRM-RNA complex.

#### ***4.3.5 Determination of optimal RNA sequence for Mei2-RRM3 binding by Isothermal Titration Calorimetry:***

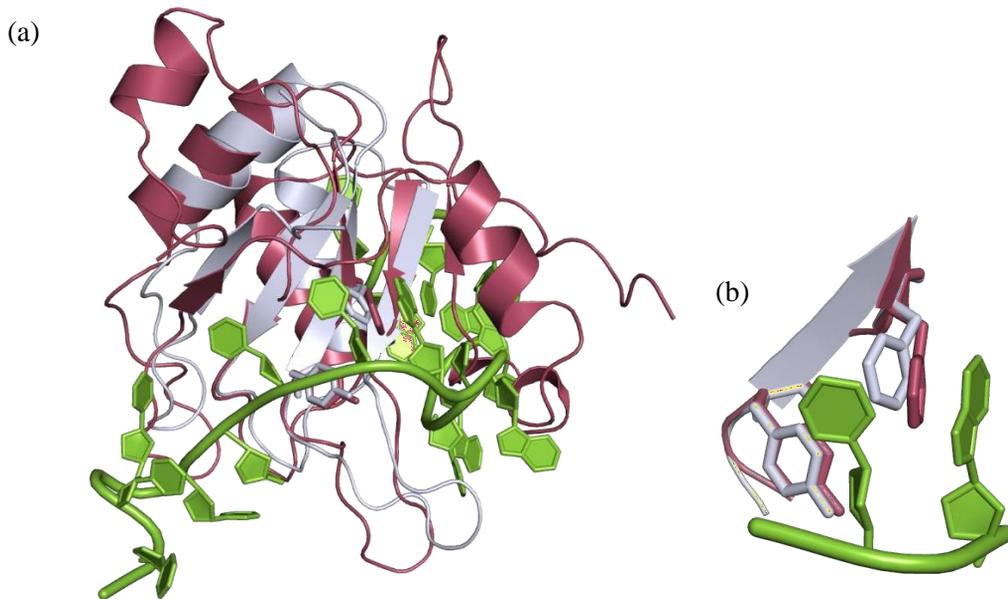


Fig. 49: Structural comparison of Mei2-RRM3 with Sxl-RRM. (a) Structural superposition of Mei2-RRM3 (raspberry) with *D. melanogaster* Sxl-RRM (lightblue)-Sxl-RNA(limon) (b) comparison of conserved aromatic residues between Sxl-RRM and Mei2-RRM3

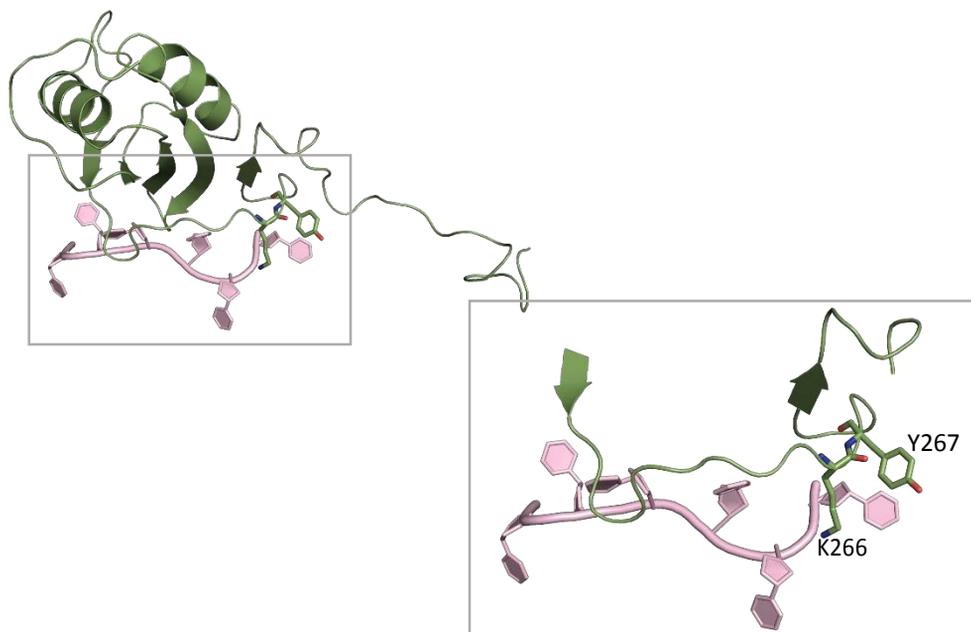


Fig. 50: The fifth  $\beta$  sheet in PTB-RRM2 domain facilitates accommodation of an extra nucleotide (PDB ID: 2adb).

The RNA binding capability of Mei2-RRM3 was investigated by ITC. Initially different lengths of Poly(U) and Poly(A) RNA oligonucleotides were tested, and U<sub>15</sub> had the best binding whereas no binding was detected for Poly(A). As our analyses of Mei2 RRM3 crystal structure revealed strong similarity with the RRM domain of Sxl-UNR protein from *D. melanogaster*, *i.e.* the Sxl-UNR RNP1 aromatic residues, which specifically recognise a G (Guanidine), are identical to the aromatic residues of Mei2-RRM3, two different constructs were tested with one and two G residues inserted in the middle: U<sub>7</sub>GU<sub>7</sub>, U<sub>6</sub>GGU<sub>7</sub>. The affinity of Mei2-RRM3 for these two RNA constructs were studied by ITC. For U<sub>7</sub>GU<sub>7</sub> Mei2-RRM3 had very high affinity (K<sub>d</sub>=104 nM) whereas for U<sub>6</sub>GGU<sub>7</sub>, the affinity was more than two folds weaker (K<sub>d</sub>=277nM) (Fig. 51). The RNA-protein stoichiometry was cross-verified with SEC-MALLS to be 1:1 (Fig. 52). Based on this information, two constructs were designed for co-crystallization of Mei2-RRM3 and RNA:

- i. 5'-GCUUUUUGUUCG-3',
- ii. 5'-GCUUUUUGUUUUUCG-3'.

The first two and last two bases, GC and CG dinucleotides at 5' and 3' extremities, respectively, were added to enhance crystallization of protein-RNA complexes by favouring crystal packing.

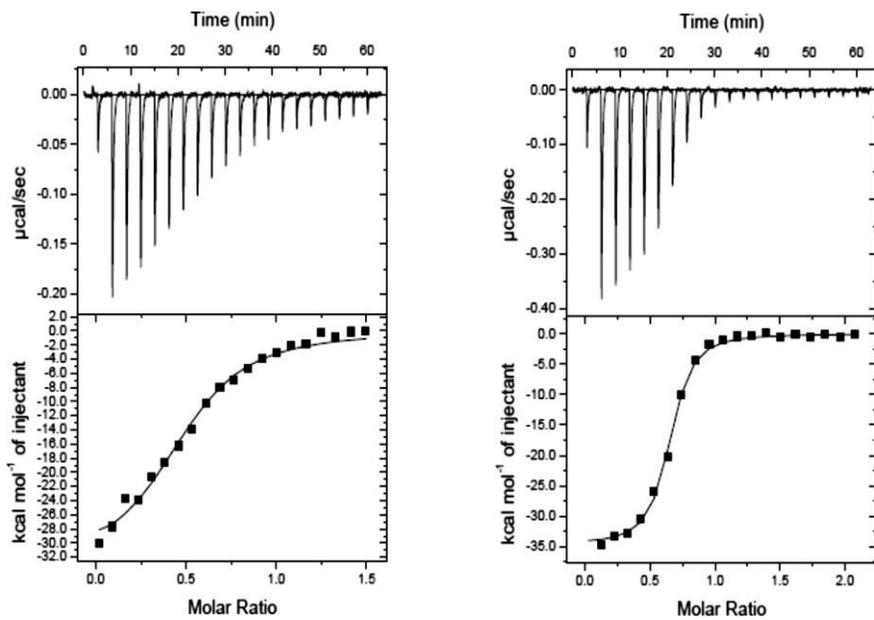
#### **4.4 Study of RRM3-RNA:**

##### **4.4.1 Co-crystallization trial of RRM3-RNA:**

Co-crystallization trials for RRM3-RNA complexes were performed by mixing 1.2 molar excess of each synthetic RNA construct with 1 mM (20 mg/ml) of protein. Preliminary trials were performed with the commercial crystallization screens: Protein Complex (Qiagen) and Nucleix (Qiagen). Crystals appeared in multiple conditions for the complex formed between Mei2-RRM3 and the 5'-GCUUUUUGUUCG-3' oligonucleotide. Diffracting crystals were obtained from the following condition: 0.2 M NaCl, 0.1 M Na/K phosphate, pH6.5, 25% w/v PEG1000 (Fig. 53).

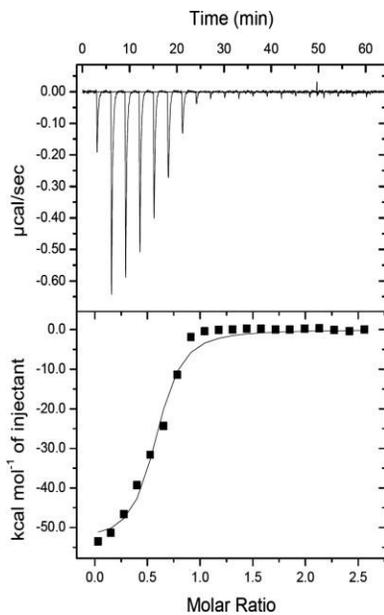
##### **4.4.2 Diffraction data collection and structure solution:**

Prior to data collection, the crystals were quick-soaked in cryo-protectant solutions containing the corresponding well solutions and supplemented with 15% (v/v) and then 30% glycerol and flash-frozen in liquid nitrogen to collect the dataset. Data were collected on PROXIMA-1 beamline at Soleil Synchrotron, France. The crystal was rotated through 360°



RRM3 vs U15

RRM3 vs U7GU7



RRM3 vs U6GGU7

Protein	RNA	$K_d$	N	$\Delta H$ (cal/mol)	$\Delta S$ (cal/mol/deg)
Mei2_RRM3	U7GU7	104nM	0.618	-3.456E4	-86.2
Mei2_RRM3	U6GGU	277nM	0.553	-5.381E4	-154
Mei2_RRM3	U15	746nM	0.497	-3.306E4	-84.7

Fig. 51: The table summarising the binding isotherms characterizing formation of RRM3-RNA complex as measured by ITC. Tabulation of thermodynamic parameters for each binding

with 0.1° rotation per frame, collecting 3600 images. The dataset was processed with XDS (Kabsch, 1993).

#### ***4421 Space group determination and calculation of the number of molecules in asymmetric unit:***

The space group of RRM3-RNA complex was determined to be I222, following the same protocol described for space group determination of RRM3. The statistics for this dataset can be found in Table 4.2.

From Matthew's coefficient the number of protein-RNA complex in an asymmetric unit was estimated to be 1.

#### ***4422 Structure solution:***

The structure was solved by molecular replacement using apo Mei2-RRM3 as a model. Analysis of the residual Fo-Fc electron density map revealed the presence of the RNA fragment in the vicinity of the classical RNA binding site in other RRM-RNA structures. The RNA was built into the density using Sxl-RNA as guide (PDB ID: 4qqb). The final structure was obtained after multiple rounds of model building and refinement using COOT (Emsley et al., 2010) and BUSTER (Bricogne et al., 2016) programs. The final model had a R and R<sub>free</sub> of 20.64% and 30.25%.

### ***4.4.3 Structure analysis of Mei2-RRM3-RNA:***

#### ***4431 Description of the structure:***

In the final structure, Mei2-RRM3 residues 580-726 and all RNA nucleotides are visible (Fig. 54a). Upon binding to Mei2-RRM3, the 5'GCUUUUUGUUCG3' RNA adopts a single stranded conformation. The superposition of the RNA bound structure onto the apo structure yields an rmsd value of 0.537Å over 134 C<sub>α</sub> atoms. The RNA binds to the positively charged region formed by the β-sheet of RRM3 in a similar way as for other RRM domains. The RNA-protein interaction is mediated by the conserved aromatic residues present in RNP1 and RNP2. Superposition of the RNA bound structure with the apo-structure reveals that these residues changed their side chain conformation to accommodate nucleotides (Fig. 54b). In this classical mode of interaction, Y629 of RNP2 from strand β<sub>2</sub> and Y642 and F644 of RNP1 from strand β<sub>3</sub> are involved in interaction with the RNA. Interestingly, F634 and

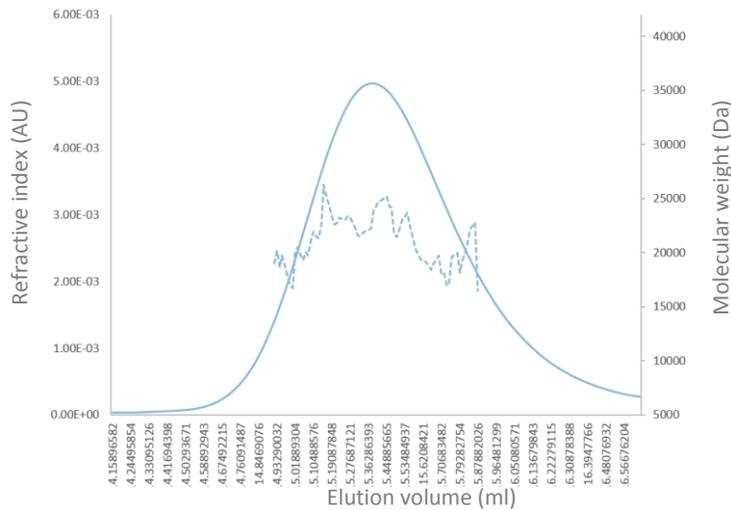


Fig. 52: Molecular weight and stoichiometry of RNA-RRM3 complex, determined by SEC-MALLS

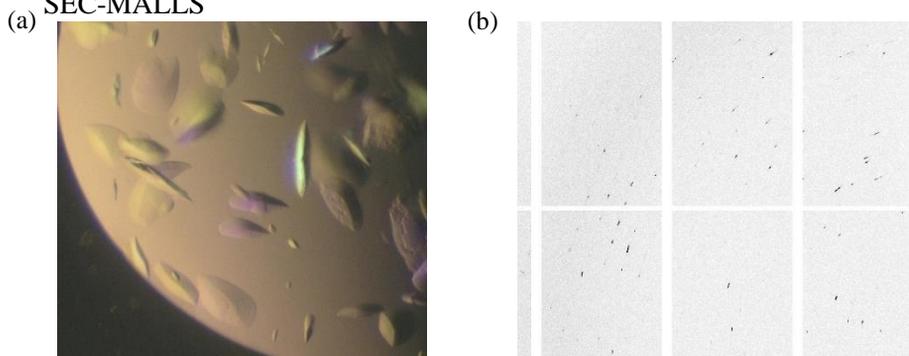


Fig. 53: Structural study of RRM3-RNA complex (a) crystals of RRM3-RNA from 0.2M NaCl, 0.1M Na/K phosphate pH6.5, 25%w/v PEG1000 (b) Diffraction spots of RRM3-RNA crystal

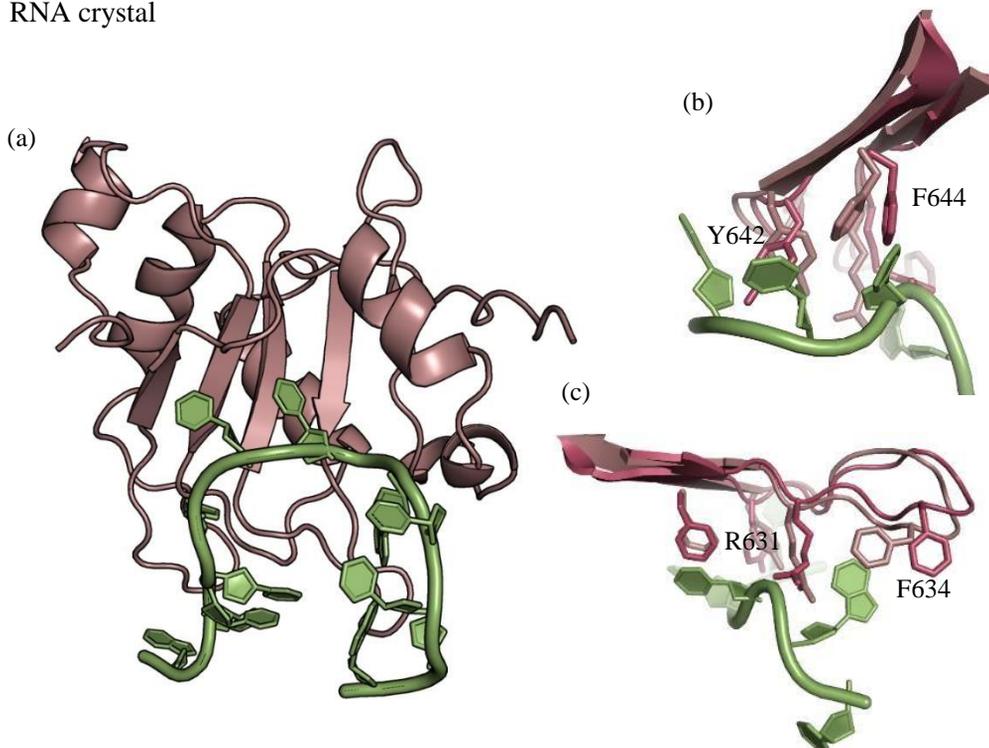


Fig. 54: Structure of RRM3-RNA (a) Representation of RRM3 (brown) and RNA (green) Superposition of apo-RRM3 (pink) and RRM3-RNA complex (dirty violet) reveals the conformational change of the conserved aromatic residues Y642 and F644 (b) & (c) conformational change of the R631 and F634 to accommodate G8.

Table 4.4: Data processing, Phasing and Refinement statistics for RRM3-RNA

<b>Data collection</b>	
Space group	I222
Unit cell parameters	74.85Å; 81.4Å; 81.4Å; 90°; 90°; 90°
Wavelength (Å)	0.980105
Resolution (Å)	40.50-2.65(2.78-2.65)
R <sub>merge</sub> (%)	25.4 (167.3)
<i>I</i> / $\sigma I$	6.41(1.2)
Completeness (%)	99.6 (97.7)
CC <sub>1/2</sub> (%)	98.6 (72.8)
Redundancy	11.4
Observed reflections	83891
Unique reflections	7387
<b>Refinement</b>	
Resolution (Å)	40.69-2.63
R / R <sub>free</sub> (%)	21.1 / 30.25
<i>Number of atoms</i>	
Protein	2353
RNA	374
Mg <sup>2+</sup>	2
Water	10
<i>R.m.s deviations</i>	
Bond lengths (Å)	
Bond angles (°)	

R631, two residues on the loop connecting strands  $\beta_3$  and  $\beta_4$ , participate in a specific interaction with G8 (Fig. 54c), which is reminiscent of quasi-RRM domains (Dominguez et al., 2010). The quasi-RRM or qRRM domains are different from classical RRM domain in their RNA binding mode. In classical RRM domains, the aromatic residues on the  $\beta$ -sheet surface are involved in protein-RNA interaction, whereas for qRRM the interaction with RNA is mediated by positively charged and aromatic residues located in loops. Human hnRNP F qRRM contacts RNA with residues located on loop (Fig. 55). However, the  $\beta_5$  strand does not interact with the RNA as the  $\alpha$ -helix (K684-R692) connecting strand  $\beta_4$  to strand  $\beta_5$  blocks the surface in front of the fifth  $\beta$  strand. One residue, K690 from this  $\alpha$  helix contacts the RNA. In classical RRM domains the position 1 residue of RNP1 (R/K) forms salt bridge with the backbone of RNA. In RRM3 it is Valine (V640). Although V640 is not involved in interaction with the RNA, *in silico* mutation of V to R/K results in steric clash with G8. Indicating presence of Valine instead of a bulky residue like arginine/Lysine helps to selectively accommodate G8.

Results from ITC experiments indicate that Mei2-RRM3 has a higher affinity ( $K_d=104\text{nM}$ ) for the 15 base RNA ( $U_7GU_7$ ) than the 12-mer RNA (GCUUUUUGUUCG) used for crystallization ( $K_d=205\text{ nM}$ ). From this observation, it is tempting to speculate that *in vivo* Mei2-RRM3 binds to a longer RNA utilizing the  $\beta_5$  strand.

To understand the RRM3-RNA interaction in more details, bases were mutated *in silico* to determine which bases might be specifically identified by the protein.

#### **4432 Protein-RNA interaction:**

The interaction between the RNA and protein is mediated by multiple H-bonds formed between protein and the backbone of the RNA.

The sugar-phosphate backbone forms multiple contacts with the protein, these non-bonded interactions are tabulated below:

Base (atom)	Amino acid (atom)	Type of interaction
C2 (O2')	F668 (O)	H-bond
U3 (OP2)	N606(N)	H-bond
U4(OP2)	K672(NZ)	Electrostatic

G8(OP2)	Y629(OH)	H-bond
G8(OP2)	R631(NE)	Electrostatic
G8(O5')	R631(NE)	H-bond
G8(O5')	R631(NH1)	H-bond
G8(O4')	R631(NH1)	H-bond

#### **4433 Recognition of bases:**

**U4 & U5:** these residues form H-bond with N606 (ND2) and K672 (NZ) respectively.

**U6:** U<sub>6</sub> is anchored in a pocket by three hydrogen bonds. The O4 of U<sub>6</sub> forms a H-bond with side chain of N582 (ND2), N3 forms a H-bond with Y678 main chain carbonyl group. O2 contacts side chain of N680 (ND2). Substitution of U<sub>6</sub> to C<sub>6</sub> would disrupt the H-bond with O4 while the presence of a purine would cause steric hindrance.

**U7:** U<sub>7</sub> is contacted by  $\pi$ - $\pi$  stacking with F644 and cation- $\pi$  with K690. Apart from stacking the side chain also forms H-bond with I681 main chain. Substitution by a cytosine would disrupt this H-bond while purines at this position would lead to steric clash (Fig. 56).

**G8:** this base forms four H-bonds, N3 of the base with R631 side chain (NH), N1 & N2 with main chain carbonyl group from I632, and O6 with F634 main chain (N). Apart from these H-bonds, G<sub>8</sub> is specifically accommodated in the pocket by  $\pi$ - $\pi$  stacking with F634. An adenosine at this position would disrupt multiple H-bonds, while substitution by C or U would disrupt H-bonds and stacking (Fig. 56).

These observations suggest that Mei2 prefers binding to a signature motif consisting of UUG. Careful examination of the long non-coding meiRNA sequence revealed repeated presence of a motif UUUUUGUU, strongly suggesting that this is the consensus binding motif for Mei2. However, further experiments are required to validate this hypothesis.

#### **4434 Identification of RNA binding residues:**

From the RNA bound structure, the residues necessary for RNA binding could be identified. The conserved aromatic residues from RNP1 Y642 and F644, play a key role in protein-RNA interaction. In a previous study, Yamamoto group has shown that cells expressing the F644A

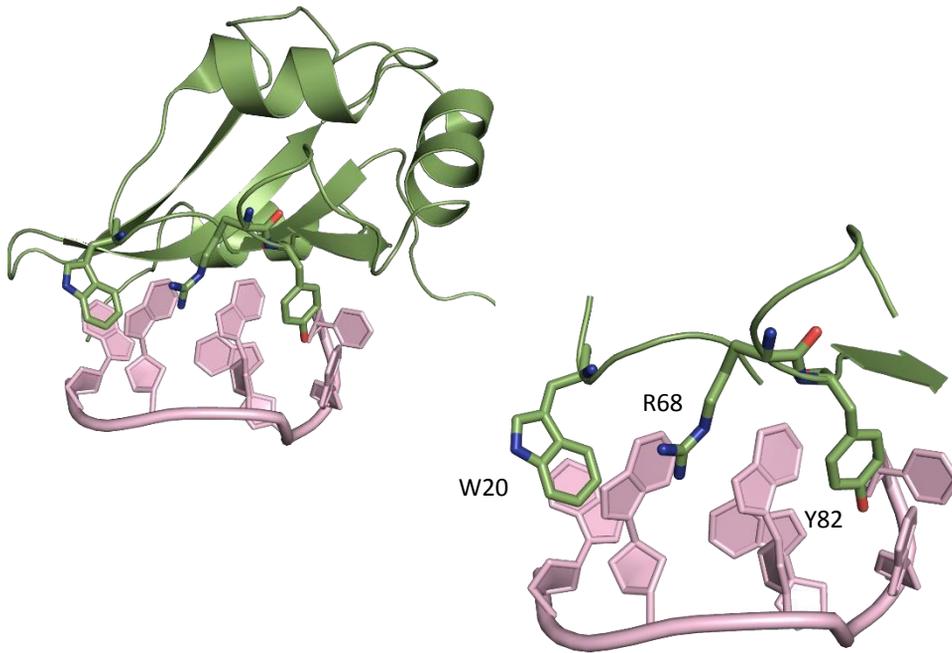


Fig. 55: Human hnRNP quasi-RRM domain interacts with RNA via residues located on loop (PDB ID: 2KFY)

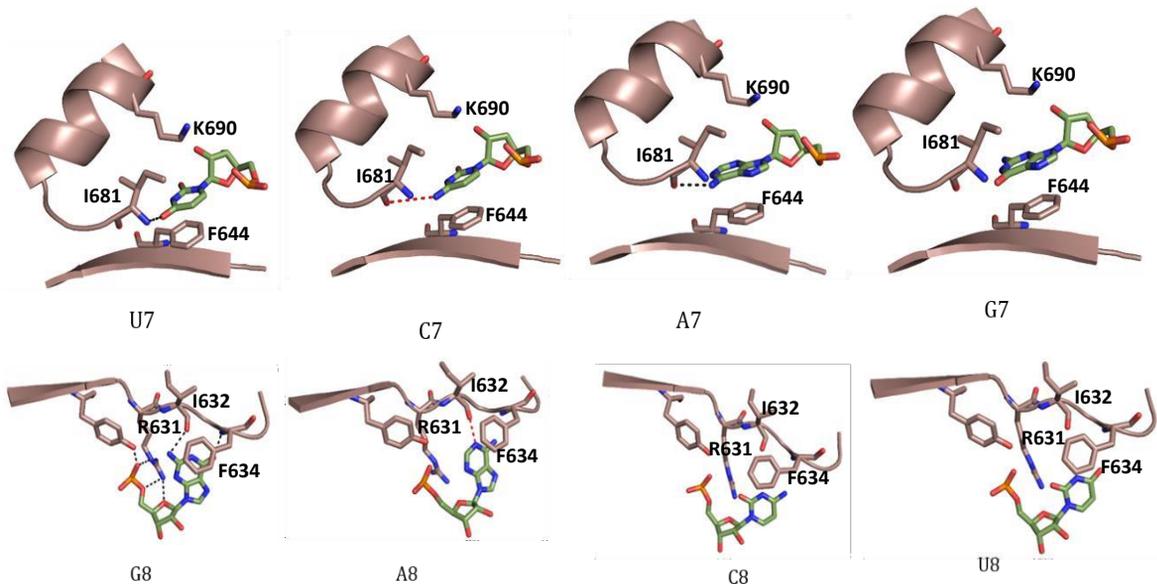


Fig. 56: Analysis of MeI2-RRM3 specificity for nucleotides at position 7 & 8 (a) Binding pocket for residue 7, U7 generates from our structure while C, A & G has been mutated in-silico. Black dashed lines indicate H-bond. Red dashed lines indicate distances too short for H-bonding. Red Ellipses indicate steric clash (b) Same as U7. Black dashed lines indicate H-bond. Red dashed lines indicate distances too short for H-bonding.

mutant is incapable of pre-meiotic DNA synthesis and meiosis (Yamashita et al., 1998). The aromatic residues from RNP1, Y642 and F644, were then mutated to alanine to investigate their role in RNA binding. Interaction of the Y642A and F644A mutants with RNA were studied by ITC, with the U<sub>7</sub>GU<sub>7</sub>, the RNA construct for which the wild-type protein has highest affinity. For both of these mutants, RNA binding was completely abolished (Fig. 57).

It has also been shown that Tor2 phosphorylates Mei2 on 9 sites. In the Mei2-RRM3 construct, only one phosphorylation site (S694) is present, ~3.8Å from the RNA backbone. To study the effect of S694 phosphorylation on RNA binding, this residue was substituted by a glutamic acid, which is a well-established structural mimetic for phosphorylated serine. The mutant was generated by site directed mutagenesis. The binding property of S694E for the U<sub>7</sub>GU<sub>7</sub> RNA was studied by ITC, revealing that mutation of Serine at position 694 by glutamic acid severely impairs the RNA binding property of Mei2-RRM3, *i.e.* the K<sub>d</sub> is 10-fold higher than that of WT protein (Fig. 57).

#### ***4.5 Mmi1- a putative partner of Mei2:***

When fission yeast enters meiosis, Mmi1 is sequestered by Mei2 and meiRNA, resulting in the stabilisation of meiotic mRNAs containing DSR motifs, thereby allowing initiation and progression of meiosis. A study by Yamamoto group proposed that Mmi1 is a putative partner of Mei2 (Harigaya et al., 2006). In their study, they have found that the C-terminal of Mei2 has high affinity for Mmi1. The interaction between Mei2-RRM3 and Mmi1-YTH was then investigated further.

##### ***4.5.1 Cloning:***

ORFs corresponding to Mmi1 full-length and YTH domain (322-488) were successfully amplified from *S. pombe* genomic DNA. The PCR product was checked by running on agarose gel. Both construct had a N-terminal 6x-His tag and were cloned in pET28b vector.

##### ***4.5.2 Solubility profiling:***

Expression test was performed for FL and YTH domain of Mmi1 in TBAI or 2YT media with BL21 Codon+ and Gold cells, where all the constructs are His<sub>6</sub> tagged. Cells grown in 2YT induced with varying amount of IPTG and in TBAI media, were grown at 18 °C. The cells were centrifuged and after a small scale purification, the elution fractions were checked

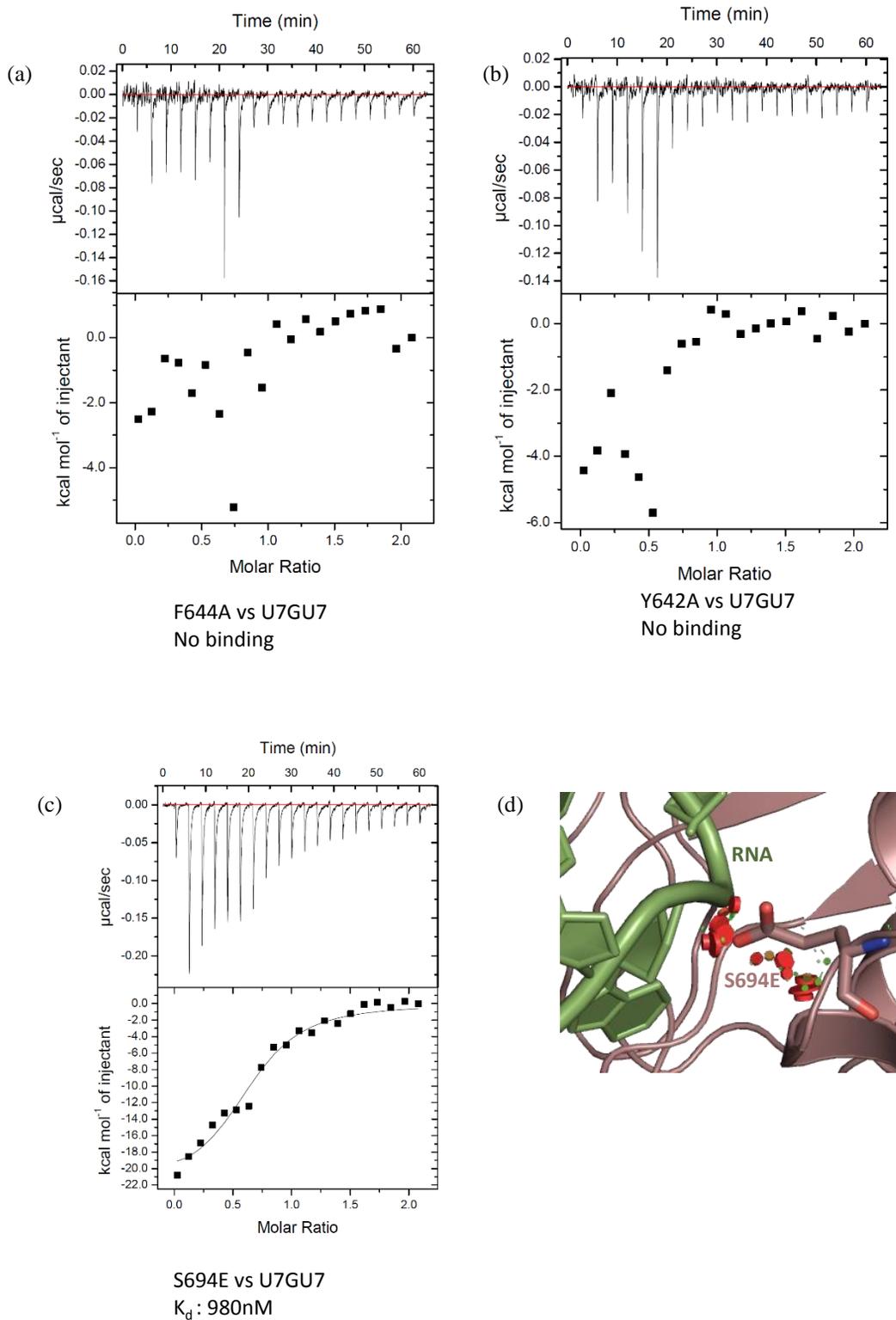


Fig. 57: Analysis of RNA binding to Mei2-RRM3. Each point mutant F644A (a) Y642A (b), S694E (c) were titrated with U7GU7. (d) S694E mutation leads to steric hindrance displayed by red disks.

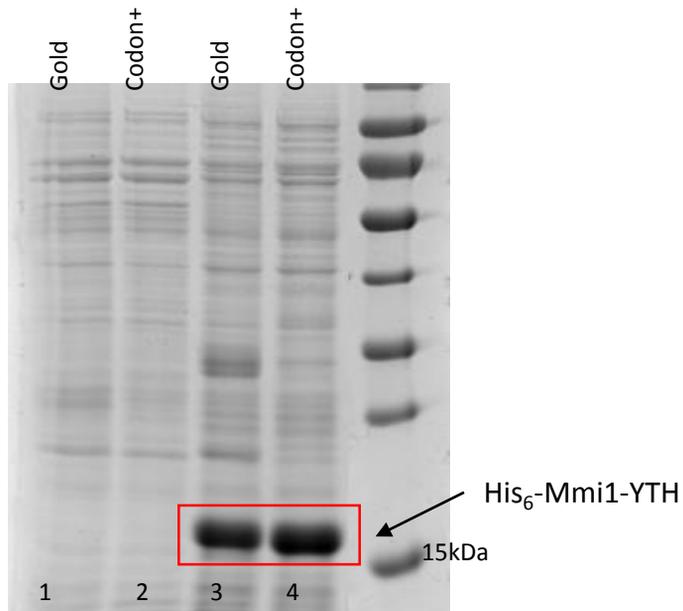


Fig. 58: Expression assay for Mmi1 constructs at 18 °C and in two *E. coli* strains (Gold, Codon+). 15% SDS-PAGE gel of proteins eluted from Ni NTA resin for expression assay for Mmi1-FL (lane 1 & 2), Mmi1-YTH (lane 3 & 4). The protein of interest is highlighted in red box

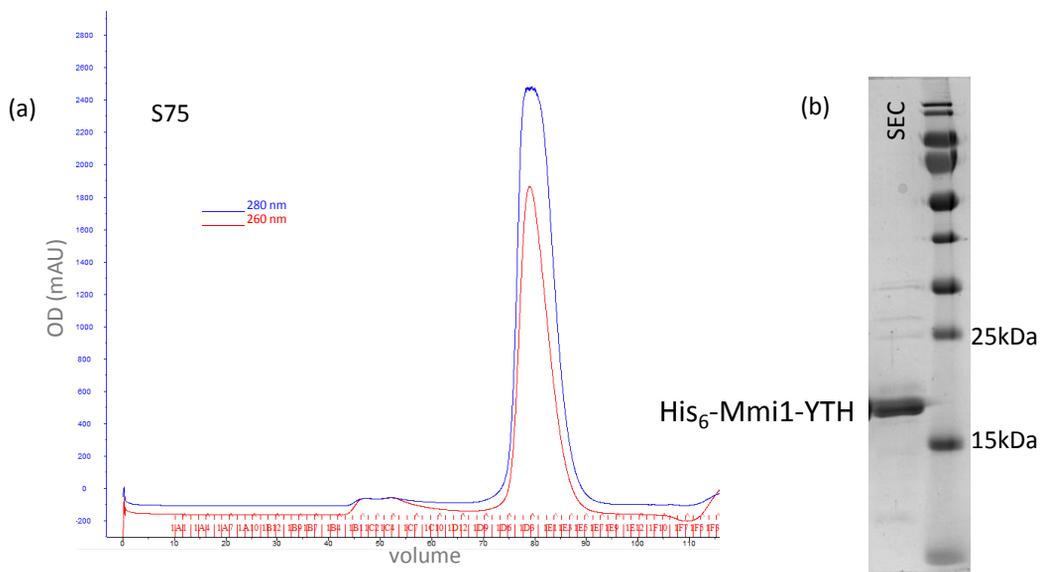


Fig. 59: Purification of Mmi1-YTH. (a) Elution profile of Mmi1-YTH from S75 gel filtration (b) SDS-PAGE gel to check homogeneity of Mmi1-YTH after purification from Size exclusion chromatography.

on 15% SDS-PAGE gel. The presence of the Mmi1-YTH was detected in both fractions, from Codon+ and Gold cells during solubility profiling. However, the full length protein was not present in the elution fraction (Fig. 58).

#### ***4.5.3 Protein production and purification of Mmi1-YTH:***

The Mmi1-YTH recombinant domain was over-expressed in BL21 (DE) Codon+ cells and in 2YT media, containing Kanamycin (25µg/ml) and Chloramphenicol (25 µg/ml). After OD<sub>600</sub> reached 0.6 at 37 °C, culture was induced by 400 µM IPTG and transferred to 18 °C for 16 h.

Mmi1 was purified by a three step purification process: Ni-NTA, Heparin and SEC. From Ni-NTA column (Protino Ni-NTA agarose MACHEREY NAGEL), the protein was eluted with elution buffer (20 mM Na citrate, pH 6.0, 200 mM NaCl, 400 mM Imidazole, 5 mM βMe), following the protocol of previous groups (Stowell et al., 2016). The protein was purified on a Heparin column and eluted using a linear gradient ranging from 100% Buffer A (50 mM Citrate, pH6.0, 50 mM NaCl, 5 mM 2-mercaptoethanol) to 100% Buffer B (50 mM Citrate, pH6.0, 1 M NaCl, 5 mM 2-mercaptoethanol). The protein eluted from Heparin column was concentrated by Vivaspin 10 concentrator (10 kDa MWCO, Sartorius) and loaded on a HiLoad 16/60 Superdex 75 prep grade size exclusion chromatography column (GE Healthcare Biosciences) on an ÄKTA Purifier system (GE Healthcare Biosciences). Gel filtration buffer consisting of 20 mM Tris-HCl, pH7.5, 200 mM NaCl, 5 mM 2-mercaptoethanol was used for size exclusion chromatography. The purified protein was concentrated using Vivaspin concentrator (10 kDa MWCO). The concentration of the protein was determined by measuring the absorbance at 280 nm with nanodrop and purity was verified by 15% SDS-PAGE (Fig. 59).

#### ***4.5.4 Determination of interaction between Mmi1-YTH and Mei2-RRM3:***

Multiple trials were made using different techniques to study the interaction between Mmi1-YTH and Mei2-RRM3.

##### ***4.5.4.1 Co-purification:***

To test if Mmi1-YTH physically interacts with the Mei2-RRM3, co-purification was performed, where the two proteins were separately expressed and the pellets (His<sub>6</sub>-Mmi1-YTH & GST-Mei2-RRM3) were mixed together before sonication. The lysate was first purified by Ni-NTA, the elution from Ni-NTA was precipitating even after filtration, therefore it was not possible to proceed further. The samples from purification were run on

SDS-PAGE gel. Surprisingly, Mmi1 was present in flow-through and wash. In the elution, there was a prominent band of GST-Mei2-RRM3 and a very faint band of His<sub>6</sub>-Mmi1-YTH (Fig. 60a). To clarify this, ITC was performed.

#### **4.5.42 ITC:**

Interaction between Mmi1-YTH and Mei2-RRM3 was studied using ITC, with 40µM Mmi1-YTH in the cell being titrated by 400µM Mei2-RRM3 in the syringe. Before the experiment, the protein samples were dialyzed in 20mM Tris-HCl pH7.5, 200mM NaCl, 5mM β-mercaptoethanol. The protocol used for Mei2-RRM3 and RNA was followed. No interaction could be observed using this experiment (Fig. 60b).

#### **4.5.43 Pull-down:**

Mmi1-YTH and Mei2-RRM3 interaction was studied by Ni-NTA pull-down performed following the standard protocol described in section 3.8. An RNA fragment harbouring binding sites for both Mmi1-YTH and Mei2-RRM3 domains was designed from meiRNA to test if their interaction is mediated by RNA. In lane 1 and 2 of input, the two domains, Mmi1-YTH and Mei2-RRM3 are present and in lane 2, the RNA harbouring binding site of both domains is added. The difference of molecular weight between His-Mmi1-YTH and Mei2-RRM3 is >0.5 kDa therefore it is difficult to find distinct bands. From this pull down, no strong interaction was detected between the two protein domains even in the presence of RNA. As the binding sites for Mmi1-YTH and Mei2-RRM3 are complementary, hence the possibility that the RNA forms a secondary structure cannot be completely excluded. It remains unclear whether the interaction between Mmi1-YTH and Mei2-RRM3 is bridged by an RNA but it can be positively concluded that they do not interact physically in the absence of RNA (Fig. 61).

#### **4.5.5 Crystallization of Mmi1-Mei2-RNA:**

During meiosis, Mmi1 is sequestered by Mei2 and meiRNA but no physical interaction was observed between Mmi1-YTH and Mei2-RRM3, indicating that the interaction between Mmi1-YTH and Mei-RRM3 might be bridged by RNA. In an attempt to co-crystallize these two domains, the RNA containing binding sites for both proteins was incubated with both domains. Preliminary crystallization trial was performed by sitting drop method by mixing 1mM of each Mmi1-YTH and Mei2-RRM3 domains with 1.2 molar excess of the RNA and plates were set by mixing 150 nl of sample with 150 nl of precipitant solution, equilibrated against 80µl of reservoir solution. The JCSG plus (Molecular Dimensions), Protein Complex Suite (Qiagen) and Nucleix (Qiagen) screens were used for primary trials. Crystals appeared

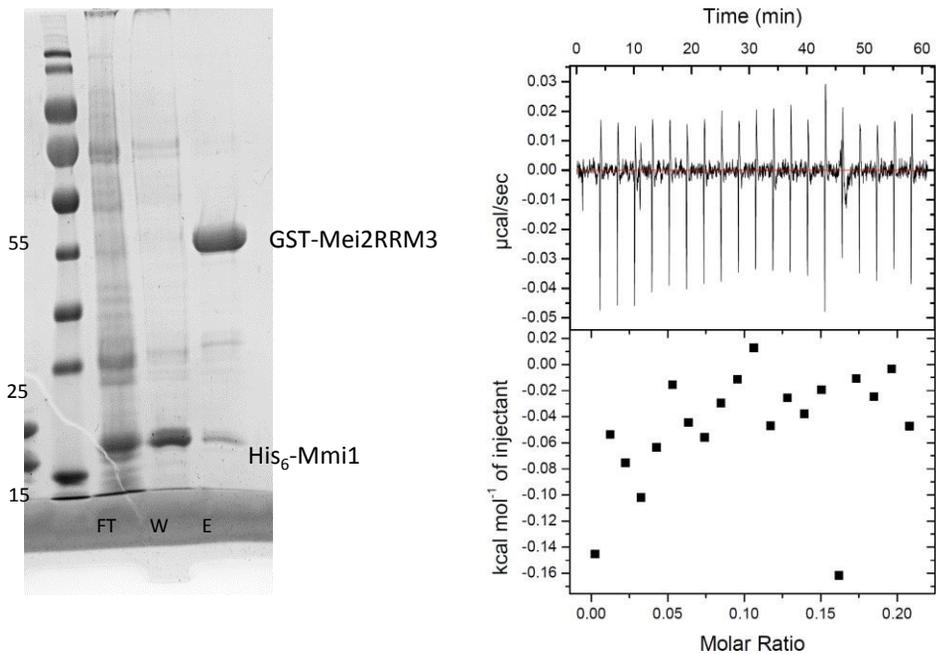


Fig. 60: Mei2-RRM3 and Mmi1-YTH do not interact directly. (a) Co-purification trial of Mmi1-YTH and Mei2-RRM3. FT: flow through, w: wash E: elution from Ni-NTA resin(b) Results from ITC measurements to study interaction between Mmi1-YTH and Mei2-RRM3.

Lane	1	2	3	4	1	2	3	4
His-Mmi1-YTH	+	+	+	-	+	+	+	-
Mei2-CTD	+	+	-	+	+	+	-	+
RNA	-	+	-	-	-	+	-	-

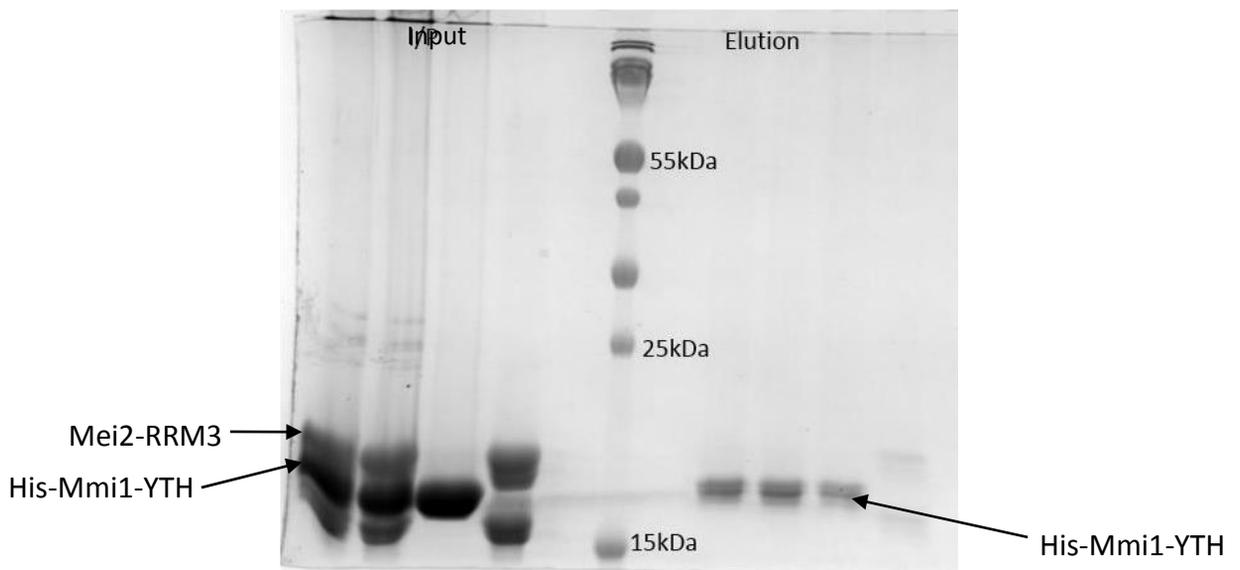


Fig. 61: Ni-NTA pull-down to study the interaction between His-Mmi1-YTH and Mei2-RRM3. in lane 1 both proteins have been added and in lane2 an RNA fragment, containing binding sites for both proteins are present

in multiple conditions. Long rod shaped crystals were obtained from the condition 0.1M Sodium-citrate pH6.0, 2M NaCl (Fig. 62) and were tested on Proxima-2A beamline from SOLEIL synchrotron. Unfortunately, all of them were found to be crystals of Mmi1.

#### **4.5.6 Mmi1 structure:**

The data was processed with XDS and structure was solved at 2.38 Å, by molecular replacement using Mmi1 (PDB ID: 5DNP) as search model. Residues 322-485 are visible in the structure. The final model was obtained by iterative cycles of building and refinement using COOT (Emsley et al., 2010) and BUSTER (Bricogne et al., 2016), respectively. Mmi1 adopts a typical YTH fold with five β sheet strands surrounded by four α helices. The structure belongs to space group C2 with two molecules present in the asymmetric unit (Fig. 63), they are identical in conformation with rmsd value of 0.288 Å over 146 C<sub>α</sub> atoms. The Mmi1 structure was compared to existing structures of Mmi1 (PDB ID:5DNP), revealing a lot of similarity with an rmsd value of 0.368 over 264 C<sub>α</sub> atoms. This structure is thereby not discussed in this manuscript. Data processing and structure refinement statistics are listed in table 4.5.

#### **4.6 Erh1:**

##### **4.6.1 Cloning:**

The sequence encoding for full length of Erh1 was amplified from *S. pombe* genomic DNA. The amplified PCR product was checked by running on 1% agarose gel. Two constructs were generated for Erh1, one with a GST-tag (pMG921) and the other one without any tag (pMG948).

##### **4.6.2 Interaction between Mmi1 & Erh1:**

###### **4.6.2.1 Expression assays:**

Although a structure of Erh1-Mmi1 complex was published recently (Xie et al., 2019) the boundaries of the Mmi1 region that interacts with Erh1 were unknown when this work was initiated, the following constructs of Mmi1 NTD: 1-75, 1-178, 1-240, 1-294, 1-322, 1-350, 60-178, 78-178 fused to a His<sub>6</sub>-tag (refer to table S3 for plasmid details), were cloned and tested to determine the Mmi1 region interacting with Erh1. Next, expression assays were performed for each of these constructs alone and in presence of Erh1. None of these Mmi1 fragments was expressed as soluble protein when expressed alone. However, upon co-expression with Erh1, two constructs Mmi1 1-322 and 1-350 were solubilized (Fig. 64a).

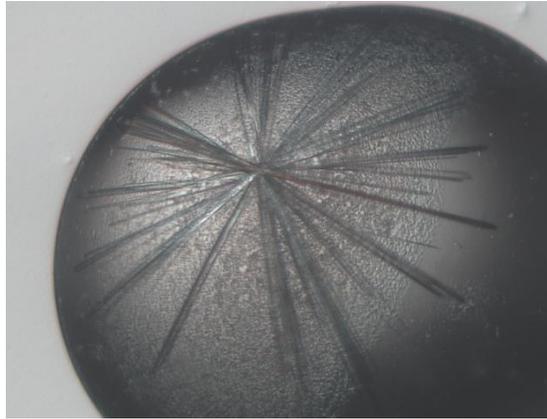


Fig. 62: Mmi1-YTH crystals grown in 0.1M Sodium-citrate pH6.0, 2M NaCl

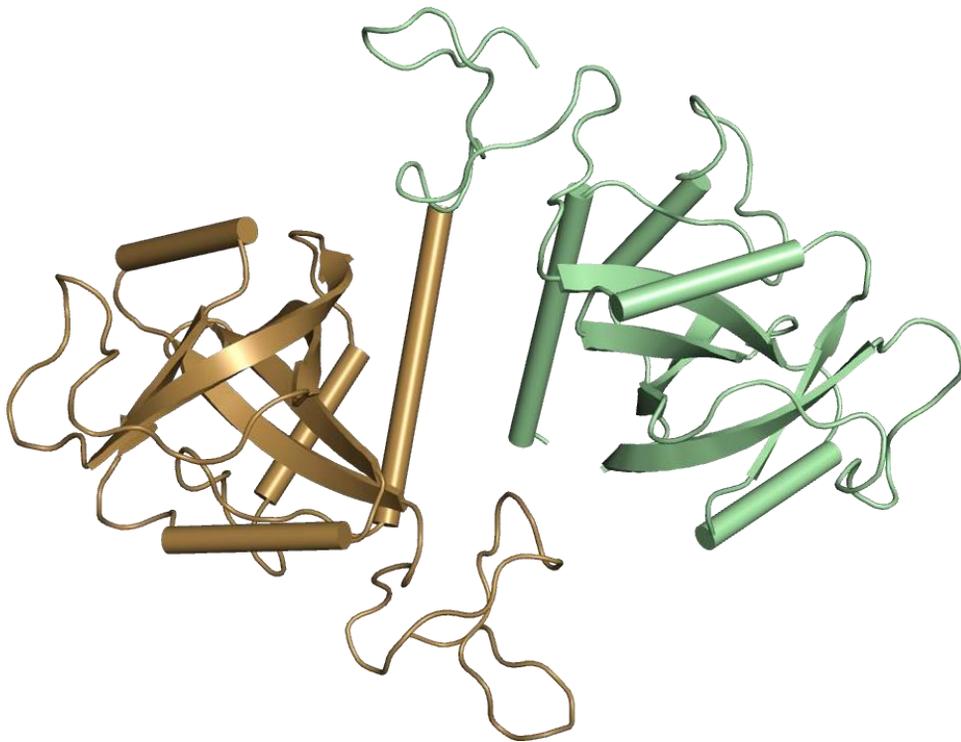


Fig. 63: Structure of Mmi1-YTH domain determined by molecular replacement from the crystals shown in Fig. 61

Table 4.5: Data processing, Phasing and Refinement for Mmi1-YTH crystal

<b><u>Data collection</u></b>	
Space group	C2
Unit cell parameters	104.6Å; 58.7Å; 67.6Å; 90°; 120°; 90°
Wavelength (Å)	0.980105
Resolution (Å)	45.05-2.38(2.69-2.38)
R <sub>merge</sub> (%)	15.1 (164.2)
<i>I</i> / $\sigma I$	6.4 (1.0)
Completeness (%)	98.9 (90.2)
CC <sub>1/2</sub> (%)	99.5 (47.4)
Redundancy	6.8
Observed reflections	96211
Unique reflections	14135
<b><u>Refinement</u></b>	
Resolution (Å)	45.09-2.38
R / R <sub>free</sub> (%)	18.9 / 24.8
<b><u>Number of atoms</u></b>	
Protein	2613
Cl <sup>-</sup> / Glycerol	1 / 6
Water	43
<b><u>R.m.s deviations</u></b>	
Bond lengths (Å)	0.01
Bond angles (°)	1.15

Among these two constructs of Mmi1-NTD, 1-350 had higher yields than 1-322. Therefore, this Mmi1 construct was selected for large scale protein purification.

#### ***4.6.2.2 Protein expression and purification of Mmi1-Erh1 complex:***

The Mmi1 NTD (1-350) and Erh1 proteins were produced in large culture by co-transforming the plasmid containing Mmi1 and the one containing Erh1 in a BL21 Gold cell. The protein production was induced by 250  $\mu$ M IPTG in 2YT media at 4 °C. A first purification step was performed by Ni-NTA affinity chromatography using the same protocol as for Mmi1 YTH domain, with the exception of elution buffer, 20 mM Tris-HCl, 200 mM NaCl, 400 mM Imidazole, 5 mM 2-mercaptoethanol. In the elution, two strong bands migrating at the expected molecular weight for Mmi1 and Erh1 were present (Fig. 64b), indicating that both proteins co-purify and most likely interact together. Furthermore, a washing step with buffer supplemented with 1 M NaCl did not dissociate the complex, indicating that the two proteins, Mmi1-NTD and Erh1 form a strong complex. However, the complex was not stable as Mmi1 kept precipitating over time while Erh1 remained soluble. Therefore, it was not possible to carry out further purification steps. Multiple trials were performed with different buffer composition but the complex could not successfully be purified.

#### ***4.7 Mmi1-Erh1-Mei2:***

Erh1 has been reported to co-localize with Mei2 but the exact nature of their interaction has not been explored yet. To see if Mei2 directly interacts with the Mmi1-Erh1 complex, a co-expression assay was performed with FL Mmi1, Mei2-CTD and Erh1. In their study, Yamamoto group showed that the Mei2-CTD (429-750) interacts with Mmi1. As discussed before, trials aimed at demonstrating a direct interaction between Mei2-CTD and Mmi1 failed, we then investigated whether Erh1 could be the missing link between these two proteins. The full length Mmi1 was cloned in a modified pET28b recombinant vector dedicated to the expression of proteins of interest fused to a His<sub>6</sub>-ZZ tag (described in details in section 5.3) at their N-terminal extremity.

The expression tests were performed only in BL21 Gold cells, at 18 °C. The elution fractions from Ni-NTA were analysed on SDS-PAGE gel (Fig. 65). Although Erh1 bands were present, significantly strong bands corresponding to molecular weight of His<sub>6</sub>-ZZ-Mmi1 and Mei2-CTD could not be detected, suggesting that Erh1 does not bridge Mei2-CTD and Mmi1.

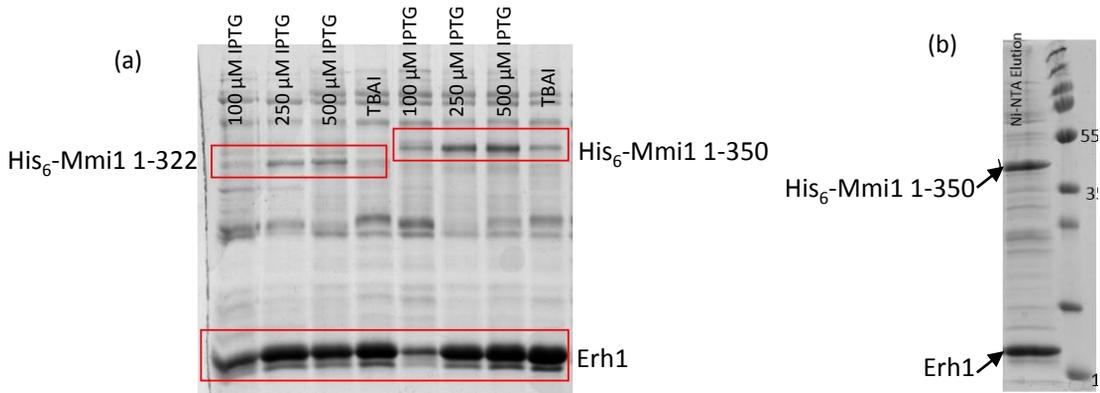


Fig. 64: Expression assay for Mmi1-Erh1-Complex at 18 °C, in Gold cells (a) SDS-PAGE analysis for the proteins present in the elution from Ni-NTA column (b) to check purity of His-Mmi1/Erh1 complex after purification from Ni-NTA.

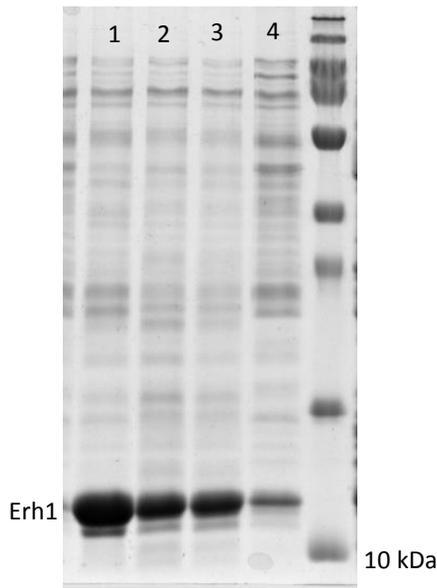


Fig. 65: 15% SDS-PAGE gel for co-expression check of His-ZZ-Mmi1, Erh1 and Mei2 proteins(429-750) at 18 °C, in Gold cells . Lanes 1: TBAI media, 2: 250 μM IPTG induction, 3: 100 μM IPTG induction, 4: no induction



Fig. 66: Erh1-crystals (left) 0.1M Citric Acid pH 4.0, 0.8M Ammonium Sulfate and (right) 0.4M Ammonium Sulfate, 0.1M Sodium Acetate pH3.8

#### ***4.8 Crystallization of Erh1:***

Preliminary crystallization trials for Erh1 were performed by sitting drop vapour diffusion method, with three different protein concentrations: 7.5mg/ml, 15mg/ml, 30mg/ml and two screens, JCSG plus (Molecular Dimension), Index (Hampton Research) at 4°C. Initial six-branched star crystals were obtained in 0.1M Citric acid, pH 4.0, 0.8M ammonium sulphate. Secondary optimization was performed by varying the ammonium sulphate concentration with pH, the precipitant was varied from 0.3-1.0M ammonium sulphate and pH was varied using two different buffers, 0.1M Citric acid (pH: 3.0, 3.5, 4.0, 4.5) and 0.1M sodium acetate (pH 3.8, 4.0, 4.2) with two different protein concentrations, 3.5mg/ml and 7.0mg/ml (Fig. 66).



## Formation of *S. pombe* Erh1 homodimer mediates gametogenic gene silencing and meiosis progression

Ditipriya Hazra<sup>1,#</sup>, Vedrana Andrić<sup>2,4,#</sup>, Benoit Palancade<sup>3</sup>, Mathieu Rougemaille<sup>2</sup> and Marc Graille<sup>1</sup>

<sup>1</sup>: BIOC, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, F-91128 Palaiseau, France.

<sup>2</sup>: Institut de Biologie Intégrative de la Cellule, CNRS, UMR9198, Université Paris-Saclay, 91190 Gif-sur-Yvette, France

<sup>3</sup>: Institut Jacques Monod, CNRS, UMR7592, Univ Paris-Diderot, Sorbonne Paris Cité, 75013 Paris, France

<sup>4</sup>: Université Paris-Saclay, 91190 Gif-sur-Yvette, France

#: equal contribution

Correspondence should be addressed to MR (mathieu.rougemaille@i2bc.paris-saclay.fr) and MG (marc.graille@polytechnique.edu).

### Abstract

Timely and accurate expression of the genetic information relies on the integration of environmental cues and the activation of regulatory networks involving transcriptional and post-transcriptional mechanisms. In fission yeast, meiosis-specific transcripts are selectively targeted for degradation during mitosis by the EMC complex, composed of Erh1, the ortholog of human ERH, and the YTH family RNA-binding protein Mmi1. Here, we present the crystal structure of Erh1 and show that it assembles as a homodimer. Mutations of amino acid residues to disrupt Erh1 homodimer formation result in loss-of-function phenotypes, similar to *erh1Δ* cells: expression of meiotic genes is derepressed in mitotic cells and meiosis progression is severely compromised. Interestingly, formation of Erh1 homodimer is dispensable for interaction with Mmi1, suggesting that only fully assembled EMC complexes consisting of two Mmi1 molecules bridged by an Erh1 dimer are functionally competent. We also show that Erh1 does not contribute to Mmi1-dependent down-regulation of the meiosis regulator Mei2, supporting the notion that Mmi1 performs additional functions beyond EMC. Overall, our results provide a structural basis for the assembly of the EMC complex and highlight its biological relevance in gametogenic gene silencing and meiosis progression.

## Introduction

Members of the ERH protein family are small proteins found in metazoan, invertebrates as well as plants. These proteins are strongly conserved (no amino acid changes between frog and human proteins and only one difference between human and zebrafish proteins), arguing for strict evolutionary constraints and for a highly important function. This gene was originally identified 25 years ago from a genetic screen as a mutant enhancing the truncated wing phenotypes of fruit flies lacking the rudimentary (*r*) gene <sup>1</sup>, which encodes for the enzyme catalyzing the first three steps of the pyrimidine biosynthesis pathway. However, the biological function of ERH remains unclear despite its strong abundance in tumors compared to human normal cells, making it a very interesting candidate for functional characterization <sup>2</sup>.

A bundle of evidences points towards a role of ERH in mRNA synthesis, maturation and nuclear export as this nuclear protein has been shown to interact with : (1) FCP1, the specific phosphatase for RNA PolII C-terminal domain <sup>3</sup>; (2) the transcription factor SPT5 <sup>3</sup>; (3) PDIP46/SKAR (now named POLDIP3, for Polymerase delta-interacting protein 3), which localizes to nuclear speckles, regions enriched in pre-mRNA splicing factors <sup>4, 5</sup>; (4) SNRPD3, a subunit of the Sm complex, which is involved in mRNA splicing. ERH also interacts with CIZ1, a zinc finger protein acting as a DNA replication factor and present in replication foci <sup>6, 7</sup>. ERH is necessary for chromosome segregation during mitosis, probably through its role on CENP-E mRNA splicing <sup>8, 9</sup>. Indeed, in the absence of ERH, the CENP-E mRNA, encoding a kinetochore protein, is incorrectly spliced and pre-mRNAs are rapidly eliminated by the nonsense-mediated mRNA decay pathway <sup>9</sup>. Such link between splicing defects, cell cycle arrest and mitotic defects has already been observed for the depletion of other splicing factors <sup>10</sup>. In *Xenopus*, ERH has been shown to act as a transcriptional repressor and to interact with DCoH/PCD (dimerization cofactor of HNF1/pterin-4a-carbinolamine dehydratase), a positive cofactor of the HNF1 homeobox transcription factor, by yeast two-hybrids <sup>11</sup>. ERH protein is absent in *S. cerevisiae* yeast but expression of human ERH in budding yeast stimulates filamentous growth in low nitrogen media <sup>12</sup>. Interestingly, this phenotype is reminiscent of the phenotype observed upon expression of the RBP7 subunit of the human RNA polymerase II in yeast <sup>13</sup>, arguing again for a potential role of ERH proteins in the control of mRNA metabolism.

Closely related proteins, sharing around 30% sequence identity with human ERH, are also present in *Schizosaccharomyces* such as *S. pombe* and in few other fungi<sup>14</sup>. Recent studies performed in *S. pombe* have enlightened the role of Erh1, the ortholog of human ERH. Initially, the *ERH1* gene was identified as a suppressor of *sme2Δ* phenotype, *i.e.* a meiotic arrest due to the lack of inactivation of Mmi1 during meiosis<sup>15</sup>. Mmi1 is a YTH-family protein<sup>16</sup>, which selectively recognizes RNA hexanucleotide motifs (*e.g.* UNAAAC) present in meiotic transcripts and triggers their nuclear retention and elimination by the nuclear exosome during mitosis<sup>17</sup>. Upon meiosis onset, Mmi1 is sequestered in a nuclear dot by the *sme2/mei* long noncoding RNA, which is assisted by the master regulator of meiosis Mei2<sup>18</sup>.

Recent works showed that Erh1 and Mmi1 form a 2:2 stoichiometric complex dubbed EMC (for Erh1-Mmi1 complex) whereby two Mmi1 peptides are physically bridged by an Erh1 homodimer<sup>17, 19, 20</sup>. EMC localizes to scattered nuclear foci in vegetative cells and associates with two distinct complexes<sup>19, 21</sup>. The first one known as MTREC (for Mtl1-Red1 core) is composed of the zinc-finger protein Red1, the Mtr4-like RNA helicase Mtl1 and Pir1/Iss10 among others subunits<sup>15, 19</sup>. MTREC cooperates with Mmi1 to mediate degradation of meiotic mRNAs by recruiting the Rrp6 subunit of the nuclear exosome<sup>22</sup>. The second complex known to interact with EMC is the CCR4-NOT complex but despite its known function as a mRNA deadenylase, it is not involved in Mmi1-dependent meiotic mRNA clearance<sup>19, 21, 23, 24, 25</sup>. Instead, it is required for the integrity of heterochromatin and regulates the abundance of Mei2 protein during mitosis through the action of its Mot2/Not4 E3 ubiquitin ligase subunit<sup>21</sup>. Interestingly, both MTREC (PAXT in human cells) and CCR4-NOT complexes are conserved in human cells, suggesting that ERH may also interact with these complexes in human cells. Furthermore, human ERH can partially rescue the sensitivity to sorbitol but neither SDS nor hydroxyurea of *S. pombe erh1Δ* cells<sup>14</sup> indicating that a partially conserved function between Erh1 and human ERH proteins.

Here, we describe the crystal structure of *S. pombe* Erh1 protein and compare it to the structures of metazoan ERH proteins that have already been solved as well as to the structure of the *S. pombe* Erh1-Mmi1 complex that has been solved while this work was in progress<sup>20</sup>. We observe that Erh1 organizes as a homodimer in which the two monomers contact each other via

hydrophobic interactions, consistent with recent work <sup>20</sup>. Structure-guided mutational analysis shows that formation of Erh1 homodimer is critical for cell growth at low temperatures and for its functions in meiotic mRNA degradation and meiosis progression. Interestingly, an Erh1 mutant (Erh1<sup>H11R,L13R</sup>) defective for dimerization still associates with Mmi1 *in vivo*, suggesting that Erh1 monomer is sufficient for interaction with Mmi1 while formation of Erh1 dimer is essential for EMC function. We also show that Erh1 does not contribute to the Mmi1-dependent down-regulation of Mei2 in mitotic cells, indicating that Mmi1 exerts functions beyond its partnership with Erh1. Overall, our results provide a structural basis for Erh1 dimerization and underscore the biological importance of Erh1 homodimer formation during both the mitotic and meiotic cell cycles.

## Results and discussion

### *S. pombe* Erh1 crystal structure

Initial polycrystals of Erh1 protein with a 6-branches star shape were obtained from an initial large screen of crystallization conditions in the following condition (0.8 M ammonium sulfate; 0.1 Na citrate pH 4). Thanks to the use of a micro-focus beamline, a complete dataset of moderate quality could be collected by shooting on a single branch of the star. Larger crystals could be obtained by increasing the drop volume, varying the ammonium sulfate concentration and the buffer. From one of these crystals, we could collect a dataset of better quality (see Table 1 for dataset statistics) from which we could determine Erh1 structure by molecular replacement using the structure of human ERH as initial model and refine it to 1.95Å resolution.

Erh1 monomer folds as an  $\alpha$ - $\beta$  protein composed of a four stranded anti-parallel  $\beta$ -sheet and three  $\alpha$  helices packed onto the same face of the  $\beta$ -sheet (Fig. 1A). This fold is similar to that of metazoan ERH (rmsd values of 1.1-1.9Å over 80 C $\alpha$  atoms; <sup>12, 26, 27, 28</sup>). The loop connecting helices  $\alpha$ 1 and  $\alpha$ 2 is not visible in our structure of Erh1. This most likely results from the intrinsic flexibility of this loop as shown for the corresponding region of metazoan ERH proteins by X-ray crystallography or NMR <sup>12, 26, 27, 28</sup>. Three copies (protomers A, B and C) of Erh1 protein are present in the asymmetric unit. Protomers A and B are virtually identical (rmsd values of 0.27Å over 84 C $\alpha$  atoms) while protomer C slightly differs as illustrated by its higher rmsd value when compared to the two other chains (1.3-1.4Å over 84 C $\alpha$  atoms; Fig. S1A). The largest differences between protomer C and the two other Erh1 molecules present in the asymmetric unit occur at helix  $\alpha$ 2 (2.5Å translation along the helix longitudinal axis), at the hinge between helix  $\alpha$ 2 and the N-terminal extremity of strand  $\beta$ 3, and to a lesser extent on helix  $\alpha$ 3 (1.3Å translation). It is noteworthy that human ERH adopts a conformation similar to that observed for Erh1 protomer C.

### An evolutionary conserved homodimer

Among the three copies present in the asymmetric unit, protomers B and C associate to form a tight homodimer with a butterfly-like shape (Fig. 1B) while the protomer A forms a similar

homodimer with a symmetry-related molecule (rmsd of 1.1Å over 160 Cα atoms). This dimeric state is consistent with the elution volume determined by size-exclusion chromatography (Fig. 1C). In the homodimer, the Erh1 β-sheet face from each monomer, which is not packed against α helices, interacts to form a β-barrel (Fig. 1B). Each monomer engages an area of 850 Å<sup>2</sup> mostly formed by hydrophobic residues, which are strongly or strictly conserved within Erh1/ERH orthologues from fungi, plants, insects, worm and animals (Fig. 1D). This rationalizes that the Erh1 homodimer is reminiscent of those observed for metazoan ERH proteins (rmsd 1.4Å over 160 Cα atoms; <sup>12, 26, 28</sup>).

An evolutionary highly conserved region present on the side of the homodimerization surface has been shown to participate in the interaction with Mmi1 (a short region encompassing residues 95-122; <sup>20</sup>) while this work was in progress (Fig. 1E). Comparison between the Erh1-Mmi1 and our apo-Erh1 structures reveals that protomer C is strongly similar to the Erh1-bound structure and hence compatible with Mmi1 binding. On the contrary, significant differences support that protomers A and B are incompatible with Mmi1 binding (Fig. S1B-C). Most interestingly, the crystal structure of Erh1-Mmi1 complex reveals a 2:2 stoichiometry <sup>20</sup>, confirming a previously proposed model of Erh1-mediated Mmi1 self-interaction <sup>17</sup>.

As in this complex, each Mmi1-[95-122] peptide interacts with Erh1 on a region spread at the homodimer interface (Fig. S1B), we have decided to investigate the functional role of Erh1 homodimerization. We then simultaneously mutated two residues located at the homodimer interface (Ile11 and Leu13) into Arg to generate the Erh1<sub>I11R,L13R</sub> double mutant with the aim of disrupting Erh1 homodimer. First, we have purified this mutant upon co-expression in *E. coli*. During purification, the Erh1<sub>I11R,L13R</sub> mutant proved much less stable than the wild-type protein and had a tendency to aggregate as demonstrated by the presence of a large peak containing Erh1 and eluting in the void volume (7-8 mL) of a S75 size-exclusion chromatography (Fig. 1C). Furthermore, the 2 mL delay in the elution volume of a fraction of this Erh1<sub>I11R,L13R</sub> double mutant (13.1 mL) compared to wild-type protein (11.1 mL) clearly indicates that the Erh1<sub>I11R,L13R</sub> double mutant is monomeric in solution as anticipated (Fig. 1C).

To obtain additional insights into EMC complex assembly, we sought to determine whether Erh1 dimerization contributes to its association with Mmi1 *in vivo*. We generated *erh1Δ* strains

expressing GFP-tagged versions of the dimeric wild-type Erh1 or the monomeric Erh1<sub>I111R,L13R</sub>. Analysis of total protein levels under denaturing conditions indicated that the two forms of Erh1 were similarly expressed (Fig. 1F). When preparing native extracts for co-immunoprecipitation assays, however, we observed a significant decrease (roughly 6-fold less) in the total amount of Erh1<sub>I111R,L13R</sub> when compared to wild-type Erh1 (Fig. 1G, WCE panel), likely as a consequence of decreased stability and in agreement with our observations during the purification of this mutant protein. Yet, surprisingly, Mmi1 still associated with Erh1<sub>I111R,L13R</sub> (Fig. 1G, IP TAP panel), suggesting that Erh1 dimerization is not a prerequisite for its association with Mmi1, contrary to what could be assumed from the crystal structure of Erh1-Mmi1-[95-122] complex<sup>20</sup>.

### **Formation of Erh1 homodimer is required for gametogenic gene silencing**

To investigate the role of Erh1 dimerization *in vivo*, we first analyzed its impact on *S. pombe* cell growth. Similar to *erh1Δ* cells, the Erh1<sub>I111R,L13R</sub> mutant displayed growth defects at all tested temperatures, especially 23°C, indicating that Erh1 dimerization is essential for its function (Fig. 2A).

In vegetative cells, Erh1 and Mmi1 assemble in the EMC complex that localizes to nuclear foci<sup>17, 19, 20</sup>. Live cell microscopy experiments revealed lower and diffuse GFP-Erh1<sub>I111R,L13R</sub> signals when compared to GFP-Erh1 (Fig. 2B). Strikingly, nuclear dots were lost in the mutant, including in cells in which the GFP signal was similar to that of the wild type (Fig. 2B, red arrows). This supports the notion that Erh1 homodimer formation is a prerequisite for its confinement into nuclear bodies.

Erh1 cooperates with Mmi1 to target meiosis-specific transcripts for degradation by the nuclear exosome<sup>17, 19, 20</sup>. To evaluate the role of Erh1 homodimer in this pathway, we measured the levels of Mmi1 RNA targets by RT-qPCR, including the *mei4+* and *ssm4+* meiotic mRNAs as well as the lncRNA *meiRNA*. Cells lacking Erh1 or expressing Erh1<sub>I111R,L13R</sub> showed a strong accumulation of *mei4+* and *ssm4+* transcripts, while *meiRNA* levels were only partially increased when compared to the *mmi1Δ* mutant (Fig. 2C). These results indicate that Erh1 dimerization is required for efficient Mmi1-dependent meiotic RNA degradation. To determine

whether this was due to defective recruitment of Mmi1 to its targets, we analyzed the levels of meiotic RNAs co-precipitated with Mmi1. In cells expressing wild-type Erh1, Mmi1 efficiently bound to the *mei4+*, *ssm4+* and meiRNA transcripts (Fig. 2D). Instead, in the Erh1<sub>IIIR,L13R</sub> mutant, the association of Mmi1 to meiotic mRNAs was abolished, while meiRNA still co-precipitated, albeit to a lower extent (Fig. 2D). Altogether, these data indicate that formation of Erh1 homodimer and hence proper EMC assembly is crucial for meiotic mRNAs recognition and degradation by Mmi1 but dispensable in the case of meiRNA. This different requirement for Erh1 dimer in the binding of Mmi1 to its RNA substrates might relate to the number of Mmi1 binding motifs (i.e. UNAAAC) within transcripts. Given *mei4+* and *ssm4+* mRNAs contain 8 and 7 binding sites, respectively, while meiRNA has up to 25<sup>29</sup>, it is possible that a pool of Mmi1 associates with the latter even in the absence of properly assembled EMC.

We previously showed that Mmi1 recruits the Ccr4-Not complex to promote ubiquitinylation and down-regulation of its own inhibitor Mei2, a master regulator of meiosis<sup>21</sup>. To determine whether Erh1 also contributes to this regulatory circuit, we analyzed Mei2 levels in cells lacking Erh1 or expressing Erh1<sub>IIIR,L13R</sub>. We found that, contrary to *mmi1Δ* cells, Mei2 levels were not increased in mutants (Fig. S2), indicating that Erh1 does not contribute to Mei2 down-regulation in mitotic cells. This is also consistent with the notion that Mmi1 can exert functions independently of Erh1, as in case of transcription termination<sup>20,30</sup>.

### **Formation of Erh1 homodimer contributes to meiosis progression**

Erh1 not only suppresses the meiotic program in vegetative cells but also stimulates meiosis progression<sup>17,19,20</sup>. To examine the requirement for Erh1 dimerization in meiosis, homothallic *erh1Δ* cells expressing wild-type Erh1 or Erh1<sub>IIIR,L13R</sub> were exposed to iodine vapor, which stains the spore wall with dark color. Cells expressing wild-type Erh1 displayed strong staining intensity and high sporulation frequency, as indicated by the prevalence of asci (Fig. 3A). On the contrary, cells lacking Erh1 or expressing the Erh1<sub>IIIR,L13R</sub> mutant showed reduced intensity in staining, consistent with lower sporulation efficiency and asci formation (Fig. 3A). However, meiosis was not completely abolished as in *mei4Δ mmi1Δ* cells, suggesting that the absence of Erh1 might be bypassed at least at low frequency.

Upon entry into meiosis, Mmi1 and Erh1 foci converge to a single nuclear dot associated with the meiosis regulator Mei2 and the lncRNA meiRNA<sup>18, 19, 31</sup>. To determine whether Erh1 homodimer is necessary for dot formation, we probed meiRNA in meiotic cells by fluorescence *in situ* hybridization. Consistent with previous studies<sup>18, 19, 31</sup>, meiRNA colocalized with wild-type GFP-tagged Erh1 in a unique nuclear dot (Fig. 3B). On the contrary, cells expressing GFP-Erh1<sup>H11R,L13R</sup> failed to form both meiRNA and Erh1 dots (Fig. 3B). From these experiments, we propose that defects in Erh1 homodimer formation prevent the assembly of the EMC-meiRNA-Mei2 nuclear dots required for Mmi1 sequestration/inactivation during meiosis, thereby rationalizing impaired meiosis progression.

## Conclusion

In this study, we report the crystal structure of Erh1 and show that it assembles as a homodimer through hydrophobic interactions between N-terminal residues. Consistent with recent work<sup>20</sup>, the structure also reveals a highly conserved region lying on the side of the dimerization domain and to which Mmi1 associates to form the heterotetrameric EMC complex. The strong similarity between Erh1 and human ERH homodimers (this study,<sup>20, 26, 28</sup>) raises the possibility that the assembly of ERH-based multimeric complexes has been maintained throughout evolution for regulatory purposes, including modulation of gene expression. Whether Mmi1-related, YTH family RNA-binding proteins or other factors bind to ERH homodimers in metazoans to regulate cellular processes remains to be investigated.

Our functional analyses also highlight the biological relevance of Erh1 homodimer formation in gametogenic gene silencing. Importantly, Erh1 dimerization is dispensable for interaction with Mmi1 but essential for Mmi1 binding to meiotic transcripts, implying that RNA recognition by the YTH domain is not sufficient *per se* and that only fully assembled EMC complexes are functionally competent. This illustrates the cooperation between the C-terminal YTH domain and the N-terminal disordered region of Mmi1, to which Erh1 associates, for optimal binding to RNA substrates. The mechanistic basis for this is presently unclear but it is possible that the assembly of large macromolecular machines (*e.g.* EMC, MTREC, Ccr4-Not) facilitates protein-RNA interactions. In line with this, it is tempting to speculate that EMC and meiotic mRNA nuclear

foci observed in vegetative cells<sup>17, 19</sup> may reflect multimerization driven by Erh1 dimers of such RNP complexes, gathering multiple transcripts in close proximity for efficient degradation. Another important finding from our work is the requirement for Erh1 dimerization in Mmi1 sequestration/inactivation by the meiRNA-Mei2 dot during meiosis. The residual binding of Mmi1 to meiRNA observed in Erh1<sup>I11R,L13R</sup> mitotic cells may not be sufficient for proper dot formation in meiosis. Therefore, Erh1 homodimer formation not only promotes Mmi1 function during vegetative growth but also contributes to its inhibition during meiosis. This cell-cycle dependent duality in functional outcomes may allow rapid changes in the activity of the complex without altering its expression levels or assembly. Whether this relates to the nature of the RNA substrate (meiotic mRNAs versus meiRNA) and/or additional factors (*e.g.* Mei2) remains to be determined. Regardless the precise mechanism, our study opens new perspectives to study the formation and activity of ERH homodimer-containing complexes in metazoan and to understand their relevance to human diseases<sup>10</sup>.

## Materials and methods

### *Cloning and protein expression*

The gene encoding for Erh1 was amplified using a *S. pombe* cDNA library by PCR with oligonucleotides oMG511 and oMG512 (Table S1) using the Phusion High-Fidelity DNA Polymerase (Thermo) according to manufacturer's instructions. The PCR products were further cloned into pGEX-6P1 vector using Fast Digest *Bam*HI and *Xho*I restriction enzymes (Thermo Fisher Scientific) to generate the plasmid pMG921 encoding for a GST-tag fused to the N-terminal extremity of the full-length Erh1 by a 3C protease cleavage site (Table S1). The plasmid encoding for the Erh1<sub>I111R,L13R</sub> double mutant was obtained by one-step site-directed mutagenesis of pMG921 using oligonucleotides oMG629/oMG630 (Table S1) to yield plasmid pMG945.

The Erh1 protein was expressed in *E. coli* BL21 (DE3) Codon+ cells upon transformation with pMG921 plasmid. Cultures were performed in 1 L of auto-inducible terrific broth media (ForMedium AIMTB0260) containing ampicillin (100 µg/mL) and chloramphenicol (25 µg/mL) first for 3 hours at 37°C and then overnight at 18°C. The cells were harvested by centrifugation at 4100 rcf for 45 minutes. The pellet was resuspended in 30 mL lysis buffer (20 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM β-mercaptoethanol) in the presence of 100 µM PMSF.

Cell lysis was performed by sonication on ice, followed by lysate clearance by centrifugation at 20000 rcf for 45 minutes. The supernatants were applied on GSH-sepharose resin pre-equilibrated with lysis buffer. After extensive washing steps with lysis buffer and with a high salinity buffer (20 mM Tris-HCl pH 7.5, 2 M NaCl, 5 mM β-mercaptoethanol), the protein was eluted with elution buffer (20 mM Tris-HCl pH 7.5, 200 mM NaCl, 20 mM GSH, 5 mM β-mercaptoethanol). The eluted protein was next incubated overnight at 4°C with GST-3C protease under dialysis conditions in lysis buffer and then passed through GSH column to remove the GST-tag as well as the GST tagged 3C protease. The unbound proteins were subjected to size exclusion chromatography using HiLoad 16/60 Superdex 75 column (GE Healthcare Biosciences) pre-equilibrated with lysis buffer on an ÄKTA Purifier system (GE Healthcare Biosciences). The Erh1<sub>I111R,L13R</sub> double mutant was purified using the same protocol.

### *Crystallization, data collection and structure determination*

Crystallization conditions were screened by the sitting-drop vapor diffusion method using JCSG+ screen (Molecular Dimensions) at 4°C by mixing 150 nL of concentrated protein (7.5mg/ml) solution with an equal volume of reservoir solution in a 96-wells TTP plates (TTPlabtech). Initial hits corresponding to star shaped crystals were obtained in 0.8M ammonium sulfate; 0.1 Na citrate pH 4. For crystal optimization, hanging-drop method was used at 4°C, by mixing 1 µL of concentrated protein with 1 µL of reservoir solution. The best dataset was collected from crystals obtained in 0.3 M ammonium sulfate; 0.1 M Na acetate pH 3.8.

Prior to data collection, the crystals were quick-soaked in cryo-protectant solutions containing 15% (v/v) and then 30% ethylene glycol in corresponding well solutions and flash-frozen in liquid nitrogen. X-ray diffraction datasets were collected on both Proxima-1 and Proxima-2a beamlines at synchrotron SOLEIL (Saint-Aubin, France) and were processed with the XDS package<sup>32</sup>. The dataset collected on Proxima-1 showed a higher resolution limit and was then used to determine and refine the structure of Erh1 protein (Table 1). The structure was solved by molecular replacement searching for 3 molecules in the asymmetric unit with the program PHASER<sup>33</sup>. The initial model for molecular replacement was generated by the PHYRE2 server<sup>34</sup> using the crystal structure of human ERH (30% sequence identify) as template<sup>28</sup>. In this model, the loop corresponding to residues 44 to 55 was removed as it is known to be highly flexible from the comparison of human and fruit fly ERH structures<sup>12, 26, 28</sup>.

The final model was obtained by iterative cycles of building and refinement using COOT<sup>35</sup> and BUSTER<sup>36</sup>, respectively (for final statistics, see Table 1). This model encompasses residues 6-46 and 55-100 for protomer A, 1-46 and 55-98 for protomer B (as well as 4 residues from the N-terminal tag) and 1-2, 7-47 and 54-98 for protomer C as well as 5 ethylene glycol molecules, 1 sulphate ion, 1 acetate molecule and 73 water molecules.

The atomic coordinates and structure factors have been deposited into the Brookhaven Protein Data Bank under the accession numbers 6S2W.

### *S. pombe strains and growth media*

The *S. pombe* strains used in this study are listed in Table S2. Strains were generated by transformation with a lithium acetate-based method. The *mmi1*Δ cells were generated from a parental strain possessing a deletion of *mei4+*, since the absence of Mmi1 leads to severe growth and viability defects due to the deleterious expression of Mei4. All experiments were performed using minimal medium (EMM Broth, Formedium, #PMD0210) supplemented with 150 mg/L of each adenine (Sigma, #A2786), L-histidine (Sigma, #H8000), uracil (Sigma, #U750) and L-lysine (Sigma, #L5501) but lacking L-leucine (EMM-LEU). To assess mating/sporulation efficiency, cells plated on EMM-LEU medium for 5 days at 30°C were exposed to iodine crystals (Sigma, #326143) for 5 min at room temperature.

#### *Co-immunoprecipitation and Total protein analyses*

Experiments were performed as described in Simonetti et al, 2017<sup>21</sup> except that detection was done with a Vilber Lourmat Fusion Fx7 imager.

#### *RNA extraction and RT-qPCR*

Experiments were performed as described in Simonetti et al, 2017<sup>21</sup> except that 100 units Maxima Reverse Transcriptase (Thermo Scientific, #EP0743) were used for reverse transcription reactions.

Oligonucleotides used in qPCR reactions are listed in Table S3.

#### *RNA-immunoprecipitation*

Experiments were performed as described in Simonetti et al, 2017<sup>21</sup> with the following modifications: 40 ODs of cells were grown to mid-log phase at 30°C in EMM-LEU and cross-linked with 0.2% formaldehyde for 20 min. Following quenching with 250 mM glycine for 5 min, cells were harvested by centrifugation. Cell pellets were resuspended in 2 ml RIPA buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 2 mM EDTA, 2 mM benzamidine, 1X Roche complete EDTA-free protease inhibitor cocktail and 80 U RNaseOUT Ribonuclease inhibitor (Invitrogen, #10777-019)) to make “pop-corn”. Lysis was performed using a Ball Mill (Retsch, MM400) for 15 min at 15 Hz frequency. Extracts were

cleared by centrifugation before immunoprecipitation with 1 mg of pre-washed rabbit IgG-conjugated M-270 Epoxy Dynabeads (Invitrogen, #14311D) for 1 hour at 4°C. Beads were then washed once with low salt buffer (10 mM Tris-HCl pH7.5, 150 mM NaCl, 0.5% Triton X-100), twice with high salt buffer (10 mM Tris-HCl pH7.5, 1 M NaCl, 0.5% Triton X-100) and once again with low salt buffer for 10 min at room temperature. Total and immunoprecipitated RNAs were decrosslinked at 70°C for 45 min in the presence of reverse buffer (10 mM Tris-HCl pH 6.8, 5 mM EDTA, 10 mM DTT, 1% SDS) and treated with proteinase K for 30 min at 37°C. RNAs samples were next extracted with phenol:chloroform 5:1 pH4.7 (Sigma, #P1944), precipitated with ethanol and treated with DNase (Ambion, #AM1906) prior to RT-qPCR analyses.

### *SmFISH*

Quasar 670-labeled meiRNA probes were designed using Stellaris Probe Designer tool (Table S4) and synthesized by Biosearch Technologies. Single molecule RNA Fluorescence In-Situ Hybridization (smFISH) was performed according to the manufacturer's protocol (Biosearch Technologies) with minor modifications.

Vegetative cells were plated on EMM-LEU and grown for 3 days at 30°C. Cells were then resuspended in 1X PBS containing 3.7% formaldehyde to an OD<sub>600nm</sub> of 0.3, treated with Zymolyase 100 T for cell wall digestion and permeabilized in 70% ethanol prior to over-night incubation with meiRNA probes. Stellaris RNA FISH hybridization and wash buffers were obtained from Biosearch Technologies. DAPI stained cells were resuspended in Vectashield antifade mounting medium (Vector laboratories) and imaged using DM6000B Leica microscope with a 100X, numerical aperture 1.4 (HCX Plan-Apo) oil immersion objective and a charge-coupled device (CCD) camera (CoolSNAP HQ; Photometrics). Optical Z sections (0.2 µm step size, 25 sections) were acquired using a piezo-electric motor (LVDT; Physik Instrument) and the MetaMorph 6.1 software prior to maximum-intensity projection into a single plane. Images were processed in ImageJ (NIH).

## References

1. Wojcik E, Murphy AM, Fares H, Dang-Vu K, Tsubota SI. Enhancer of rudimentaryp1, e(r)p1, a highly conserved enhancer of the rudimentary gene. *Genetics* **138**, 1163-1170 (1994).
2. Zafrakas M, Losen I, Knuchel R, Dahl E. Enhancer of the rudimentary gene homologue (ERH) expression pattern in sporadic human breast cancer and normal breast tissue. *BMC cancer* **8**, 145 (2008).
3. Amente S, *et al.* Identification of proteins interacting with the RNAPII FCP1 phosphatase: FCP1 forms a complex with arginine methyltransferase PRMT5 and it is a substrate for PRMT5-mediated methylation. *FEBS letters* **579**, 683-689 (2005).
4. Smyk A, Szuminska M, Uniewicz KA, Graves LM, Kozlowski P. Human enhancer of rudimentary is a molecular partner of PDIP46/SKAR, a protein interacting with DNA polymerase delta and S6K1 and regulating cell growth. *The FEBS journal* **273**, 4728-4741 (2006).
5. Ma XM, Yoon SO, Richardson CJ, Julich K, Blenis J. SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs. *Cell* **133**, 303-313 (2008).
6. Coverley D, Marr J, Ainscough J. Ciz1 promotes mammalian DNA replication. *Journal of cell science* **118**, 101-112 (2005).
7. Lukasik A, Uniewicz KA, Kulis M, Kozlowski P. Ciz1, a p21 cip1/Waf1-interacting zinc finger protein and DNA replication factor, is a novel molecular partner for human enhancer of rudimentary homolog. *The FEBS journal* **275**, 332-340 (2008).
8. Fujimura A, Kishimoto H, Yanagisawa J, Kimura K. Enhancer of rudimentary homolog (ERH) plays an essential role in the progression of mitosis by promoting mitotic chromosome alignment. *Biochemical and biophysical research communications* **423**, 588-592 (2012).
9. Weng MT, *et al.* Evolutionarily conserved protein ERH controls CENP-E mRNA splicing and is required for the survival of KRAS mutant cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E3659-3667 (2012).
10. Weng MT, Luo J. The enigmatic ERH protein: its role in cell cycle, RNA splicing and cancer. *Protein & cell* **4**, 807-812 (2013).
11. Pogge von Strandmann E, Senkel S, Ryffel GU. ERH (enhancer of rudimentary homologue), a conserved factor identical between frog and human, is a transcriptional repressor. *Biological chemistry* **382**, 1379-1385 (2001).

12. Jin T, Guo F, Serebriiskii IG, Howard A, Zhang YZ. A 1.55 Å resolution X-ray crystal structure of HEF2/ERH and insights into its transcriptional and cell-cycle interaction networks. *Proteins* **68**, 427-437 (2007).
13. Khazak V, Sadhale PP, Woychik NA, Brent R, Golemis EA. Human RNA polymerase II subunit hsRPB7 functions in yeast and influences stress survival and cell morphology. *Molecular biology of the cell* **6**, 759-775 (1995).
14. Krzyzanowski MK, Kozłowska E, Kozłowski P. Identification and functional analysis of the *erh1(+)* gene encoding enhancer of rudimentary homolog from the fission yeast *Schizosaccharomyces pombe*. *PLoS one* **7**, e49059 (2012).
15. Yamashita A, Takayama T, Iwata R, Yamamoto M. A novel factor Iss10 regulates Mmi1-mediated selective elimination of meiotic transcripts. *Nucleic acids research* **41**, 9680-9687 (2013).
16. Hazra D, Chapat C, Graille M. m(6)A mRNA Destiny: Chained to the rhYTHm by the YTH-Containing Proteins. *Genes* **10**, (2019).
17. Shichino Y, Otsubo Y, Kimori Y, Yamamoto M, Yamashita A. YTH-RNA-binding protein prevents deleterious expression of meiotic proteins by tethering their mRNAs to nuclear foci. *eLife* **7**, (2018).
18. Harigaya Y, *et al.* Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature* **442**, 45-50 (2006).
19. Sugiyama T, *et al.* Enhancer of Rudimentary Cooperates with Conserved RNA-Processing Factors to Promote Meiotic mRNA Decay and Facultative Heterochromatin Assembly. *Molecular cell* **61**, 747-759 (2016).
20. Xie G, *et al.* A conserved dimer interface connects ERH and YTH family proteins to promote gene silencing. *Nature communications* **10**, 251 (2019).
21. Simonetti F, Candelli T, Leon S, Libri D, Rougemaille M. Ubiquitination-dependent control of sexual differentiation in fission yeast. *eLife* **6**, (2017).
22. Sugiyama T, Sugioka-Sugiyama R. Red1 promotes the elimination of meiosis-specific mRNAs in vegetatively growing fission yeast. *The EMBO journal* **30**, 1027-1039 (2011).
23. Cotobal C, *et al.* Role of Ccr4-Not complex in heterochromatin formation at meiotic genes and subtelomeres in fission yeast. *Epigenetics & chromatin* **8**, 28 (2015).

24. Stowell JAW, Webster MW, Kogel A, Wolf J, Shelley KL, Passmore LA. Reconstitution of Targeted Deadenylation by the Ccr4-Not Complex and the YTH Domain Protein Mmi1. *Cell reports* **17**, 1978-1989 (2016).
25. Ukleja M, *et al.* The architecture of the Schizosaccharomyces pombe CCR4-NOT complex. *Nature communications* **7**, 10433 (2016).
26. Arai R, *et al.* Crystal structure of an enhancer of rudimentary homolog (ERH) at 2.1 Angstroms resolution. *Protein science : a publication of the Protein Society* **14**, 1888-1893 (2005).
27. Li H, *et al.* Solution structure of the mouse enhancer of rudimentary protein reveals a novel fold. *Journal of biomolecular NMR* **32**, 329-334 (2005).
28. Wan C, Tempel W, Liu ZJ, Wang BC, Rose RB. Structure of the conserved transcriptional repressor enhancer of rudimentary homolog. *Biochemistry* **44**, 5017-5023 (2005).
29. Yamashita A, *et al.* Hexanucleotide motifs mediate recruitment of the RNA elimination machinery to silent meiotic genes. *Open biology* **2**, 120014 (2012).
30. Touat-Todeschini L, *et al.* Selective termination of lncRNA transcription promotes heterochromatin silencing and cell differentiation. *The EMBO journal* **36**, 2626-2641 (2017).
31. Shichino Y, Yamashita A, Yamamoto M. Meiotic long non-coding meiRNA accumulates as a dot at its genetic locus facilitated by Mmi1 and plays as a decoy to lure Mmi1. *Open biology* **4**, 140022 (2014).
32. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* **26**, 795-800 (1993).
33. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674 (2007).
34. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10**, 845-858 (2015).
35. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
36. Bricogne G, *et al.* BUSTER version 2.10.2. (ed<sup>^</sup>(eds) (2016).
37. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research* **42**, W320-324 (2014).

38. Ashkenazy H, *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research* **44**, W344-350 (2016).

## **Acknowledgments**

We acknowledge SOLEIL for provision of synchrotron radiation facilities and we would like to thank the beamline scientists for their assistance in using beamlines Proxima-1 and Proxima-2a.

We are indebted to Dr Y. Mechulam for his help with data collection and to Pr. C. Gaillardin (INA-PG, Thiverval-Grignon, France) for the gift of the *S. pombe* cDNA library.

This work was supported by Ecole Polytechnique, the Institut de Biologie Intégrative de la Cellule, the Centre National pour la Recherche Scientifique and the Agence Nationale pour la Recherche [grants ANR-16-CE11-0003 to M.G and ANR-16-CE12-0031-01 to M.R.]. D.H. was supported by a PhD fellowship from the French Ministère de l'Enseignement Supérieur et de la Recherche (MESR).

## **Authors' Contributions**

DH and VA performed experiments. DH, VA, BP, MR and MG designed the study, analyzed the data, drafted the article and approved the version to be published.

## **Competing Interests**

We have no competing interests.

## Legends to figures.

### **Figure 1: *S. pombe* Erh1 structure.**

- A. Cartoon representation of Erh1 monomer. The protein is colored from blue (N-terminus) to red (C-terminus). The loop encompassing residues 47 to 54, which not defined in electron density maps, probably due to high flexibility, is depicted as a dashed line.
- B. Homodimer representation of Erh1. The Ile11 and Leu13 residues mutated in this study are shown as sticks.
- C. Analytical S75 size-exclusion chromatography profile of WT (solid line) and I11R/L13R double mutant (dashed line) Erh1 proteins.
- D. Sequence alignment of Erh1/ERH proteins from *Schizosaccharomyces* fungi (*S. pombe*, *S. cryophilus*, *S. octosporus* and *S. japonicus*), insects (*Apis mellifera* and *D. melanogaster*), plants (*Zea mays* and *Glycine max*), *C. elegans* worm and animals (*Danio rerio*, *Salmo salar* and *Homo sapiens*). Strictly conserved residues are in white on a black background. Partially conserved residues are boxed and in bold. This panel was generated using the ESPript server 37. Secondary structure elements detected from Erh1 apo structure are shown above the alignment. Residues involved in homodimer formation are indicated by black stars below the alignment.
- E. Sequence conservation score mapped at the surface of Erh1 homodimer. The conservation score has been calculated using the ConSurf server <sup>38</sup> from the alignment shown in Fig. 1D. Coloring is from white (low conservation) to red (strictly conserved). The Mmi1 region interacting with Erh1 is shown as a cyan cartoon. The side chain for the Mmi1 tryptophan residue (W112) present at the heart of the interface is shown as sticks in panels A and B.
- F. Expression levels of dimeric GFP-Erh1 and monomeric GFP-Erh1<sub>I11R,L13R</sub> in *erh1Δ* cells. Western blot showing total GFP-Erh1 and GFP-Erh1<sub>I11R,L13R</sub> levels expressed from the P<sub>nmt41</sub> promoter (pREP41 vector) and obtained in denaturing conditions. An anti-CDC2 antibody was used for loading control.
- G. Monomeric Erh1<sub>I11R,L13R</sub> protein still interacts with Mmi1. Western blot showing that both GFP-Erh1 and GFP-Erh1<sub>I11R,L13R</sub> co-immunoprecipitate with Mmi1-TAP. (WCE) Whole Cell Extract; (IP) Immunoprecipitation.

**Figure 2: *Erh1* dimerization is required for meiotic mRNAs recognition and degradation by *Mmi1***

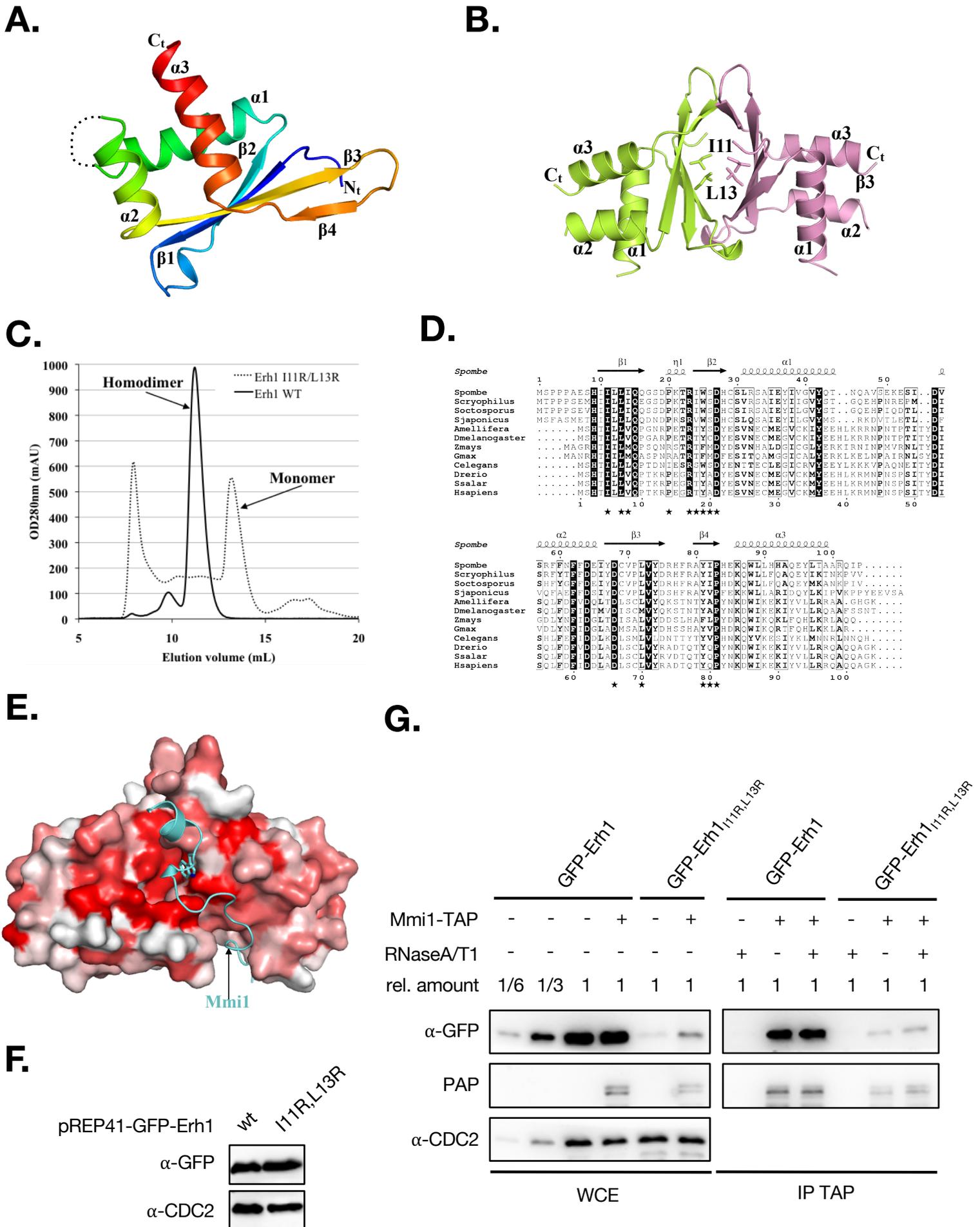
- A. The  $Erh1_{I111R,L13R}$  mutant phenocopies the deletion of *erh1+* gene. Spotting assays at 23°C, 30°C and 37°C. Cells of the indicated genotypes were grown until mid-log phase and plated on EMM-LEU medium at an initial OD = 0.25 followed by 5-fold serial dilutions.
- B. Live cell microscopy of GFP-tagged *Erh1* and  $Erh1_{I111R,L13R}$  in strains of the indicated genotypes. Cells were imaged by differential interference contrast (DIC) and with a GFP filter. Red arrows indicate  $Erh1_{I111R,L13R}$  cells for which GFP signals are similar to that of wild type *Erh1*. Squares with white dashed lines lying on the bottom right of GFP panels show enlarged images of the small squares.
- C. RT-qPCR analysis of the *mei4+*, *ssm4+* and *meiRNA* transcripts in strains of the indicated genotypes. Signals were normalized to *act1+* mRNA levels and expressed relative to wild-type cells. Error bars represent the standard deviation from four independent experiments.
- D. RNA-immunoprecipitation experiments in *erh1Δ* cells expressing GFP-*Erh1* or GFP- $Erh1_{I111R,L13R}$ . Shown are the enrichments (% input) of *act1+*, *mei4+*, *ssm4+* and *meiRNA* transcripts upon pulldown of TAP-tagged *Mmi1*. Error bars represent the standard deviation of four independent immunoprecipitations from biological duplicates.

**Figure 3: *Erh1* dimerization is required for efficient meiosis progression and *meiRNA* dot formation**

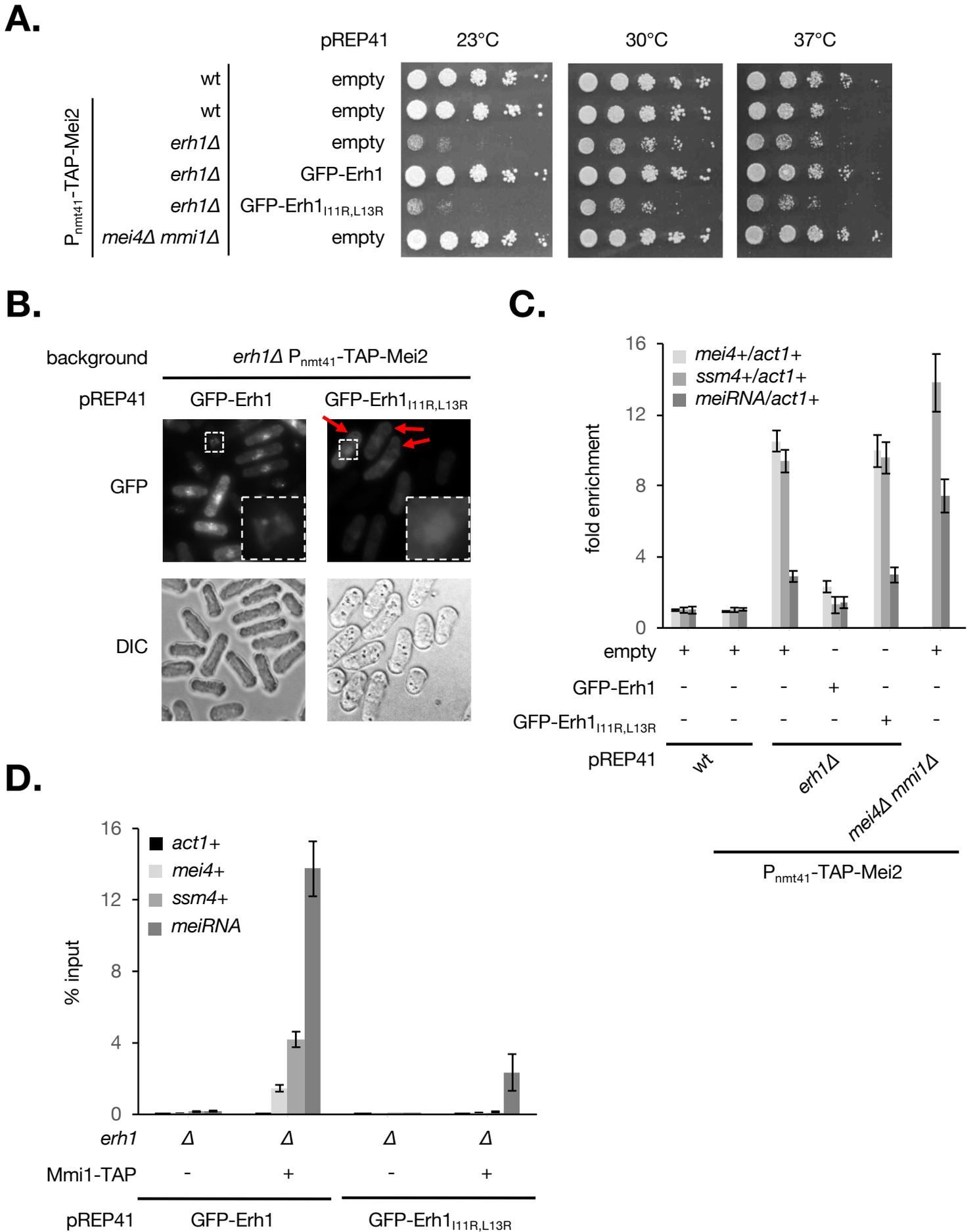
- A. Homothallic strains of the indicated genotypes were spotted on EMM-LEU plates and incubated for 5 days at 30°C. The presence or absence of asci was determined by iodine staining and live cell imaging (bright field). The mating/sporulation efficiency is indicated for each strain and represents the percentage of asci among 500 cells.
- B. Representative images of *meiRNA* (red) detected by Single molecule RNA Fluorescence In-Situ Hybridization (SmFISH) in meiotic cells. DNA was stained with DAPI (blue). GFP-tagged wild-type *Erh1* or  $Erh1_{I111R,L13R}$  were visualized in parallel. Images are shown as the maximum-intensity projections of Z-stacks.

**Table 1: Data collection and refinement statistics**

<b><u>Data collection</u></b>	
Space group	C222 <sub>1</sub>
Unit cell parameters	71.9Å; 123.9Å; 68.2Å; 90°; 90°; 90°
Wavelength (Å)	0.97857
Resolution (Å)	50-1.95 (2.07-1.95)
R <sub>merge</sub> (%)	7.7 (190.8)
<i>I</i> / $\sigma I$	14.3 (1.0)
Completeness (%)	99.6 (98)
CC <sub>1/2</sub> (%)	99.9 (50.4)
Redundancy	7.9
Observed reflections	176672
Unique reflections	22390
<b><u>Refinement</u></b>	
Resolution (Å)	50-1.95
R / R <sub>free</sub> (%)	20 / 23.2
<i>Number of atoms</i>	
Protein	2308
SO <sub>4</sub> <sup>2-</sup> / Ethylene glycol / Tris / Acetate	5 / 20 / 8 / 4
Water	73
<i>B-factors (Å<sup>2</sup>)</i>	
Protein	61
SO <sub>4</sub> <sup>2-</sup> / Ethylene glycol / PEG	90.2 / 70.1 / 90 / 97
Water	58.6
<i>R.m.s deviations</i>	
Bond lengths (Å)	0.010
Bond angles (°)	0.94

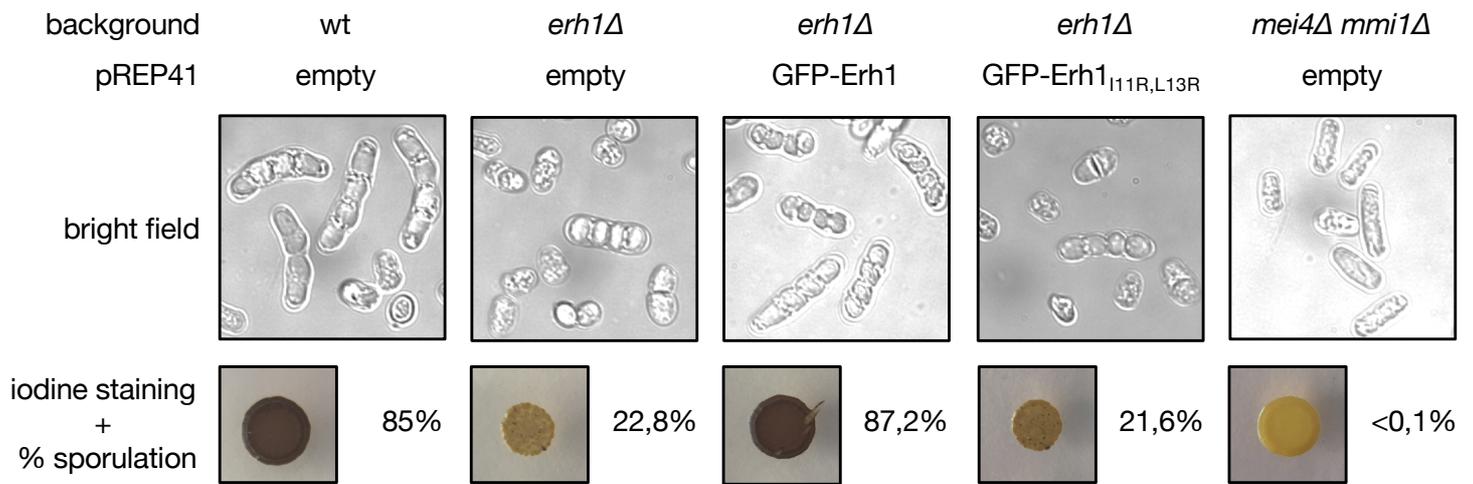


**Figure 1: *S. pombe* Erh1 structure**

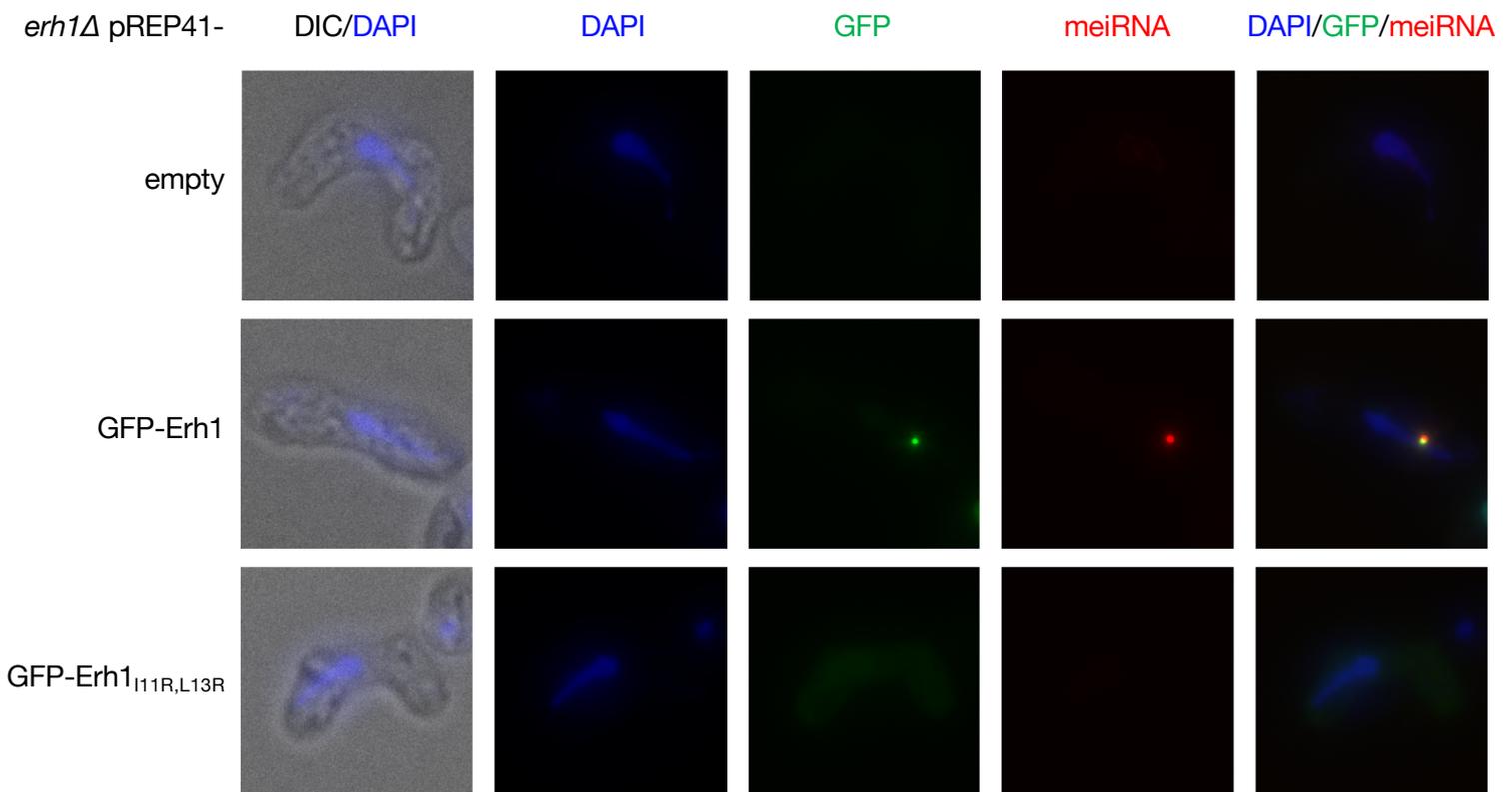


**Figure 2: Erh1 dimerization is required for meiotic mRNAs recognition and degradation by Mmi1**

**A.**



**B.**



**Figure 3: Erh1 dimerization is required for efficient meiosis and meiRNA dot formation**



## Supplementary files

### **Formation of *S. pombe* Erh1 homodimer mediates gametogenic gene silencing and meiosis progression**

**Ditipriya Hazra<sup>1,#</sup>, Vedrana Andrić<sup>2,4,#</sup>, Benoit Palancade<sup>3</sup>, Mathieu Rougemaille<sup>2</sup> and Marc Graille<sup>1</sup>**

<sup>1</sup>: BIOc, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, F-91128 Palaiseau, France.

<sup>2</sup>: Institut de Biologie Intégrative de la Cellule, CNRS, UMR9198, Université Paris-Saclay, 91190 Gif-sur-Yvette, France

<sup>3</sup>: Institut Jacques Monod, CNRS, UMR7592, Univ Paris-Diderot, Sorbonne Paris Cité, 75013 Paris, France

<sup>4</sup>: Université Paris-Saclay, 91190 Gif-sur-Yvette, France

#: equal contribution

Correspondence should be addressed to MR ([mathieu.rougemaille@i2bc.paris-saclay.fr](mailto:mathieu.rougemaille@i2bc.paris-saclay.fr)) and MG ([marc.graille@polytechnique.edu](mailto:marc.graille@polytechnique.edu)).

**Table S1: Oligonucleotides and plasmids used to over-express proteins in *E. coli***

ORF	Primer name	Primer sequence (the restriction sites are underlined)	Plasmid
Erh1	oMG511	TATAGGATCCAGCCCCCACCCGCG	pMG921
	oMG512	GCGGCTCGAGTTACGGAATCTGACGAGCCGC	
Erh1 <sub>I11R,L13R</sub>	oMG629	CGAATCTCATATCAGGCTGAGGATTCAGCAAGGT TCTGACCCT	pMG945
	oMG630	CCTTGCTGAATCCTCAGCCTGATATGAGATTCGGC GGGTGGGGGGC	
Mmi1-[95-122]	oMG605	TATA GGATCC GTAAATATGATTTTAGCAGGC	pMG915
	oMG606	TATA CTCGAG TCAAGACTCACGACGAAGG	

**Table S2: *S. pombe* strains used in this study**

Strain	Genotype
PR040	h90, <i>ura4-DS/E</i> , <i>ade6-M210</i> , <i>leu1-32</i> , <i>mat3M::ura4+</i>
PR808	PR040, <i>pREP41::LEU2</i>
PR1026	PR040, <i>mei4::nat<sup>RMX</sup> mmi1::hph<sup>RMX</sup> kan<sup>RMX</sup>::P<sub>nmt41</sub>-TAP-Mei2 pREP41::LEU2</i>
PR1316	PR040, <i>mei4::nat<sup>RMX</sup> mmi1::hph<sup>RMX</sup> pREP41::LEU2</i>
PR1413	PR040, <i>kan<sup>RMX</sup>::P<sub>nmt41</sub>-TAP-Mei2 pREP41::LEU2</i>
PR1414	PR040, <i>erh1::nat<sup>RMX</sup> kan<sup>RMX</sup>::P<sub>nmt41</sub>-TAP-Mei2 pREP41::LEU2</i>
PR1415	PR040, <i>erh1::nat<sup>RMX</sup> kan<sup>RMX</sup>::P<sub>nmt41</sub>-TAP-Mei2 pREP41-GFP-Erh1::LEU2</i>
PR1416	PR040, <i>erh1::nat<sup>RMX</sup> kan<sup>RMX</sup>::P<sub>nmt41</sub>-TAP-Mei2 pREP41-GFP-Erh1<sub>I11R,L13R</sub>::LEU2</i>
PR1420	PR040, <i>erh1::nat<sup>RMX</sup> pREP41::LEU2</i>
PR1421	PR040, <i>erh1::nat<sup>RMX</sup> pREP41-GFP-Erh1::LEU2</i>
PR1422	PR040, <i>erh1::nat<sup>RMX</sup> pREP41-GFP-Erh1<sub>I11R,L13R</sub>::LEU2</i>
PR1440	PR040, <i>erh1::nat<sup>RMX</sup> Mmi1-TAP::hph<sup>RMX</sup> pREP41-GFP-Erh1::LEU2</i>
PR1441	PR040, <i>erh1::nat<sup>RMX</sup> Mmi1-TAP::hph<sup>RMX</sup> pREP41-GFP- Erh1<sub>I11R,L13R</sub>::LEU2</i>

**Table S3: Oligonucleotides used in this study**

<b>Primers</b>	<b>Sequence</b>	<b>Related figures</b>
P249: <i>mei4</i> + fwd	5'-TGGATCAGATCCGTGGAATC-3'	2C, 2D
P250: <i>mei4</i> + rev	5'-AACGCTCGATTAGAAGGCAT-3'	2C, 2D
P253: <i>act1</i> + fwd	5'-AACCCCTCAGCTTTGGGTCTT-3'	2C, 2D
P254: <i>act1</i> + rev	5'-TTTGCATACGATCGGCAATA-3'	2C, 2D
P325: <i>ssm4</i> + fwd	5'-ACACAGTTTACGGGATTCTA-3'	2C, 2D
P326: <i>ssm4</i> + rev	5'-GATTGTGATGAAAACCTGGGT-3'	2C, 2D

**Table S4: meiRNA probes for SmFISH**

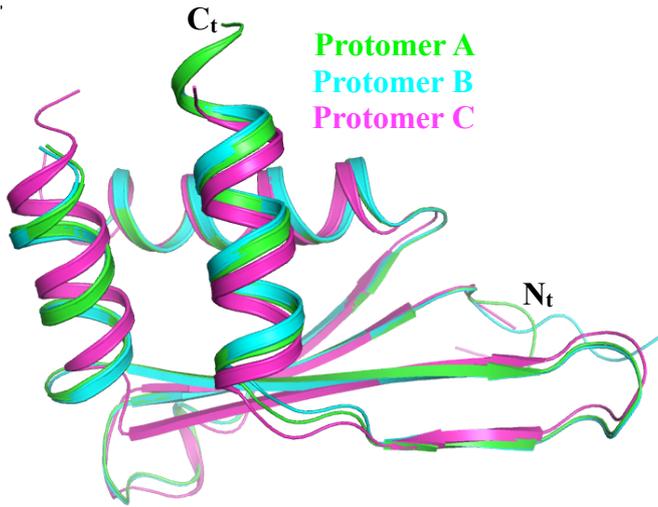
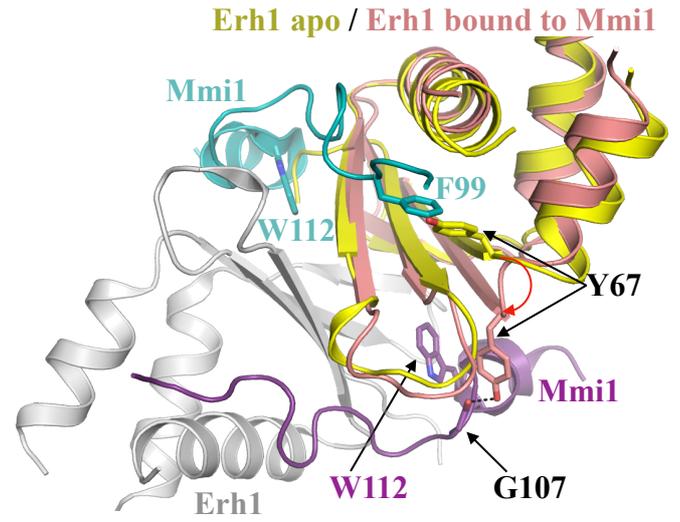
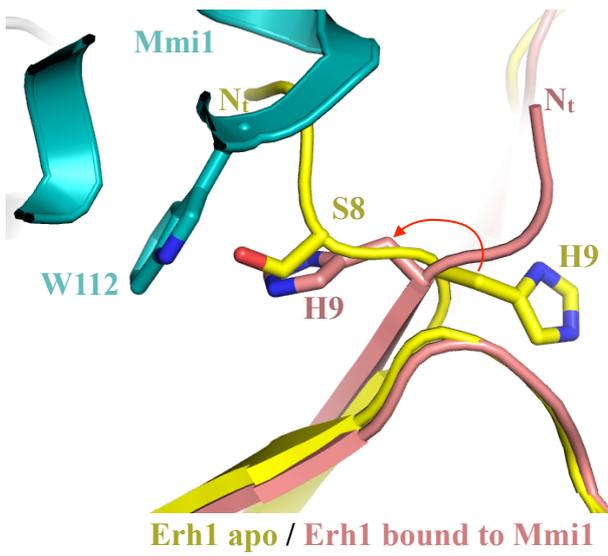
<b>Probes</b>	<b>Sequence</b>
#1	5'-ATACCCACTAAGTCTGTTTA-3'
#2	5'-CGGCAGAAGATTGACCAACA-3'
#3	5'-GCATATTCCGTCTTACAATA-3'
#4	5'-ACCAACTAAAGCGATCTTGC-3'
#5	5'-GACCATTTCAAAATGTTGCA-3'
#6	5'-TACCGAATCCAGCTTTTTGA-3'
#7	5'-CAGAGCTTAGAAGACAAGGT-3'
#8	5'-TAACTGGACCCCATCAAGAA-3'
#9	5'-TAAACCAACTTGGGGGTTGG-3'
#10	5'-TCTAAGCTACTATTCATCCA-3'
#11	5'-AGTAGATTCCATCAGTCATA-3'
#12	5'-TGCAGCCAAAAAGTGACCA-3'
#13	5'-CATTGTAAGTGCTTTCAAGG-3'
#14	5'-TTCAGTCATTCGCAAAGTTT-3'
#15	5'-AGTCGTTTTATTTCTTTTCT-3'
#16	5'-GTTTCAACAATAGTTCAGGT-3'
#17	5'-TCTGTTTCAGGAATACGTTT-3'
#18	5'-TGTTTCGCATCAAACCTTCA-3'
#19	5'-GCGTTTAAACAAACTGCGGG-3'
#20	5'-TGGTTTCAGCACGTTTTCAA-3'
#21	5'-TTGGTTTGCAGGGTTTAACG-3'
#22	5'-CTTGCTGTGGTTATTGTTTA-3'

***Expanded view Figure 1: Structural rearrangements of Erh1 upon binding to Mmi1***

- A. Superimposition of the three copies of Erh1 present in the asymmetric unit.
- B. Comparison of apo and Mmi1-bound Erh1 structures (rmsd values ranging from 0.8-1.2Å over 80 C $\alpha$  atoms). This reveals a large conformational change of the  $\alpha$ 2- $\beta$ 3 hinge characterized by a 4Å translation of Ile66 C $\alpha$  atom and a 180° rotation of Tyr67 side chain (indicated by a red arrow). As a result, this renders accessible a hydrophobic cavity at the surface of Erh1, in which Phe99 side chain from one Mmi1 accommodates in the Erh1-bound structure. Concomitantly, Erh1 Tyr67 hydroxyl group forms a hydrogen bond (depicted as a black dashed line) with the Gly107 carbonyl group from the second Mmi1 peptide.
- C. Upon Mmi1 binding, the N-terminal extremity of Erh1 strand  $\beta$ 1 rearranges and His9 side chain flips by 180° (red arrow) to stack with Mmi1 Trp112 side chain.

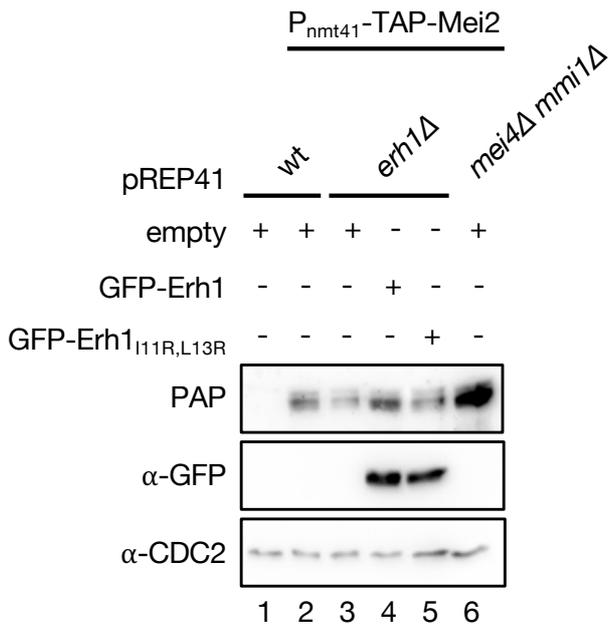
***Expanded view Figure 2: Erh1 does not contribute to Mmi1-dependent downregulation of Mei2***

Western blot showing total TAP-Mei2 levels expressed from the P<sub>nmt41</sub> promoter in strains of the indicated genotypes. anti-GFP and anti-CDC2 antibodies were used to evaluate Erh1 levels and loading, respectively.

**A.****B.****C.**

**Supplemental Figure 1: Structural rearrangements of Erh1 upon binding to Mmi1**

**A.**



**Supplemental Figure 2: Erh1 does not contribute to Mmi1-dependent downregulation of Mei2**



# CHAPTER V

Structural and functional  
characterization of Pho92 and Not1



## ***Not1-Pho92 results:***

### ***5.1 Cloning:***

Preliminary yeast two-hybrid results from our collaborator's lab (Seraphin's lab, IGBMC, Strasbourg) indicated that the Not1 fragment covering residues 1080 to 2108 is strongly interacting with Pho92, the *S. cerevisiae* YTH-containing m<sup>6</sup>A reader (Fig. 67). As the YTH domain proteins interact with effector proteins with their N-terminal, for instance human NOT1 interacts with a N-terminal region from YTHDF2, the NTD of Pho92 (1-142) was assumed to be interacting with Not1 (Fig. 68a). The DNA sequences encoding for these *S. cerevisiae* Not1 and Pho92 regions were successfully amplified and cloned into pET28b (with a His<sub>6</sub>-tag) and pGEX-6P1 vectors, respectively. All plasmids were verified by sequencing.

### ***5.2 Solubility profiling:***

*E. coli* BL21 (DE) Gold and Codon plus cells were induced with varying amount of IPTG and subjected to grow at 18 °C. The cells were harvested and after small-scale purification, elution and soluble fraction were checked separately and the samples were run on 12% or 15% SDS-PAGE gel.

The first constructs to be checked were Not1 (1348-2093), (1565-2093) as well as Pho92 (FL) or (1-142). From co-expression of Not1 (1565-2093) and Pho92 1-142, very faint bands were observed at desired molecular weight (Fig. 68b). However, the expression level of Not1 was too poor to determine if there was a strong interaction between these two constructs.

### ***5.3 Construction of a recombinant plasmid for improving the yield of Not1:***

To validate the interaction between these two domains, it was important to improve the expression of Not1 fragments. To do that, a solubility tag was designed consisting of a N-terminal His<sub>6</sub>-tag followed by two Z domains and a 3C protease cleavage site, referred to as His<sub>6</sub>-ZZ. The Z domain is an IgG binding domain of Protein A from *S. aureus* that is used to improve solubility of a protein (Inouye and Sahara, Protein. Expr. Purif., 2009). This tag allows two modes of affinity purification, using Ni-NTA or IgG. The DNA sequence encoding for the His<sub>6</sub>-ZZ tag was introduced into pET28b vector, between NcoI and BamHI restriction site (Fig. 69). This recombinant pET28b-His<sub>6</sub>-ZZ vector was used to clone all the Not1 constructs using BamHI and XhoI restriction sites. The recombinant protein, containing ZZ domain and an N-terminal His tag had significantly improved yield (Fig. 70).

## **5.4 Purification:**

### **5.4.1 Not1 protein production and purification:**

To get a confirmatory result of Not1(1565-2093) and Pho92(1-142) interaction, Not1 (1565-2093) was cloned into recombinant pET28b vector containing His-ZZ tag. Recombinant protein was expressed in BL21 (DE) Gold cells in TBAI media at 18°C containing Kanamycin (50µg/ml). The His-ZZ tagged protein was purified by two step purification: Ni-NTA affinity and Size exclusion chromatography.

### **5.4.2 Pho92 purification:**

GST-tagged Pho92 constructs were expressed in BL21 (DE) Gold cells in TBAI media at 18°C with Ampicillin (100µg/ml). All Pho92 constructs were purified by three purification steps: GST affinity purification, ion-exchange and gel filtration. The purifications were performed following the standard protocol described in section 3.4. The protein was further purified using anion-exchange chromatography (HiTrap Q SP FF, GE healthcare). The GST tagged protein was finally purified by Size Exclusion Chromatography by Superdex 75 column (GE healthcare). The concentration of the protein was estimated by the method of Bradford (1976) and measuring absorbance at 280 nm. The protein purity and molecular weight was verified using 15% SDS-PAGE stained with Coomassie brilliant blue R-250.

### **5.4.3 Identification of interacting domains:**

The full-length Not1 is a large protein (>200kDa) and cannot be produced in *E. coli*. As we had evidences from our collaborators that the Not1 fragment encompassing residues 1080 to 2108 strongly interacts with Pho92, a series of pull down experiments were performed to narrow down the Not1 region required for interaction with Pho92 NTD (1-142). To determine whether Not1(1565-2093) interacts with Pho92, a GST pull down was performed by mixing His<sub>6</sub>-ZZ-Not1 (1565-2093) with GST-Pho92 (1-142) and GST alone. From this pull down, no strong interaction was detected between these two constructs (Fig. 71).

### **5.4.4 Boundary determination for Not1:**

To narrow down the boundary of Not1, three other constructs were designed encompassing residues 1071-1565, 1071-1282 and 1343-1565. The determination of domain boundary was based on secondary structure prediction by HCA (hydrophobic cluster analysis). All these constructs were cloned into recombinant pET28b vector, carrying His<sub>6</sub>-ZZ tag. Best condition for expression was searched by transforming the corresponding plasmids in two different

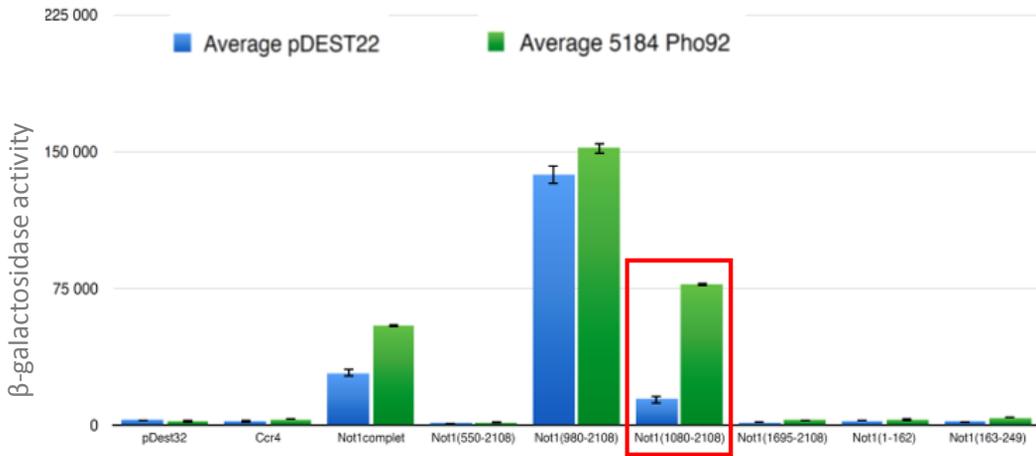


Fig. 66: Yeast two hybrid result from collaborators (Seraphin Lab, IGBMC, Strasbourg) helps to narrow down the interacting region from Not1. The domain boundary of Not1 that shows highest affinity for Pho92 is highlighted in the red box

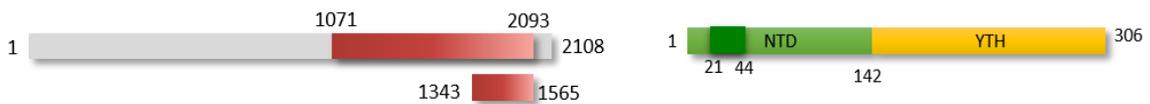


Fig. 67: Domain organization of ScNot1 (left) and ScPho92 (right). The domain boundaries required for interaction from both Not1 and Pho92 are highlighted in the box

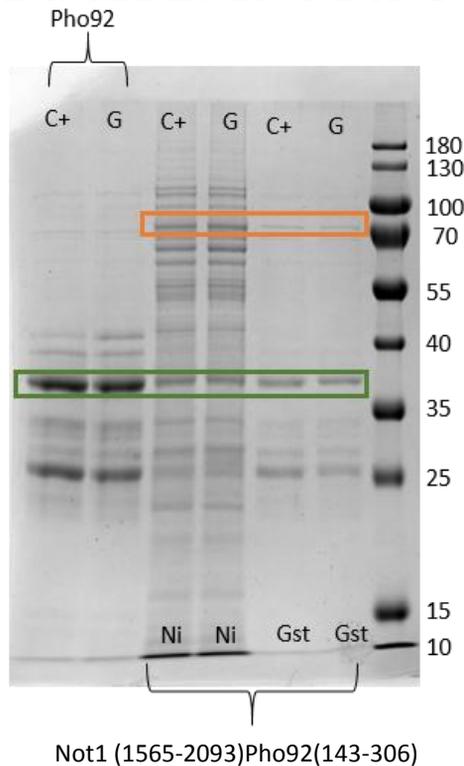


Fig. 68: Expression assay of Not1(1565-2093) along with Pho92(1-142) at 18°C in TBAI media with Codon+ and Gold cells. The elution fraction from the first affinity column was analysed by SDS-PAGE. Presence of very faint bands at desired molecular weight can be detected. Orange and green box highlight bands corresponding to His<sub>6</sub>-Not1(1565-2093) and GST-Pho92(1-142) respectively

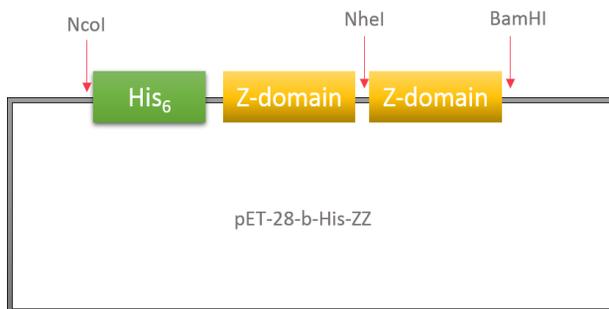


Fig. 69: Simple representation of the His-ZZ-pET-28-B plasmid used to express some proteins

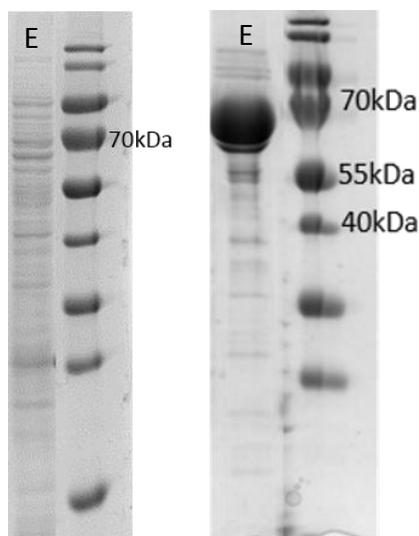


Fig. 70: SDS-PAGE analysis from Ni-NTA elution for Not1(1565-2093) expressed(left) without His<sub>6</sub>-ZZ tag, (right) with His<sub>6</sub>-ZZ tag

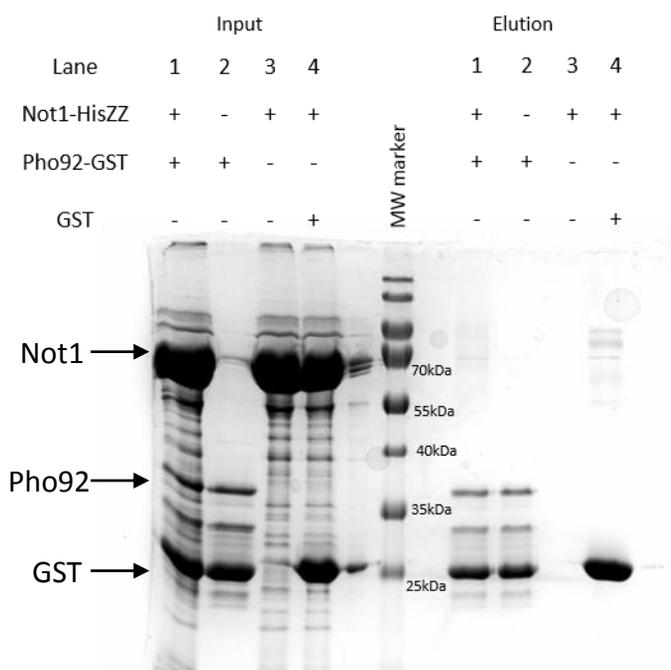


Fig. 71: GST-Pull down to study interaction between HisZZ-Not1 (1565-2093) and either GST-Pho92 (1-142). GST was used as control. Proteins present in input and elution from GST purification has been analysed by SDS-PAGE

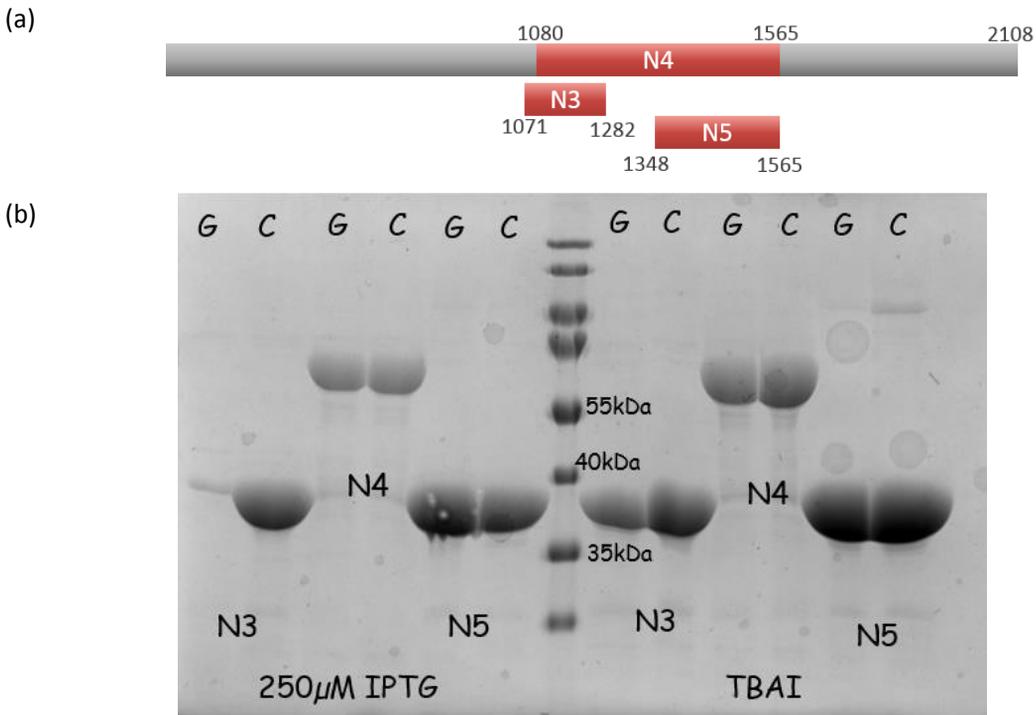


Fig.72: (a) Schematic representation of the Not1 boundaries studied (b)SDS-PAGE analysis of the proteins eluted from Ni NTA expression assay. Abbreviations indicate the following: N3: 1071-1282, N4: 1071-1565, N5: 1343-1565, G: Gold cells, C: Codon plus cells

Lane	1	2	3	4	1	2	3	4
N4	+	+	-	+	+	+	-	+
P3-GST	+	-	+	-	+	-	+	-
GST	-	-	-	+	-	-	-	+

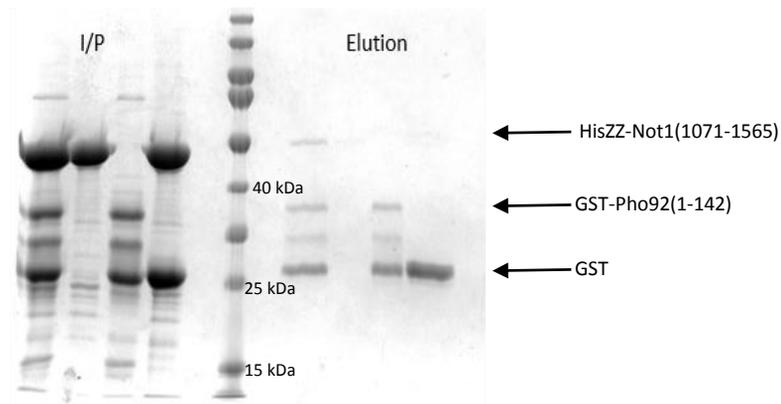


Fig. 73: HisZZ-Not1 (1071-1565) interacts with GST-Pho92 (1-142) and incubated according to GST pull-down. Binding was assessed by SDS-PAGE. N4: 1071-1565, P3: 1-142 . Same as described for Fig. 71

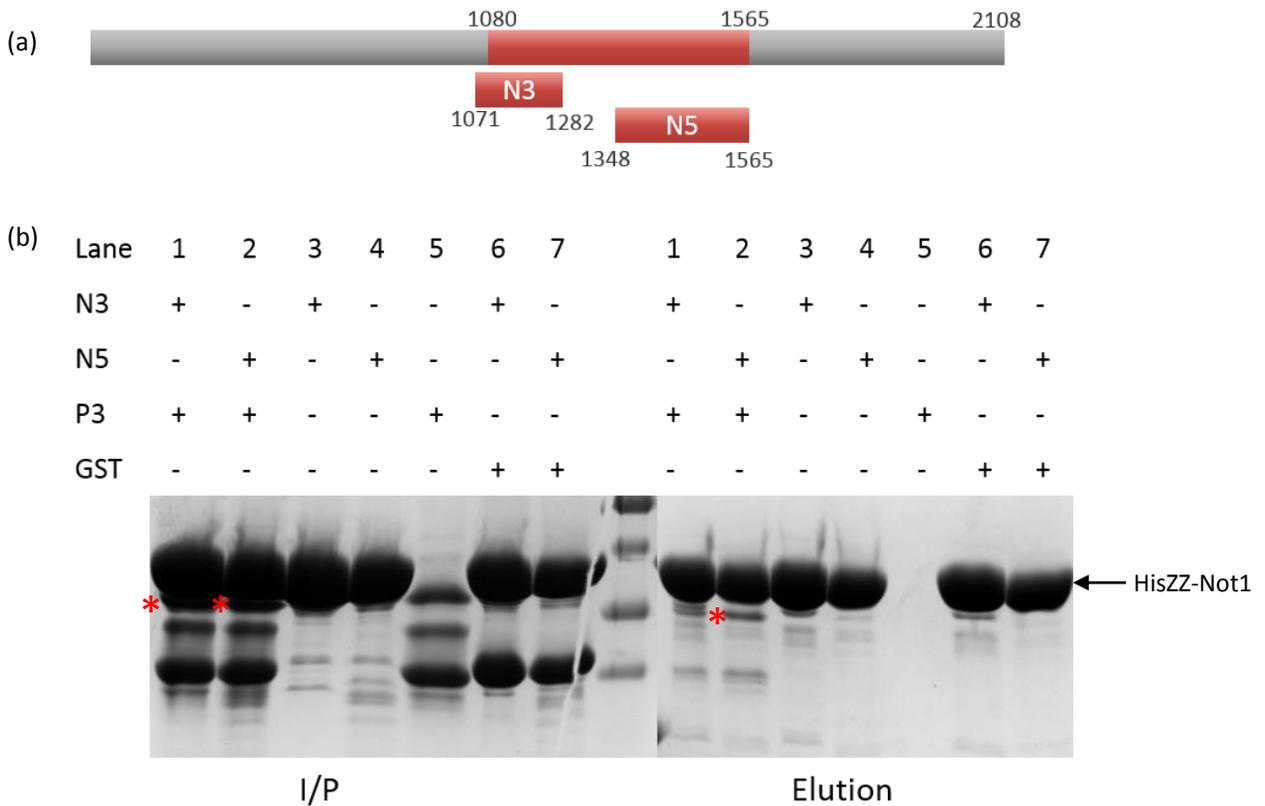


Fig. 74: (a) Schematic representation of the Not1 boundaries studied. Two different constructs of HisZZ-Not1 (N3: Not1 1071-1282, N5= 1343-1565) were mixed with GST-Pho92 (1-142) or GST alone. Their interaction was studied by Ni-NTA pull-down. Results were analysed by SDS PAGE. The arrow-head indicates bands for Not1 constructs, the red asterisk indicates band corresponding to the molecular weight of Pho92 (1-142) in elution.

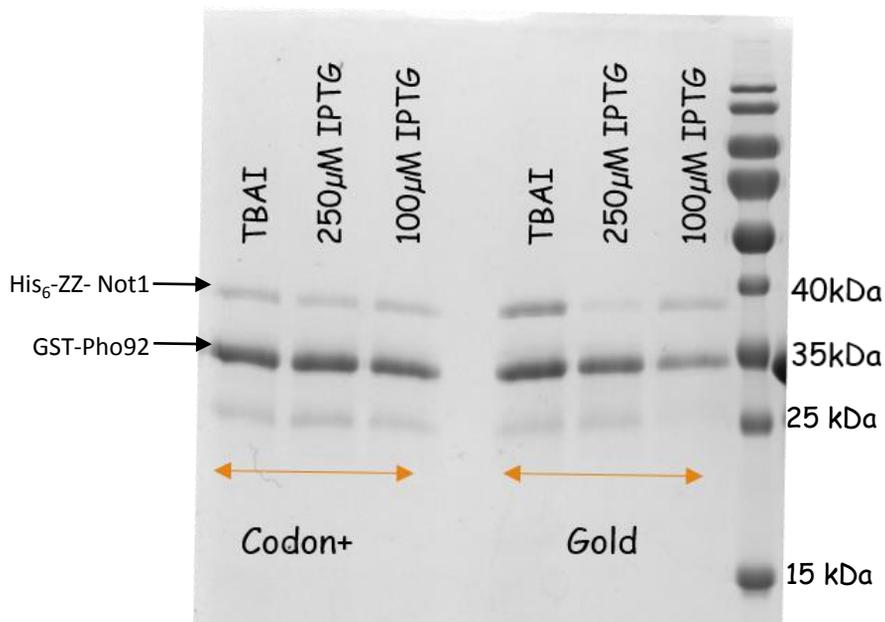


Fig.75: Co-expression assay for Not1 (1343-1565) and Pho92 (1-69) using codon+ or gold cells and TBAI media or different concentrations of IPTG.

cells, BL21 Gold and BL21 Codon Plus and they were subjected to grow in 2YT media with 250  $\mu$ M IPTG induction and in TBAI media at 18 °C over-night. The elution from Ni-NTA was run on a gel (Fig. 72) and the best condition was found to be BL21 Gold cells in TBAI media.

Recombinant protein was expressed in BL21 (DE) Gold cells in TBAI media at 18 °C containing Kanamycin (50  $\mu$ g/ml) and purified with two step purification process following the protocol described in previous section for purifying His<sub>6</sub>-ZZ-Not1 (1565-2093).

To determine the interacting region of Not1, a GST pull-down was performed by mixing His<sub>6</sub>-ZZ-Not1 (1071-1565) with GST-Pho92 (1-142) and GST alone. In the elution, a band corresponding to the molecular weight of His<sub>6</sub>-ZZ-Not1 (1071-1565) was observed with GST-Pho92 (Fig. 73). To narrow down the boundaries further, two HisZZ-Not1 constructs (1071-1282, 1343-1565) were mixed with GST-Pho92 and subjected to Ni-NTA pulldown. Both the input samples and elution samples were run on gel. Then, a direct interaction was detected between Not1 (1343-1565) and Pho92 (1-142) fragments (Fig. 74).

For further studies, Not1(1343-1565) was purified by three levels of purification steps using Ni-NTA affinity chromatography, ion-exchange (HiTrap Q SP FF) and Gel filtration chromatography. The purifications were performed following the protocol described in section 3.4. After elution from Ni-NTA (20 mM Tris-HCl, 200 mM NaCl, 400 mM Imidazole, 5mM 2-mercaptoethanol), the His<sub>6</sub>-ZZ tag was removed by overnight digestion with 3C protease under dialysis conditions (buffer: 20 mM NaCl, 100 mM KCl, 5 mM 2-mercaptoethanol). The tag was removed by ion-exchange, exploiting the difference of affinity of the tag and the protein for anion exchange resin at pH 8.0. At this pH, the His<sub>6</sub>-ZZ tag binds to anion-exchange resin (HiTrap Q) whereas Not1 does not, hence the protein of interest is present in the flow through. The protein was finally purified by size exclusion chromatography with an Superdex 75 16/60 column (GE healthcare).

#### ***5.4.5 A conserved region of Pho92:***

NTD (1-142) of Pho92 is highly unstructured, which is an obstacle for crystallization, a shorter and more compact fragment (Pho92 residues 1 to 69), containing a short region conserved among Pho92 fungal orthologues, was designed in parallel, to test interaction with Not1 (1343-1565). The best condition for Not1(1343-1565) and Pho92 (1-69) co-expression was identified by transforming the plasmids in two different cells (BL21 Gold and BL21 Codon Plus) and different media (2YT supplemented with 100 $\mu$ M or 250 $\mu$ M IPTG induction

or TBAI) at 18°C overnight. The elution fraction from Glutathione-sepharose resin (Pho92 is GST tagged) was run on a gel (Fig. 75) and the best condition was found to be BL21 Gold cells in TBAI media. As mentioned earlier, the NTD of Pho92 is a low complexity region. As low complexity regions do not have a compact folding, it can be difficult to crystallize them. Therefore, to favour crystallization, it was important to find a structured region of Pho92. To do that multiple sequence alignment was performed with Pho92 proteins from diverse fungi. This revealed the presence of a conserved region (21-44 in Pho92; Fig. 76). This peptide was cloned into pGEX-6P-1 and expressed as a GST fusion.

#### ***5.4.6 Reconstitution of the Not1-Pho92 complex:***

##### ***5.4.6.1 Co-purification of Not1 & Pho92:***

To reconstitute the Not1(1343-1565)/Pho92 (21-44) complex, each protein was expressed separately. Untagged Not1(1343-1565) was produced in BL21-Gold cells, TBAI media at 18°C in presence of Kanamycin (50 µg/ml) and GST-Pho92(21-44) in Gold cells, TBAI media, 18°C in presence of Ampicillin (100 µg/ml) and before sonication, the resuspended pellets were mixed together. The complex was co-purified by two step purification, GST, and size exclusion chromatography (Superdex 75). The supernatant was loaded in a Glutathione-sepharose column allowing the protein complex to bind to the column. After draining the flow-through fraction, the column was washed with 10 CV of lysis buffer, removing all non-specific impurities. 3C protease with 1 ml of lysis buffer was added to the protein bound matrix and incubated overnight at 4 °C to cleave the GST tag. The protein complex was obtained in the flow through and finally purified with gel filtration (20 mM TrisHCl pH8.0, 100 mM KCl, 5 mM 2-mercaptoethanol). The purity of the protein was checked by 15%SDS-PAGE (Fig. 77). Lane 1 corresponds to the elution from Glutathione-sepharose purification and Lane 2 to the peak fraction from gel filtration. Comparison of the two samples indicate that the intensity of the Pho92 band is much less in the sample from gel filtration, suggesting that Pho92 degrades or dissociates from Not1 with time.

##### ***5.4.6.2 Anisotropy:***

To assess the strength of the interaction between Pho92 (21-44) and Not1 (1349-1565), fluorescence anisotropy measurements were performed by using a FITC-tagged synthesized Pho92 peptide (21-44). The anisotropy value of Pho92 (21-44)-FITC significantly increased as higher concentrations of Not1 were added. The experiment was performed in triplicate and from the titration curve, the dissociation constant,  $K_d$ , was determined to be 4.46 µM, indicative of a significant binding (Fig. 78).

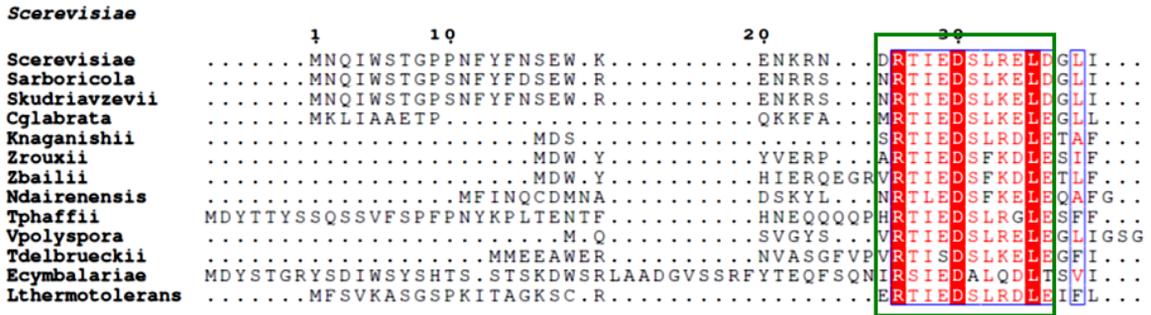


Fig. 76: Multiple sequence alignment reveals a conserved motif present in the N-terminal of Pho92 fungal orthologues. The conserved region is highlighted in the green box.

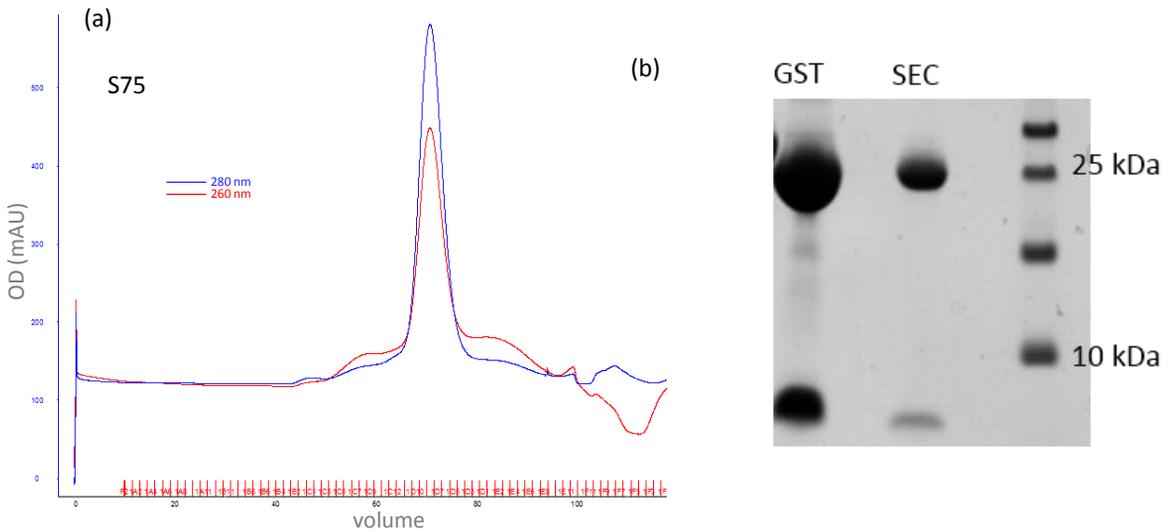


Fig. 77: (a) Elution profile of the complex from gel filtration in S 75 gel filtration column (b) 15% SDS-PAGE to check homogeneity of the Not1-Pho92 protein complex after purification from GST and gel filtration.

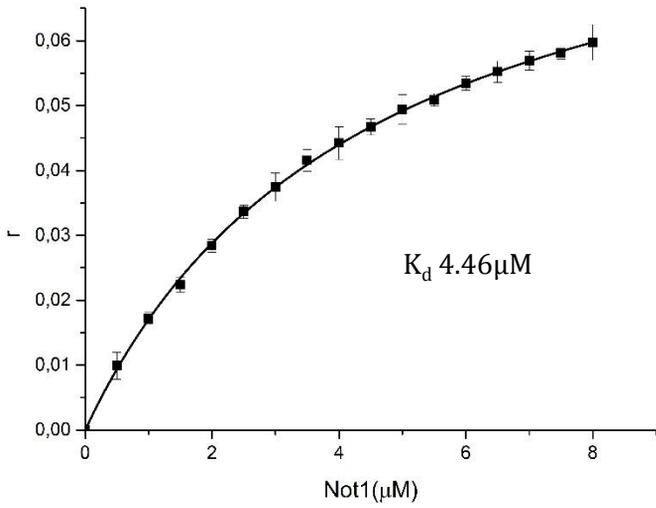


Fig. 78: Binding curve characterizing formation of Not1 (1343-1565)/ PM194 complex as measured by steady state fluorescence anisotropy. Error bars were determined from three different experiments

However, this affinity is considered to be too weak to maintain the complex under gel filtration conditions, rationalizing the dissociation observed between Pho92 peptide and Not1 during S75 purification step.

#### **5.4.6.3 Crystallization trial:**

The co-purified Not1-Pho92 complex was concentrated to 17 mg/ml and preliminary crystallization trials were performed by sitting drop method with commercially available crystallisation kits, Crystal Screen1-2 (Hampton Research), JCSG plus (Molecular Dimensions) and Index (Hampton Research). Unfortunately, no crystals appeared so far. Careful observation of the results obtained from purification indicate that the low-complexity Pho92 peptide degrades with time, suggesting that the low stability of the complex could be the reason for crystallization failure. To eliminate this problem many approaches were explored.

#### **5.4.7 Reconstitution of the complex from different organisms- *C. glabrata* and *Z. rouxii*:**

Multiple sequence alignments were performed to find corresponding regions from Pho92 and Not1 proteins from other organisms that might be more compact than in *S. cerevisiae*. To perform these multiple sequence alignment, organisms were selected on the basis of presence of the conserved Pho92 region (21-44) (Fig. 79). YTH domain proteins from divergent organisms containing this conserved region were then subjected to secondary structure prediction using HCA (Callebaut et al., 1997). Based on the HCA diagrams, two organisms were selected for further studies, *Candida glabrata* and *Zygosaccharomycesrouxii*. In both organisms, the NTD of YTH domain protein is enriched in hydrophobic residues compared to the corresponding region of *S. cerevisiae* Pho92 (Fig. 80). The following boundaries were selected from the HCA diagram prediction: 1-53 for *C. Glabrata* and 1-20 for *Z. rouxii*. The Not1 boundary was selected from both of these organisms, based on the sequence alignment with *S. cerevisiae* Not1 (1343-1565). For protein production, constructs were designed to include coding sequence for GST tag followed by a 3C cleavage site then the fragment from Pho92 protein followed by a ribosome binding site, Shine Dalgarno sequence and then coding sequence for Not1 (Fig. 81). The artificially synthesized constructs were cloned following the standard protocol.

The protein complex was produced in TBAI media, at 18 °C from in Gold cells. The GST tagged complex was purified with two step purification: GST and gel filtration

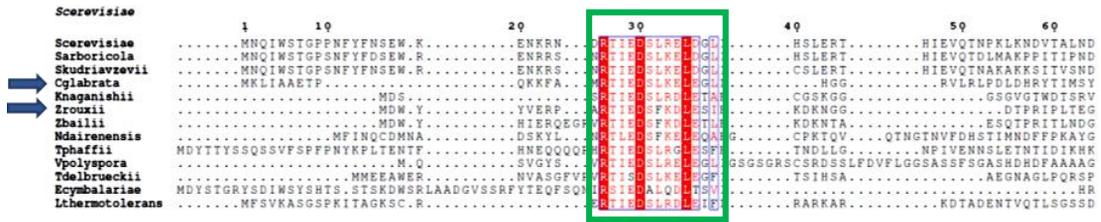


Fig. 79: Multiple sequence alignment to find proteins from other organisms carrying the conserved Pho92 region, highlighted in green box

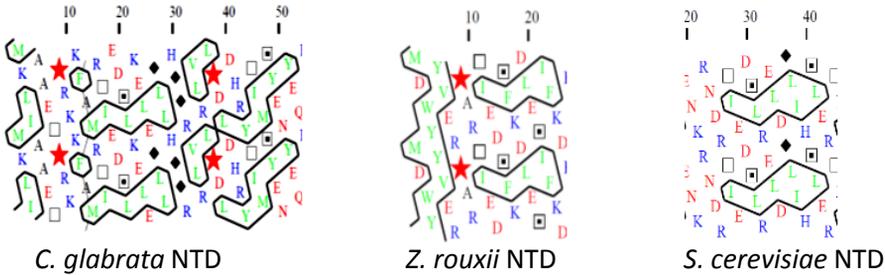


Fig. 80: HCA diagram for secondary structure prediction of domains selected from different organisms



Fig. 81: Schematic representation of construction of synthetic gene coding for desired protein domains

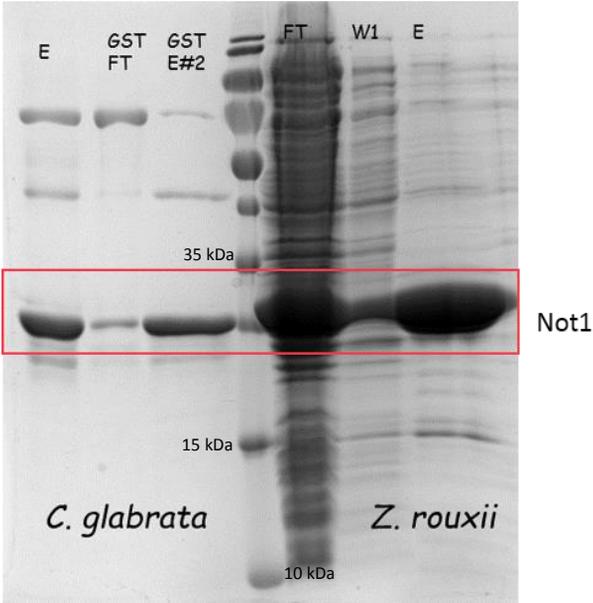


Fig. 82: SDS-PAGE gel with purification samples from *C. glabrata* and *Z. rouxii*. The red box highlights band corresponding to molecular weight of Not1. Abbreviations: E: elution, FT: flowthrough from GST column, W: wash fraction

chromatography. The protocol described for purification of *S. cerevisiae* Not1-Pho92 complex was followed.

Unfortunately, at the end of these purifications, the yields were considerably lower than those obtained for *S. cerevisiae* Not1-Pho92. Band corresponding to molecular weight of the Pho92 homologue protein was not visible on SDS-PAGE gel by Coomassie staining (Fig 82). Therefore, further experiments were not performed with this protein complex.

#### **5.4.8 Limited proteolysis:**

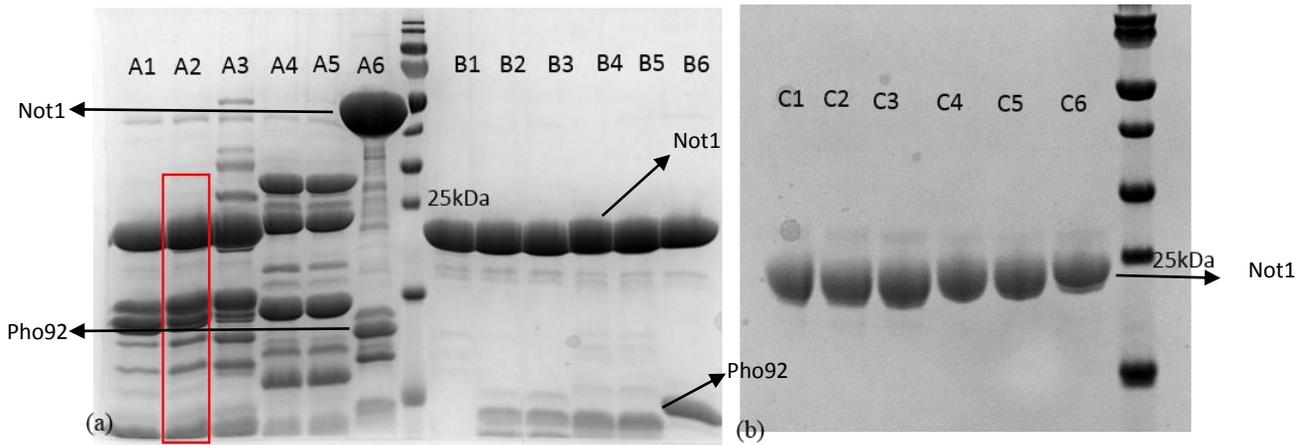
Limited proteolysis is a widely used technique to trim the flexible parts of the protein and this approach was tried next to chop off the flexible regions. Limited proteolysis by various proteases was performed with all Not1-Pho92 complexes of constructs with different lengths:

- A: Not1(1071-1565)/Pho92 (1-142)
- B: Not1(1343-1565)/Pho92 (1-69)
- C: Not1(1343-1565)/Pho92 (21-44)

The results were analysed by running the samples on 15% SDS-PAGE (Fig. 83). Undigested protein is used as the control. The effect of limited proteolysis is most prominent on the largest fragments of Not1(1071-1565) and Pho92(1-142). Appearance of bands of lower molecular weight than the undigested proteins indicate that the flexible parts have been chopped off by the proteases. This was observed for all three sets of Not1/Pho92. Trypsin: Not1(1071-1565)/Pho92 (1-142) with a ratio of 1:200 (w/w) was selected to be the best concentration of the protease. This complex [Not1(1071-1565)/Pho92 (1-142)] was digested with Trypsin (with Trypsin: Protein 1:200 ratio) for 20 minutes in 20 °C and loaded to analytical 10/300 S75 gel filtration column (GE Healthcare) (Fig. 84). The peak fraction containing the complex was concentrated and preliminary crystallization trials were performed by sitting drop vapour diffusion method with commercially available crystallization kits JCSG+ (Molecular Dimensions), Index (Hampton Research). But no crystals appeared till now.

#### **5.4.9 Co-crystallization trial with Not1-PM193 peptide:**

Co-purification of the Not1(1343-1565)/Pho92(21-44) complex results in dissociation of Pho92 with time. To circumvent this, excess amount of synthetic Pho92 peptide (PM193) was mixed with Not1 for crystallization trials. Crystallization trials were performed by



- Lanes:
- |   |                               |                                |
|---|-------------------------------|--------------------------------|
| A | Not1(1071-1565)/Pho92 (1-142) | 1: Trypsin: Protein 1:100      |
| B | Not1(1343-1565)/Pho92 (1-69)  | 2: Trypsin: Protein 1:200      |
| C | Not1(1343-1565)/Pho92 (21-44) | 3: Trypsin: Protein 1:1000     |
|   |                               | 4: Chymotrypsin: Protein 1:100 |
|   |                               | 5: Chymotrypsin: Protein 1:200 |
|   |                               | 6: Control                     |

Fig. 83: Limited proteolysis of various Not1-Pho92 complexes. 15% SDS-PAGE gels for analysis of limited proteolysis (a) limited proteolysis of Not1(1071-1565)/Pho92 (1-142) and Not1(1343-1565)/Pho92 (1-69) (b) Not1(1343-1565)/Pho92 (21-44)

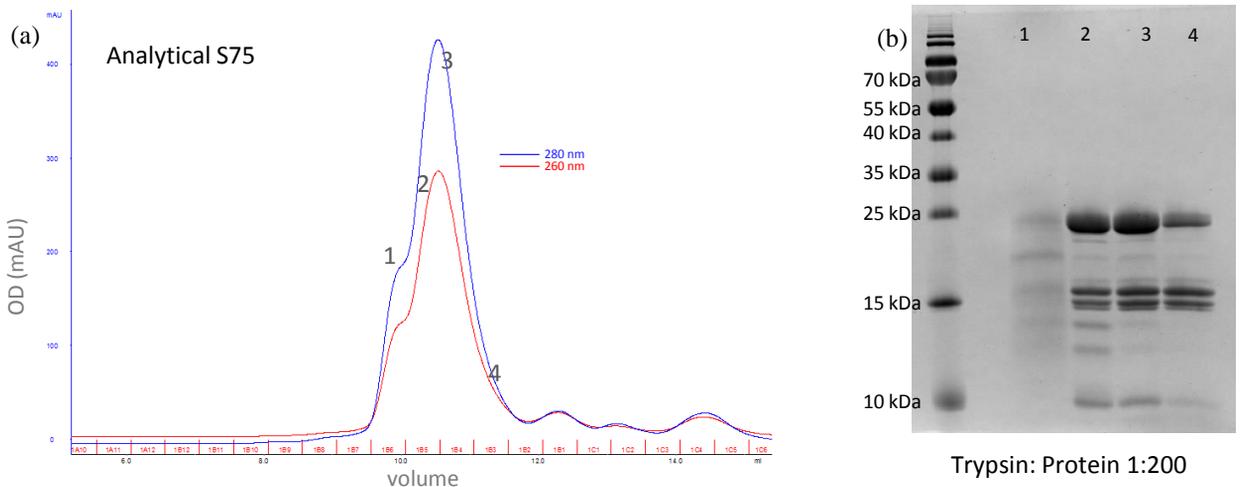


Fig. 84:(a) Chromatogram of the Trypsin digested Not1-Pho92 complex from analytical S75 (b) 15% SDS-PAGE to analyse fractions from gel filtration

mixing 2mM of Not1 (1343-1565) with 1.2 and 1.5 molar excess of PM193 [Pho92 (21-44)] peptide. Primary crystallization trials were performed using JCSG+ (Molecular Dimensions) and Protein Complex Suite (Qiagen) at either 24 °C or 4 °C by sitting drop method. Crystals were obtained in two conditions: (i) 0.1 M CHES pH9.5, 20% PEG 8,000 at 4°C and (ii) vapour diffusion at 24 °C (Fig. 85). These crystals were very small and impossible to fish. For further optimization of the crystals, sitting drops were set with a gradient of 10-30% PEG8000 vs 0.1M CHES buffer of pH 9.3, 9.5, 9.7. 24 hours after setting the plate, clear drops were seeded with the sitting drop crystals. After a day, crystals appeared in one drop 10% PEG8000, 0.1 M CHES pH9.5. The crystals were under a hard skin and it was still difficult to fish them. Further optimization is ongoing.

#### ***5.4.10 Not1-Pho92 fusion:***

As multiple crystallization trials failed using co-purified Not1 and Pho92 protein complex, probably due to the micromolar affinity between these two proteins, two constructs were designed so as to fuse Not1 (1343-1565) and Pho92 (21-44) by a flexible linker of Gly-Ser-Ser repeated 5 times, (GSS)<sub>5</sub>. In one construct, the Pho92 peptide is on the C-terminal end of Not1 (gbMG21, where gb stands for G-block) and in the other one on the N-terminal of Not1 (gbMG22) (Fig. 86). The DNA sequences encoding for these constructs were commercially synthesized and cloned in pGEX-6P-1 vector. Cloning and expression checking was done following the standard protocol described above. Both constructs were expressed in Codon plus, TBAI media, at 4°C in presence of Ampicillin (100 µg/ml) and Chloramphenicol (25µg/ml). These proteins were purified by two step purification process, GST and SEC following the protocol described for purification of the Not1(1343-1565)/Pho92(1-69) complex. Before solving the structure, it is impossible to be certain about which construct Pho92 peptide interacts with Not1. To get a preliminary idea, fluorescence anisotropy was performed with the FITC-tagged Pho92 peptide (PM194). The rationale behind this, was that if in one of the constructs, the fused Pho92 peptide binds to Not1, it will not interact with the FITC-tagged peptide. From this fluorescence polarization study, it was observed that gb21 interacts with the FITC-tagged Pho92 peptide (PM194) in a manner comparable to Not1 (1343-1565) alone, however, gb22 does not interact at all. This led to the conclusion that in gb22 construct, the Pho92 and Not1 fragments are interacting intra- molecularly. The protein (gb22) was concentrated up to 20 mg/ml and preliminary crystallization trial was performed with JCSG (molecular Dimensions). No crystals appeared yet and further trials are ongoing.

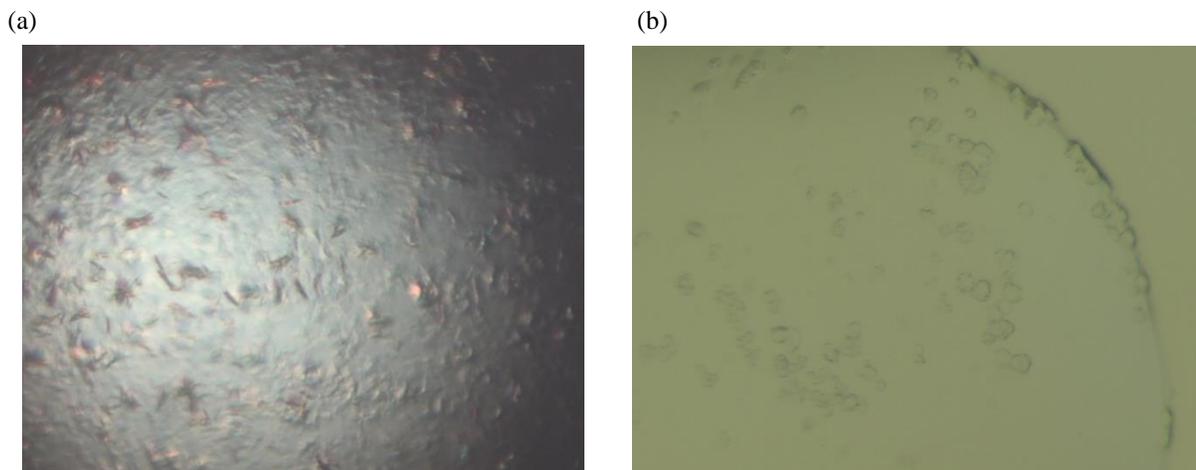


Fig. 85: Crystals of Not1-PM193 (Pho92(21-44) peptide) grown from by vapour diffusion at 24°C (a) or at 4°C in 0.1M CHES pH9.5, 20% PEG 8,000 (b)



Fig. 86: Schematic representation of the constructs designed to express Pho92-Not1 fusion. Two constructs were designed (left) Not1 (1343-1565)- (GSS)<sub>5</sub> linker-Pho92(21-44) and (right) Pho92(21-44) )- (GSS)<sub>5</sub> linker-Not1 (1343-1565). They were cloned in pGex-6P-1 vector



# DISCUSSION



## CHAPTER VI

Discussion, conclusion and  
Future prospect of Mmi1,  
Mei2, Erh1



## **6.1 Mmi1-Erh1 and Mei2:**

Mmi1 inhibits meiotic entry by selectively degrading meiotic transcripts. Erh1 is a small protein with unknown function that forms a complex with Mmi1. Mei2 is a RNA binding protein essential for pre-meiotic DNA synthesis and entry to meiosis I. The following discussion is based on our experimental findings on these three proteins.

### **6.1.1 Mei2-RRM3 is a novel RNA binding domain:**

The C-terminal RRM domain of Mei2, RRM3 is a novel RRM domain. Comparison of the apo-RRM3 structure with existing RRM domains indicated presence of N-terminal and C-terminal extensions. The RNA substrate of RRM3 was determined by ITC experiments with different RNA constructs and rigorous analysis of structure. Examination of meiRNA sequence revealed repeated presence of a UUUUUGUU motif, suggesting this might be the binding site of Mei2. Finally, RRM3 was crystallised with a 12 base RNA: GCUUUUUGUUCG. Analysis of the RNA bound structure, revealed that at least three residues, U<sub>6</sub>, U<sub>7</sub> and G<sub>8</sub> are specifically recognized by the protein. Like most classical RRM domains, the RNA protein interaction between RRM3 and RNA is also mediated by the conserved aromatic residues of the two RNP motifs: Y642 and F644 of RNP1 and Y629 of RNP2. RRM3 has the identical aromatic residues in RNP1 as *D. melanogaster* Sxl-UNR RRM domain. The Sxl-UNR RRM domain aromatic residues Y168/F170 specifically recognize a guanidine residue whereas the RRM3 Y642/F644 specifically recognize a uridine residue, U<sub>7</sub>. G to U mutation at this position decreases the affinity of RRM3, as validated by ITC. The G<sub>8</sub> is inserted in a pocket created by R631, I632 and F634, residues located on the loop connecting  $\beta_3$  and  $\beta_4$ . The R631 forms multiple H-bonds with G<sub>8</sub> and F634 is involved in a stacking interaction. This interaction is comparable to quasiRRM-RNA interaction, where residues R81 and Y82 interact with G<sub>3</sub> as R81 specifically recognizes O6 of G<sub>3</sub> and Y82 is involved in stacking (Fig. 87). RRM3 possesses a fifth  $\beta$  strand. So far, only two structures have been reported to have a fifth  $\beta$ -strand, RBD2 and RBD3 of human Polypyrimidine tract binding protein (PTB) (Oberstrass et al., 2005). The RNA bound structure of RBD2 has been solved, and from the structure it is evident that the fifth  $\beta$ -strand helps in accommodating a longer stretch of RNA. However, in the RRM3-RNA crystal structure the RNA is not interacting with the  $\beta_5$  which is guarded by  $\alpha_C$ . RRM3 has been crystallized with a 12 base RNA but in ITC, it showed higher affinity for a 15 base RNA. As the 15 base RNA and RRM3 complex did not crystallize, it is difficult to predict the role of  $\beta_5$  in RNA-RRM3 interaction. It is possible that RRM3 might undergoes a conformational change

when it binds RNA, longer than 12 bases and allows  $\beta_5$  to interact with the RNA. Further structural study is required to reveal this mechanism. These interactions between RRM3 and the RNA are reminiscent of quasi-RRM domain and classical-RRM domain, indicating that RRM3 is able to bind an RNA using the classical RRM RNA binding mode but it also has other non-canonical interactions. Most of the RRM domain containing proteins have multiple RRM repeats, allowing binding of a longer stretch of an RNA. Mei2 is also comprised of three RRM domains, among them RRM3 is essential for biological function of the protein (Watanabe et al., 1997). The known natural substrate of Mei2 is a lncRNA, meiRNA of approximately 0.5kb in size. Therefore, it can be proposed that Mei2 is able to bind a long stretch of RNA with the help of all three RRM domains, and probably the non-canonical RNA binding mode of RRM3 helps in that. But further experimental evidences are required to decide this hypothesis.

#### ***6.1.2 Mmi1-Mei2 interaction:***

Mmi1 is known to be sequestered in meiotic cells by Mei2 and meiRNA while in mitotic cells, Mmi1 degrades Mei2 by recruiting Ccr4-Not for ubiquitination of Mei2. In a study by Yamamoto group they performed a pull down with Mmi1 and Mei2 full length, N-terminal (1-429), C-terminal (429-750) (Harigaya et al., 2006). From their result it was evident that Mmi1 interacts with both N-terminal and C-terminal domains of Mei2, and it has a higher affinity for C-terminal RRM domain. To validate complex formation between Mmi1-YTH and Mei2, multiple experiments were performed. But no direct interaction was observed between these two proteins. As these two proteins have a common RNA substrate, meiRNA, it is possible that their interaction is bridged by an RNA molecule. To investigate that, an RNA construct was designed, harbouring the binding site of both proteins. A pull-down was performed by mixing the two proteins with this RNA but no interaction was detected. The binding site of Mmi1 is A-rich (UUAAAC) and Mei2 is U-rich (UUUUUGUU), therefore the chance of forming a double stranded structure cannot be omitted and this might be the reason why the complex did not form. Hence, the mode of interaction between Mmi1 and Mei2 still remains to be deciphered.

#### ***6.1.3 Mmi1 purification trial:***

Purifying Mmi1, either YTH domain or N-terminal domain was a challenging job. During purification of Mmi1-YTH, a strange phenomenon was observed. Concentrated protein precipitated when kept in ice or at 4°C. When the tube was transferred to room temperature,

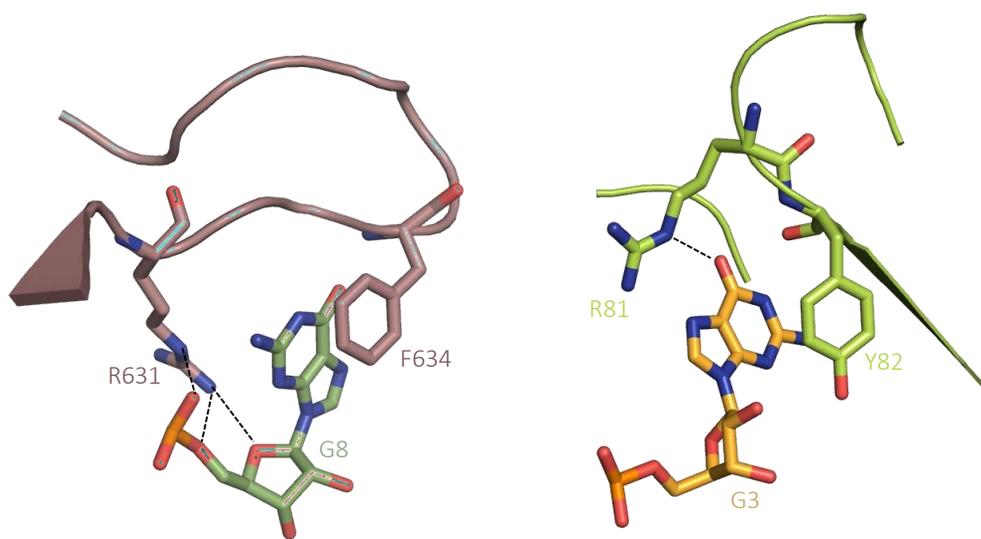


Fig. 87: Comparison of nucleotide recognition by residues located on loop of RRM domain. RRM3 Arg631 and Phe634 specifically recognizes G<sub>8</sub> (left) and quasiRRM Arg81 and Tyr82 specifically recognize G<sub>3</sub> (right). H-bonds are indicated by black dotted lines

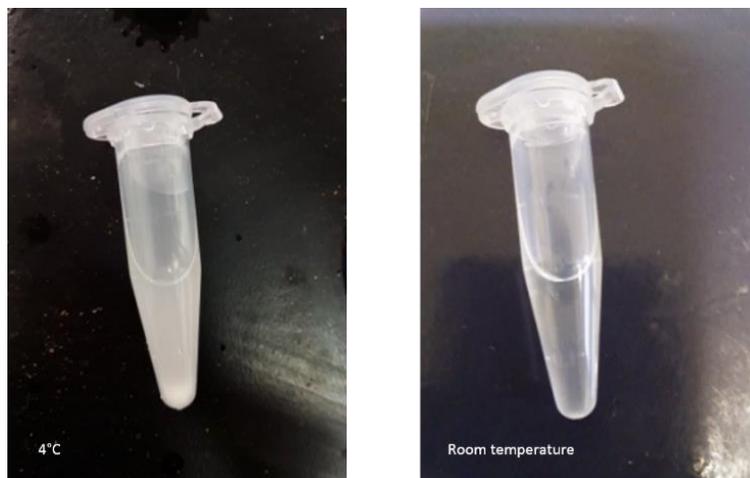


Fig. 88: Cold induced precipitation of Mmi1-YTH. Concentrated Mmi1 precipitates when kept in ice or 4°C (left); the precipitation disappears at room temperature

the precipitation disappeared (Fig. 88). This could be cold-triggered precipitation of protein or liquid-liquid phase separation (LLPS). Inside cell, biochemical reactions are compartmentalized in membrane-less liquid droplets that are phase separated from the cytoplasm. Stress granules and P-bodies are examples of such membrane-less liquid compartments. The process of initiation of LLPS is not well understood but there are evidences that constituent proteins with low-complexity domains are capable of mediating the process (Nott et al., 2015; Molliex et al., 2015). As Mmi1 is known to associate with stress granules (Rinnerthaler et al., 2013), it'd be interesting to investigate whether Mmi1 undergoes LLPS or if it is just cold induced precipitation.

#### **6.1.4 Role of Erh1:**

Erh1 is a small protein that was found to co-localize with Mmi1 during both mitosis and meiosis and it has also been shown by previous studies that EMC is important for down-regulation of meiotic genes, either by degradation or by heterochromatin formation. When this work was initiated, no information was available about which region of Mmi1 is interacting with Erh1. Like other YTH domain proteins, Mmi1 has an unstructured N-terminal domain. For crystallization, it is important to find a well-folded region. To do that multiple constructs of Mmi1-N terminal domain were generated. Using these constructs, co-expression assays were performed where Mmi1 constructs were His-tagged and Erh1 was untagged. From co-expression assays of different domains of Mmi1, it was found that Erh1 interacts with N-terminal domain (1-322 and 1-350) of Mmi1. When expression assays were performed for the individual constructs of Mmi1, no expression was detected, suggesting that Erh1 enhances the solubility of these Mmi1 constructs, probably due a direct interaction between both proteins. Multiple co-purification trials were performed with Mmi1 NTD and Erh1. After affinity purification, Mmi1 starts precipitating, numerous buffer optimizations were tried but it was not possible to prevent Mmi1 precipitation. To know the structure of Erh1, it was crystallized alone and the structure was solved. Erh1 was found to exist as a dimer. While this work was ongoing, another group published a structure of the Mmi1 (95-122)-Erh1 complex where Mmi1 is attached to Erh1 with a flexible linker (Xie et al., 2019). This structure revealed that dimerization of Mmi1 is mediated by Erh1 dimer. In the complex, Erh1 forms a dimer, and Mmi1 interacts with the dimer interface forming a dimer of heterodimers. The Erh1 dimerization surface is highly conserved but the role of Erh1 dimerization in its function was never investigated. In order to do that, a double mutant (I11R/L13R) was designed to disrupt Erh1 dimerization. *In vivo* studies performed by our collaborators indicated that disruption of

Erh1 dimerization affects some aspects of Mmi1's function but not all. When an *erh1+* deleted cell was supplemented by the double mutant, the meiotic transcripts accumulated, meaning Mmi1 failed to degrade the meiotic transcripts when Erh1 dimerization is impaired. On the other hand, Mmi1 was still capable of degrading Mei2 protein in absence Erh1 dimer. In a pull-down experiment performed by mixing Erh1 double mutant with Mmi1 (95-122), no interaction was detected among these two constructs. Surprisingly, *in vivo*, the Erh1 double mutant is still able to interact with full-length Mmi1, suggesting that the interaction between Erh1 and Mmi1 spans more than the residues identified (95-122).

### **6.1.5 *Erh1 and Mei2:***

In a former study (Harigaya et al., 2006), Erh1 and Mei2 were found to co-localize during meiosis and loss of Erh1 led to disruption of *mei2* loci. There was no evidence to show if these two proteins interact. To investigate the mode of interaction between these two proteins, a co-expression was performed with Mmi1, Erh1 and Mei2-CTD. But no interaction was detected between the two proteins.

The switch between mitosis and meiosis is tightly regulated by a complex network in *S. pombe* and the key players are Mmi1 and Mei2. The work performed as part of this thesis was focused in deciphering the mechanism by which the three proteins, Mmi1, Erh1 and Mei2 regulate each other. There is scope of further investigation to unravel the complex process and other partners involved in this process to regulate mitosis and meiosis switch.

### **6.2 *Conclusion:***

A complex regulatory mechanism exists between Mmi1, Mei2 and Erh1. Mmi1 and Erh1 form a tight complex, EMC which is responsible for degradation of meiotic transcripts as well as heterochromatin island formation at these genetic loci in mitotically growing cells. Mei2 sequesters Mmi1 during meiosis thereby inhibiting its function. Mei2 also binds to *mmi1+* transcript, repressing Mmi1 at mRNA level (Mukherjee et al., 2018). In an attempt to decipher this complex regulation, we tried to characterize Mmi1, Mei2 and Erh1. From our studies, a novel RRM domain was found that has a RNA binding mode different from any known RRM domain. No direct interaction was detected between Mmi1 and Mei2 RRM1-2 or RRM3. Physical interaction was detected between Erh1 and Mmi1 and the domain boundary was determined. Further studies revealed that dimerization of Erh1 is necessary for some functions of Mmi1. From experimental results, our hypothesis is, that the interaction between Mei2 and

Mmi1 is bridged by RNA. The regulation of mitosis to meiosis switch is an active field of research and further experiments are required to decipher this complex mechanism and all the proteins involved in it.

### ***6.3 Future prospect:***

The work carried out in this thesis sheds light on different aspects of the relationship between Mmi1, Mei2 and Erh1. Mei2 is the master regulator of meiosis. It is comprised of three RRM domains, RRM3 being the essential one. Our RNA bound structure of RRM3 shows that this RRM domain has a non-canonical RNA binding mode. The recognition of nucleotide by residues located on the loop (R631, F634) is similar to RNA recognition mode of quasiRRM domains. Additionally, RRM3 has a fifth  $\beta$  strand which may extend the RNA binding  $\beta$ -surface. But due to the lack of crystals of RRM3 with long RNA, the role of this fifth  $\beta$  remain unexplored thus opens a prospect for future work on this topic. Co-crystallization with a longer RNA fragment, site directed mutagenesis and ITC might be helpful to understand the role of the C-terminal extremities with more details

The first two RRM domains of Mei2 are closely spaced. RRM1-2 has been crystallized but not good enough to collect diffraction data sets. Crystal optimization to obtain the structure along with biophysical analysis like ITC, fluorescence polarization study could be performed to determine the RNA substrate of these RRM domains. Structural study of these two RRM domains, in apo and RNA bound state will contribute to better understanding of this protein.

Mmi1 forms a tight complex with Erh1. The recently published paper shows that 95 – 122 amino acid region of Mmi1 interacts with the dimeric interface of Erh1 (Xie et al., 2019). In vitro it fails to interact with the monomeric Erh1. But our in vivo studies show that Mmi1 is capable of interacting with monomeric Erh1, suggesting that the domain boundary of Mmi1 that interacts with Erh1 can be longer than 95-122 amino acid region. Determination of the interacting region could be interesting to have more detailed information about this complex. Different constructs of Mmi1 NTD (1 – 350) with a solubility tag might be helpful to find out the interacting region by pull down experiments. Also, it is worth to investigate whether Mmi1 is capable of liquid-liquid-phase separation, this can be performed by fluorescence spectroscopy and microscopy.

Mmi1 and Mei2 control each other's level of expression in cell. But the mode of interaction between these two proteins is still a mystery. So far, the only known common substrate of these

two proteins, is meiRNA. Our studies indicate there is no direct interaction between Mmi1-YTH domain and Mei2-RRM3 domain. However, it would be of interest to use in vitro transcribed meiRNA to investigate formation of the complex. The complex formation can be verified by SEC-MALLS and the components of the complex can be identified by mass spectrometry, this will help to identify new partners of this complex. It'd be interesting to determine the structure of this complex either by crystallography or Cryo-electron microscopy. This could provide insight about the mechanism of mitosis to meiosis switch regulation in *S. pombe*.



## CHAPTER VII

Discussion, conclusion and  
Future prospect of Pho92  
and Not1



### 7.1 *Pho92 and Not1:*

Pho92 is the only YTH domain protein in budding yeast. This protein has been selected for study as it is homologous to human YTHDF2. Both YTHDF2 and Pho92 are known to interact with Not1 of Ccr4-Not complex. Like *S. pombe* Mmi1, Pho92 degrades mRNAs during meiosis. So far, only one target of Pho92 has been identified, that is PHO4. An unpublished yeast two hybrid result from our collaborator Dr Bertrand Séraphin, indicated that Pho92 interacts with Not1 of Ccr4-Not complex. Not1 is a large protein of more than 2000 amino acids and it is not possible to produce a protein of this size in bacteria. Therefore, the main challenge was to identify the domain boundary of Not1 that interacts with Pho92. Not1 constructs initially were not soluble. A solubility tag was designed containing Z domain of protein-A from *S. aureus*. A recombinant vector was created by inserting this solubility tag and all Not1 constructs were strongly over-expressed using this His<sub>6</sub>-ZZ tag.

To determine the boundaries a series of pull down was performed with different constructs of Not1, in presence of benzonase. A domain encompassing residues 1348-1565 was found to be interacting with the N-terminal domain of Pho92. Like other YTH domain proteins, the N-terminal domain of Pho92 is a low complexity region. Hence, it was crucial to find a short compact region that will allow crystallization. In this pursuit, a short conserved stretch was discovered that interacts with Not1. Initial crystallization trials were performed with co-purified Not1 (1343-1565) and Pho92 (21-44). But the short peptide was degrading over time, changing the stoichiometry of the complex. To overcome this problem, we decided to try different organisms. A multiple sequence alignment was performed on diverse organisms that had the conserved Pho92 region. Then on the basis of secondary structure prediction, two organisms were selected as they seemed to have a more structured Pho92 homologous protein: *Candida glabrata* and *Zygosaccharomyces rouxii*. But co-purification trial from either of these organisms was not successful. The Pho92 homologue was degrading and the yield of Not1 was also lower than Not1 of *S. cerevisiae*.

The next attempt was limited proteolysis. Limited proteolysis is a standard technique to trim the flexible or unfolded parts of a protein to facilitate crystallization. For this trial, different constructs of Not1 and Pho92 were used. From the result, Not1 (1071-1565) and Pho92 (1-142) digested by Trypsin (1:200) was selected to be the optimal condition. After digestion by Trypsin, the protein complex was loaded onto a gel filtration column and peak fractions

containing the digested product of the complex were concentrated for crystallization trial. But no crystals were obtained from this trial.

In the next trial, purified Not1 was mixed with synthetic Pho92 peptide for co-crystallization trial. Crystals were obtained in two conditions from this trial. These crystals were very small and impossible to fish. Further trials are required for optimization of the crystallization condition.

In another attempt, following the Mmi1 (95-122)-Erh1 complex crystallization (Xie et al., 2019), Pho92 peptide was attached to Not1 by a (Gly-Ser-Ser)<sub>5</sub> linker. This construct was purified and crystallization trials are ongoing.

### **7.2 Conclusion:**

Our study has established that there is physical contact between Not1 and Pho92 and the domain boundary for this interaction were identified. This work is in early stages and further study is required to investigate the role of Pho92 in regulation of meiosis in yeast and to establish the mechanism by which Pho92 degraded its target transcripts. This knowledge will help us to understand the role of YTHDF2 in mRNA degradation in human beings.

### **7.3 Future prospect:**

In our study, we concluded that Not1 (1343-1565) is physically interacting with N-terminal domain of Pho92. Crystallization trials of this complex are ongoing. Structural study of Not1-Pho92 complex will give an insight of the mechanism by which Pho92 degrades its target transcripts by recruiting Ccr4-Not complex. Human YTHDF2 was also found to be interacting with NOT1. Therefore, this study will help to understand how the process is carried out in human beings as well. Ccr4-Not complex is a large multi-subunit complex and Pho92 is known to interact with at least one other member of this complex, Pop2. It'd be interesting to investigate if the three proteins, Not1, Pho92 and Pop2 form a complex which will give an insight about mRNA degradation process.

Human YTHDF2, *S. pombe* Mmi1 and *S. cerevisiae* Pho92 all three YTH domain proteins from three different organisms interact with Not1 of Ccr4-Not complex and recruits Ccr4-Not complex for degradation of target mRNAs. We identified the region of Pho92 that interacts with Not1. But this region is not conserved, either in human beings or in *S. pombe*. Structural studies of these YTH-NTD and Not1 will help to have a better understanding of their

interaction and to explore the mechanism by which YTH domain proteins degrade their substrate RNAs by recruiting Ccr4-Not complex.



# REFERENCES



Abby, E., Tourpin, S., Ribeiro, J., Daniel, K., Messiaen, S., Moison, D., ... & Toth, A. (2016). Implementation of meiosis prophase I programme requires a conserved retinoid-independent stabilizer of meiotic transcripts. *Nature communications*, 7, 10324.

Achsel, T., & Bagni, C. (2016). Cooperativity in RNA–protein interactions: the complex is more than the sum of its partners. *Current opinion in neurobiology*, 39, 146-151.

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

Agarwala, S. D., Blitzblau, H. G., Hochwagen, A., & Fink, G. R. (2012). RNA methylation by the MIS complex regulates a cell fate decision in yeast. *PLoS genetics*, 8(6), e1002732.

Álvarez, B., & Moreno, S. (2006). Fission yeast Tor2 promotes cell growth and represses cell differentiation. *Journal of cell science*, 119(21), 4475-4485.

Anders, M., Chelysheva, I., Goebel, I., Trenkner, T., Zhou, J., Mao, Y., ... & Ignatova, Z. (2018). Dynamic m6A methylation facilitates mRNA triaging to stress granules. *Life science alliance*, 1(4), e201800113.

Anderson, S. J., Kramer, M. C., Gosai, S. J., Yu, X., Vandivier, L. E., Nelson, A. D., ... & Gregory, B. D. (2018). N6-methyladenosine inhibits local ribonucleolytic cleavage to stabilize mRNAs in Arabidopsis. *Cell reports*, 25(5), 1146-1157.

Arango, D., Sturgill, D., Alhusaini, N., Dillman, A. A., Sweet, T. J., Hanson, G., ... & Fox, S. D. (2018). Acetylation of cytidine in mRNA promotes translation efficiency. *Cell*, 175(7), 1872-1886.

Arribas-Hernández, L., Bressendorff, S., Hansen, M. H., Poulsen, C., Erdmann, S., & Brodersen, P. (2018). An m6A-YTH module controls developmental timing and morphogenesis in Arabidopsis. *The Plant Cell*, 30(5), 952-967.

Bailey, A. S., Batista, P. J., Gold, R. S., Chen, Y. G., de Rooij, D. G., Chang, H. Y., & Fuller, M. T. (2017). The conserved RNA helicase YTHDC2 regulates the transition from proliferation to differentiation in the germline. *Elife*, 6, e26116.

- Baltz, A. G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., ... & Wyler, E. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Molecular cell*, *46*(5), 674-690.
- Basquin, J., Roudko, V. V., Rode, M., Basquin, C., Séraphin, B., & Conti, E. (2012). Architecture of the nuclease module of the yeast Ccr4-not complex: the Not1-Caf1-Ccr4 interaction. *Molecular cell*, *48*(2), 207-218.
- Bhaskar, V., Basquin, J., & Conti, E. (2015). Architecture of the ubiquitylation module of the yeast Ccr4-Not complex. *Structure*, *23*(5), 921-928.
- Bhaskar, V., Roudko, V., Basquin, J., Sharma, K., Urlaub, H., Séraphin, B., & Conti, E. (2013). Structure and RNA-binding properties of the Not1–Not2–Not5 module of the yeast Ccr4–Not complex. *Nature structural & molecular biology*, *20*(11), 1281.
- Bhat, S. S., Bielewicz, D., Jarmolowski, A., & Szweykowska-Kulinska, Z. (2018). N6-methyladenosine (m6A): Revisiting the Old with Focus on New, an Arabidopsis thaliana Centered Review. *Genes*, *9*(12), 596.
- Bodi, Z., Button, J. D., Grierson, D., & Fray, R. G. (2010). Yeast targets for mRNA methylation. *Nucleic acids research*, *38*(16), 5327-5335.
- Boland, A., Chen, Y., Raisch, T., Jonas, S., Kuzuoğlu-Öztürk, D., Wohlbold, L., ... & Izaurralde, E. (2013). Structure and assembly of the NOT module of the human CCR4–NOT complex. *Nature structural & molecular biology*, *20*(11), 1289.
- Bresch, C., Müller, G., & Egel, R. (1968). Genes involved in meiosis and sporulation of a yeast. *Molecular and General Genetics MGG*, *102*(4), 301-306.
- Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., ... & Mornon, J. P. (1997). Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cellular and Molecular Life Sciences CMLS*, *53*(8), 621-645.
- Carlile, T. M., Rojas-Duran, M. F., Zinshteyn, B., Shin, H., Bartoli, K. M., & Gilbert, W. V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*, *515*(7525), 143.

Castello, A., Fischer, B., Frese, C. K., Horos, R., Alleaume, A. M., Foehr, S., ... & Hentze, M. W. (2016). Comprehensive identification of RNA-binding domains in human cells. *Molecular cell*, 63(4), 696-710.

Chen, Y., & Varani, G. (2013). Engineering RNA-binding proteins for biology. *The FEBS journal*, 280(16), 3734-3754.

Clancy, M. J., Shambaugh, M. E., Timpte, C. S., & Bokar, J. A. (2002). Induction of sporulation in *Saccharomyces cerevisiae* leads to the formation of N<sup>6</sup>-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. *Nucleic acids research*, 30(20), 4509-4518.

Cléry, A., & Allain, F. H. T. (2012). From structure to function of RNA binding domains. *RNA binding proteins*, 137-58.

Collart, M. A., & Panasenko, O. O. (2012). The Ccr4–not complex. *Gene*, 492(1), 42-53.

Collart, M. A., & Struhl, K. (1993). CDC39, an essential nuclear protein that negatively regulates transcription and differentially affects the constitutive and inducible HIS3 promoters. *The EMBO journal*, 12(1), 177-186.

Conte, M. R., Grüne, T., Ghuman, J., Kelly, G., Ladas, A., Matthews, S., & Curry, S. (2000). Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. *The EMBO journal*, 19(12), 3132-3141.

Cotobal, C., Rodríguez-López, M., Duncan, C., Hasan, A., Yamashita, A., Yamamoto, M., ... & Mata, J. (2015). Role of Ccr4-Not complex in heterochromatin formation at meiotic genes and subtelomeres in fission yeast. *Epigenetics & chromatin*, 8(1), 28.

Daubner, G. M., Cléry, A., & Allain, F. H. (2013). RRM–RNA recognition: NMR or crystallography... and new findings. *Current opinion in structural biology*, 23(1), 100-108.

Denis, C. L., & Chen, J. (2003). The CCR4-NOT complex plays diverse roles in mRNA metabolism. *Progress in nucleic acid research and molecular biology*, 73, 221-250.

Desrosiers, R., Friderici, K., & Rottman, F. (1974). Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proceedings of the National Academy of Sciences*, 71(10), 3971-3975.

- Dominguez, C., Fiset, J. F., Chabot, B., & Allain, F. H. (2010). Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nature structural & molecular biology*, 17(7), 853.
- Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., ... & Sorek, R. (2012). Topology of the human and mouse m<sup>6</sup>A RNA methylomes revealed by m<sup>6</sup>A-seq. *Nature*, 485(7397), 201.
- Dominissini, D., Nachtergaele, S., Moshitch-Moshkovitz, S., Peer, E., Kol, N., Ben-Haim, M. S., ... & Zheng, G. (2016). The dynamic N<sup>1</sup>-methyladenosine methylome in eukaryotic messenger RNA. *Nature*, 530(7591), 441.
- Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., ... & Wu, L. (2016). YTHDF2 destabilizes m<sup>6</sup>A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex. *Nature communications*, 7, 12626.
- Edupuganti, R. R., Geiger, S., Lindeboom, R. G., Shi, H., Hsu, P. J., Lu, Z., ... & Müller, M. (2017). N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) recruits and repels proteins to regulate mRNA homeostasis. *Nature structural & molecular biology*, 24(10), 870.
- Egan, E. D., Braun, C. R., Gygi, S. P., & Moazed, D. (2014). Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex. *Rna*, 20(6), 867-881.
- Forsburg, S. L., & Nurse, P. (1991). Cell cycle regulation in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Annual review of cell biology*, 7(1), 227-256.
- Gilbert, W. V., Bell, T. A., & Schaening, C. (2016). Messenger RNA modifications: form, distribution, and function. *Science*, 352(6292), 1408-1412.
- Gokhale, N. S., McIntyre, A. B., McFadden, M. J., Roder, A. E., Kennedy, E. M., Gandara, J. A., ... & Ilkayeva, O. R. (2016). N<sup>6</sup>-methyladenosine in Flaviviridae viral RNA genomes regulates infection. *Cell host & microbe*, 20(5), 654-665.
- Harigaya, Y., Tanaka, H., Yamanaka, S., Tanaka, K., Watanabe, Y., Tsutsumi, C., ... & Yamamoto, M. (2006). Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature*, 442(7098), 45.
- Hartmann, A. M., Nayler, O., Schwaiger, F. W., Obermeier, A., & Stamm, S. (1999). The interaction and colocalization of Sam68 with the splicing-associated factor YT521-B in

nuclear dots is regulated by the Src family kinase p59fyn. *Molecular biology of the cell*, 10(11), 3909-3926.

Hausmann, I. U., Bodi, Z., Sanchez-Moran, E., Mongan, N. P., Archer, N., Fray, R. G., & Soller, M. (2016). m<sup>6</sup>A potentiates Sxl alternative pre-mRNA splicing for robust *Drosophila* sex determination. *Nature*, 540(7632), 301.

Hazra, D., Chapat, C., & Graille, M. (2019). m<sup>6</sup>A mRNA Destiny: Chained to the rhYTHm by the YTH-Containing Proteins. *Genes*, 10(1), 49.

Higuchi, T., Watanabe, Y., & Yamamoto, M. (2002). Protein kinase A regulates sexual development and gluconeogenesis through phosphorylation of the Zn finger transcriptional activator Rst2p in fission yeast. *Molecular and cellular biology*, 22(1), 1-11.

Hiriart, E., Vavasseur, A., Touat-Todeschini, L., Yamashita, A., Gilquin, B., Lambert, E., ... & Lachuer, J. (2012). Mmi1 RNA surveillance machinery directs RNAi complex RITS to specific meiotic genes in fission yeast. *The EMBO journal*, 31(10), 2296-2308.

Holm, L., & Laakso, L. M. (2016). Dali server update. *Nucleic acids research*, 44(W1), W351-W355.

Houseley, J., LaCava, J., & Tollervey, D. (2006). RNA-quality control by the exosome. *Nature reviews Molecular cell biology*, 7(7), 529.

Hsu, P. J., Zhu, Y., Ma, H., Guo, Y., Shi, X., Liu, Y., ... & Cheng, Y. (2017). Ythdc2 is an N<sup>6</sup>-methyladenosine binding protein that regulates mammalian spermatogenesis. *Cell research*, 27(9), 1115.

Imai, Y., Matsuo, N., Ogawa, S., Tohyama, M., & Takagi, T. (1998). Cloning of a gene, YT521, for a novel RNA splicing-related protein induced by hypoxia/reoxygenation. *Molecular brain research*, 53(1-2), 33-40.

Inouye, S., & Sahara, Y. (2009). Expression and purification of the calcium binding photoprotein mitrocomin using ZZ-domain as a soluble partner in *E. coli* cells. *Protein expression and purification*, 66(1), 52-57.

Jain, D., Puno, M. R., Meydan, C., Lailier, N., Mason, C. E., Lima, C. D., ... & Keeney, S. (2018). ketu mutant mice uncover an essential meiotic function for the ancient RNA helicase YTHDC2. *Elife*, 7, e30919.

- Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., ... & He, C. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature chemical biology*, 7(12), 885.
- Kadosh, D., & Struhl, K. (1998). Targeted recruitment of the Sin3-Rpd3 histone deacetylase complex generates a highly localized domain of repressed chromatin in vivo. *Molecular and cellular biology*, 18(9), 5121-5127.
- Kang, H. J., Jeong, S. J., Kim, K. N., Baek, I. J., Chang, M., Kang, C. M., ... & Yun, C. W. (2014). A novel protein, Pho92, has a conserved YTH domain and regulates phosphate metabolism by decreasing the mRNA stability of PHO4 in *Saccharomyces cerevisiae*. *Biochemical Journal*, 457(3), 391-400.
- Kasowitz, S. D., Ma, J., Anderson, S. J., Leu, N. A., Xu, Y., Gregory, B. D., ... & Wang, P. J. (2018). Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. *PLoS genetics*, 14(5), e1007412.
- Kennedy, E. M., Bogerd, H. P., Kornepati, A. V., Kang, D., Ghoshal, D., Marshall, J. B., ... & Cullen, B. R. (2016). Posttranscriptional m6A editing of HIV-1 mRNAs enhances viral gene expression. *Cell host & microbe*, 19(5), 675-685.
- Kitamura, K., & Shimoda, C. (1991). The *Schizosaccharomyces pombe* mam2 gene encodes a putative pheromone receptor which has a significant homology with the *Saccharomyces cerevisiae* Ste2 protein. *The EMBO journal*, 10(12), 3743-3751.
- Kitamura, K., Katayama, S., Dhut, S., Sato, M., Watanabe, Y., Yamamoto, M., & Toda, T. (2001). Phosphorylation of Mei2 and Ste11 by Pat1 kinase inhibits sexual differentiation via ubiquitin proteolysis and 14-3-3 protein in fission yeast. *Developmental cell*, 1(3), 389-399.
- Kjærulff, S., Andersen, N. R., Borup, M. T., & Nielsen, O. (2007). Cdk phosphorylation of the Ste11 transcription factor constrains differentiation-specific transcription to G1. *Genes & development*, 21(3), 347-359.
- Kretschmer, J., Rao, H., Hackert, P., Sloan, K. E., Höbartner, C., & Bohnsack, M. T. (2018). The m6A reader protein YTHDC2 interacts with the small ribosomal subunit and the 5'-3' exoribonuclease XRN1. *RNA*, 24(10), 1339-1350.

Krug, R. M., Morgan, M. A., & Shatkin, A. J. (1976). Influenza viral mRNA contains internal N6-methyladenosine and 5'-terminal 7-methylguanosine in cap structures. *Journal of virology*, *20*(1), 45-53.

Kunitomo, H., Higuchi, T., Iino, Y., & Yamamoto, M. (2000). A zinc-finger protein, Rst2p, regulates transcription of the fission yeast *ste11+* gene, which encodes a pivotal transcription factor for sexual development. *Molecular biology of the cell*, *11*(9), 3205-3217.

Lee, N. N., Chalamcharla, V. R., Reyes-Turcu, F., Mehta, S., Zofall, M., Balachandran, V., ... & Grewal, S. I. (2013). Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. *Cell*, *155*(5), 1061-1074.

Lence, T., Akhtar, J., Bayer, M., Schmid, K., Spindler, L., Ho, C. H., ... & Roignant, J. Y. (2016). m<sup>6</sup>A modulates neuronal functions and sex determination in *Drosophila*. *Nature*, *540*(7632), 242.

Li, A., Chen, Y. S., Ping, X. L., Yang, X., Xiao, W., Yang, Y., ... & Bhattarai, D. P. (2017). Cytoplasmic m<sup>6</sup>A reader YTHDF3 promotes mRNA translation. *Cell research*, *27*(3), 444.

Li, F., Zhao, D., Wu, J., & Shi, Y. (2014). Structure of the YTH domain of human YTHDF2 in complex with an m<sup>6</sup>A mononucleotide reveals an aromatic cage for m<sup>6</sup>A recognition. *Cell research*, *24*(12), 1490.

Li, P., & McLeod, M. (1996). Molecular mimicry in development: identification of *ste11+* as a substrate and *mei3+* as a pseudosubstrate inhibitor of *ran1+* kinase. *Cell*, *87*(5), 869-880.

Li, Z., Wang, R., Gao, Y., Wang, C., Zhao, L., Xu, N., ... & Crawford, N. M. (2017). The *Arabidopsis* CPSF30-L gene plays an essential role in nitrate signaling and regulates the nitrate transceptor gene NRT1. 1. *New Phytologist*, *216*(4), 1205-1222.

Lichinchi, G., Zhao, B. S., Wu, Y., Lu, Z., Qin, Y., He, C., & Rana, T. M. (2016). Dynamics of human and viral RNA methylation during Zika virus infection. *Cell host & microbe*, *20*(5), 666-673.

Lingel, A., & Sattler, M. (2005). Novel modes of protein-RNA recognition in the RNAi pathway. *Current opinion in structural biology*, *15*(1), 107-115.

- Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., ... & Dai, Q. (2014). A METTL3–METTL14 complex mediates mammalian nuclear RNA N<sup>6</sup>-adenosine methylation. *Nature chemical biology*, *10*(2), 93.
- Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., & Pan, T. (2015). N<sup>6</sup>-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature*, *518*(7540), 560.
- Lunde, B. M., Moore, C., & Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nature reviews Molecular cell biology*, *8*(6), 479.
- Luo, S., & Tong, L. (2014). Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. *Proceedings of the National Academy of Sciences*, *111*(38), 13834-13839.
- Maeda, T., Mochizuki, N., & Yamamoto, M. (1990). Adenylyl cyclase is dispensable for vegetative cell growth in the fission yeast *Schizosaccharomyces pombe*. *Proceedings of the National Academy of Sciences*, *87*(20), 7814-7818.
- Maris, C., Dominguez, C., & Allain, F. H. T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *The FEBS journal*, *272*(9), 2118-2131.
- Masliah, G., Barraud, P., & Allain, F. H. T. (2013). RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cellular and Molecular Life Sciences*, *70*(11), 1875-1895.
- Matsuo, T., Otsubo, Y., Urano, J., Tamanoi, F., & Yamamoto, M. (2007). Loss of the TOR kinase Tor2 mimics nitrogen starvation and activates the sexual development pathway in fission yeast. *Molecular and cellular biology*, *27*(8), 3154-3164.
- McLeod, M., & Beach, D. (1988). A specific inhibitor of the ran1<sup>+</sup> protein kinase regulates entry into meiosis in *Schizosaccharomyces pombe*. *Nature*, *332*(6164), 509.
- Mendel, M., Chen, K. M., Homolka, D., Gos, P., Pandey, R. R., McCarthy, A. A., & Pillai, R. S. (2018). Methylation of structured RNA by the m<sup>6</sup>A writer METTL16 is essential for mouse embryonic development. *Molecular cell*, *71*(6), 986-1000.

- Meyer, K. D., Patil, D. P., Zhou, J., Zinoviev, A., Skabkin, M. A., Elemento, O., ... & Jaffrey, S. R. (2015). 5' UTR m6A promotes cap-independent translation. *Cell*, *163*(4), 999-1010.
- Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., & Jaffrey, S. R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, *149*(7), 1635-1646.
- Mitchell, P., & Tollervey, D. (2000). mRNA stability in eukaryotes. *Current opinion in genetics & development*, *10*(2), 193-198.
- Mitchell, S. F., Jain, S., She, M., & Parker, R. (2013). Global analysis of yeast mRNPs. *Nature structural & molecular biology*, *20*(1), 127.
- Morohashi, K., Sahara, H., Watashi, K., Iwabata, K., Sunoki, T., Kuramochi, K., ... & Shimotohno, K. (2011). Cyclosporin A associated helicase-like protein facilitates the association of hepatitis C virus RNA polymerase with its cellular cyclophilin B. *PLoS one*, *6*(4), e18285.
- Moschall, R., Gaik, M., & Medenbach, J. (2017). Promiscuity in post-transcriptional control of gene expression: Drosophila sex-lethal and its regulatory partnerships. *FEBS letters*, *591*(11), 1471-1488.
- Mukherjee, K., Futcher, B., & Leatherwood, J. (2018). Mmi1 and rep2 mRNAs are novel RNA targets of the Mei2 RNA-binding protein during early meiosis in *Schizosaccharomyces pombe*. *Open biology*, *8*(9), 180110.
- Nayler, O., Hartmann, A. M., & Stamm, S. (2000). The ER repeat protein YT521-B localizes to a novel subnuclear compartment. *The Journal of cell biology*, *150*(5), 949-962.
- Nicastro, G., Taylor, I. A., & Ramos, A. (2015). KH-RNA interactions: back in the groove. *Current opinion in structural biology*, *30*, 63-70.
- Nurse, P., Thuriaux, P., & Nasmyth, K. (1976). Genetic control of the cell division cycle in the fission yeast *Schizosaccharomyces pombe*. *Molecular and General Genetics MGG*, *146*(2), 167-178.
- Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., ... & Allain, F. H. T. (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science*, *309*(5743), 2054-2057.

- Ok, S. H., Jeong, H. J., Bae, J. M., Shin, J. S., Luan, S., & Kim, K. N. (2005). Novel CIPK1-associated proteins in Arabidopsis contain an evolutionarily conserved C-terminal region that mediates nuclear localization. *Plant physiology*, *139*(1), 138-150.
- Otsubo, Y., & Yamamoto, M. (2010). TOR and sexual development in fission yeast. In *The Enzymes* (Vol. 27, pp. 229-250). Academic Press.
- Otsubo, Y., & Yamamoto, M. (2012). Signaling pathways for fission yeast sexual differentiation at a glance.
- Otsubo, Y., Yamashita, A., Ohno, H., & Yamamoto, M. (2014). S. pombe TORC1 activates the ubiquitin-proteasomal degradation of the meiotic regulator Mei2 in cooperation with Pat1 kinase. *J Cell Sci*, *127*(12), 2639-2646.
- Peer, E., Rechavi, G., & Dominissini, D. (2017). Epitranscriptomics: regulation of mRNA metabolism through modifications. *Current opinion in chemical biology*, *41*, 93-98.
- Pendleton, K. E., Chen, B., Liu, K., Hunter, O. V., Xie, Y., Tu, B. P., & Conrad, N. K. (2017). The U6 snRNA m6A methyltransferase METTL16 regulates SAM synthetase intron retention. *Cell*, *169*(5), 824-835.
- Ping, X. L., Sun, B. F., Wang, L., Xiao, W., Yang, X., Wang, W. J., ... & Zhao, X. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell research*, *24*(2), 177.
- Primig, M., Williams, R. M., Winzeler, E. A., Tevzadze, G. G., Conway, A. R., Hwang, S. Y., ... & Esposito, R. E. (2000). The core meiotic transcriptome in budding yeasts. *Nature genetics*, *26*(4), 415.
- Rao, S., Dinkins, R. D., & Hunt, A. G. (2009). Distinctive interactions of the Arabidopsis homolog of the 30 kD subunit of the cleavage and polyadenylation specificity factor (AtCPSF30) with other polyadenylation factor subunits. *BMC cell biology*, *10*(1), 51.
- Reyes-Turcu, F. E., & Grewal, S. I. (2012). Different means, same end—heterochromatin formation by RNAi and RNAi-independent RNA processing factors in fission yeast. *Current opinion in genetics & development*, *22*(2), 156-163.
- Roost, C., Lynch, S. R., Batista, P. J., Qu, K., Chang, H. Y., & Kool, E. T. (2015). Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. *Journal of the American Chemical Society*, *137*(5), 2107-2115.

- Roundtree, I. A., Luo, G. Z., Zhang, Z., Wang, X., Zhou, T., Cui, Y., ... & He, E. (2017). YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. *Elife*, *6*, e31311.
- Rundlett, S. E., Carmen, A. A., Suka, N., Turner, B. M., & Grunstein, M. (1998). Transcriptional repression by UME6 involves deacetylation of lysine 5 of histone H4 by RPD3. *Nature*, *392*(6678), 831.
- Safra, M., Sas-Chen, A., Nir, R., Winkler, R., Nachshon, A., Bar-Yaacov, D., ... & Schwartz, S. (2017). The m<sup>1</sup>A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature*, *551*(7679), 251.
- Salz, H. K., Maine, E. M., Keyes, L. N., Samuels, M. E., Cline, T. W., & Schedl, P. (1989). The *Drosophila* female-specific sex-determination gene, *Sex-lethal*, has stage-, tissue-, and sex-specific RNAs suggesting multiple modes of regulation. *Genes & development*, *3*(5), 708-719.
- Samuelsson, E., Moks, T., Uhlen, M., & Nilsson, B. (1994). Enhanced in vitro refolding of insulin-like growth factor I using a solubilizing fusion partner. *Biochemistry*, *33*(14), 4207-4211.
- Sato, M., Shinozaki-Yabana, S., Yamashita, A., Watanabe, Y., & Yamamoto, M. (2001). The fission yeast meiotic regulator Mei2p undergoes nucleocytoplasmic shuttling. *FEBS letters*, *499*(3), 251-255.
- Schwartz, S., Agarwala, S. D., Mumbach, M. R., Jovanovic, M., Mertins, P., Shishkin, A., ... & Carr, S. A. (2013). High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell*, *155*(6), 1409-1421.
- Schwartz, S., Bernstein, D. A., Mumbach, M. R., Jovanovic, M., Herbst, R. H., León-Ricardo, B. X., ... & Fink, G. (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell*, *159*(1), 148-162.
- Scutenaire, J., Deragon, J. M., Jean, V., Benhamed, M., Raynaud, C., Favory, J. J., ... & Bousquet-Antonelli, C. (2018). The YTH domain protein ECT2 is an m<sup>6</sup>A reader required for normal trichome branching in *Arabidopsis*. *The Plant Cell*, *30*(5), 986-1005.

- Shah, J. C., & Clancy, M. J. (1992). IME4, a gene that mediates MAT and nutritional control of meiosis in *Saccharomyces cerevisiae*. *Molecular and cellular biology*, *12*(3), 1078-1086.
- Shi, H., Zhang, X., Weng, Y. L., Lu, Z., Liu, Y., Lu, Z., ... & Wu, Y. (2018). m6A facilitates hippocampus-dependent learning and memory through YTHDF1. *Nature*, *563*(7730), 249.
- Shichino, Y., Otsubo, Y., Kimori, Y., Yamamoto, M., & Yamashita, A. (2018). YTH-RNA-binding protein prevents deleterious expression of meiotic proteins by tethering their mRNAs to nuclear foci. *Elife*, *7*, e32155.
- Shichino, Y., Yamashita, A., & Yamamoto, M. (2014). Meiotic long non-coding meiRNA accumulates as a dot at its genetic locus facilitated by Mmi1 and plays as a decoy to lure Mmi1. *Open biology*, *4*(6), 140022.
- Shima, H., Matsumoto, M., Ishigami, Y., Ebina, M., Muto, A., Sato, Y., ... & Igarashi, K. (2017). S-Adenosylmethionine synthesis is regulated by selective N6-adenosine methylation and mRNA degradation involving METTL16 and YTHDC1. *Cell reports*, *21*(12), 3354-3363.
- Simonetti, F., Candelli, T., Leon, S., Libri, D., & Rougemaille, M. (2017). Ubiquitination-dependent control of sexual differentiation in fission yeast. *Elife*, *6*, e28046.
- Simpson, P. J., Monie, T. P., Szendrői, A., Davydova, N., Tyzack, J. K., Conte, M. R., ... & Curry, S. (2004). Structure and RNA interactions of the N-terminal RRM domains of PTB. *Structure*, *12*(9), 1631-1643.
- Soh, Y. S., Mikedis, M. M., Kojima, M., Godfrey, A. K., de Rooij, D. G., & Page, D. C. (2017). Meioc maintains an extended meiotic prophase I in mice. *PLoS genetics*, *13*(4), e1006704.
- Squires, J. E., Patel, H. R., Nusch, M., Sibbritt, T., Humphreys, D. T., Parker, B. J., ... & Preiss, T. (2012). Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic acids research*, *40*(11), 5023-5033.
- Stoilov, P., Rafalska, I., & Stamm, S. (2002). YTH: a new domain in nuclear proteins. *Trends in biochemical sciences*, *27*(10), 495-497.

- Stowell, J. A., Wagstaff, J. L., Hill, C. H., Yu, M., McLaughlin, S. H., Freund, S. M., & Passmore, L. A. (2018). A low-complexity region in the YTH domain protein Mmi1 enhances RNA binding. *Journal of Biological Chemistry*, 293(24), 9210-9222.
- Stowell, J. A., Webster, M. W., Kögel, A., Wolf, J., Shelley, K. L., & Passmore, L. A. (2016). Reconstitution of targeted deadenylation by the Ccr4-Not complex and the YTH domain protein Mmi1. *Cell reports*, 17(8), 1978-1989.
- Sugiyama, T., & Sugioka-Sugiyama, R. (2011). Red1 promotes the elimination of meiosis-specific mRNAs in vegetatively growing fission yeast. *The EMBO Journal*, 30(6), 1027-1039.
- Sugiyama, T., Thillainadesan, G., Chalamcharla, V. R., Meng, Z., Balachandran, V., Dhakshnamoorthy, J., ... & Grewal, S. I. (2016). Enhancer of rudimentary cooperates with conserved RNA-processing factors to promote meiotic mRNA decay and facultative heterochromatin assembly. *Molecular cell*, 61(5), 747-759.
- Sukegawa, Y., Yamashita, A., & Yamamoto, M. (2011). The fission yeast stress-responsive MAPK pathway promotes meiosis via the phosphorylation of Pol II CTD in response to environmental and feedback cues. *PLoS genetics*, 7(12), e1002387.
- Tanabe, A., Konno, J., Tanikawa, K., & Sahara, H. (2014). Transcriptional machinery of TNF- $\alpha$ -inducible YTH domain containing 2 (YTHDC2) gene. *Gene*, 535(1), 24-32.
- Tanabe, A., Tanikawa, K., Tsunetomi, M., Takai, K., Ikeda, H., Konno, J., ... & Mori, M. (2016). RNA helicase YTHDC2 promotes cancer metastasis via the enhancement of the efficiency by which HIF-1 $\alpha$  mRNA is translated. *Cancer letters*, 376(1), 34-42.
- Tanaka, K., Davey, J., Imai, Y., & Yamamoto, M. (1993). *Schizosaccharomyces pombe* map3+ encodes the putative M-factor receptor. *Molecular and cellular biology*, 13(1), 80-88.
- Tashiro, S., Asano, T., Kanoh, J., & Ishikawa, F. (2013). Transcription-induced chromatin association of RNA surveillance factors mediates facultative heterochromatin formation in fission yeast. *Genes to Cells*, 18(4), 327-339.
- Theiler, D., Dominguez, C., Blatter, M., Boudet, J., & Allain, F. H. T. (2014). Solution structure of the YTH domain in complex with N6-methyladenosine RNA: a reader of methylated RNA. *Nucleic acids research*, 42(22), 13911-13919.

- Tirumuru, N., Zhao, B. S., Lu, W., Lu, Z., He, C., & Wu, L. (2016). N6-methyladenosine of HIV-1 RNA regulates viral infection and HIV-1 Gag protein expression. *Elife*, 5, e15528.
- Toda, T., Shimanuki, M., & Yanagida, M. (1991). Fission yeast genes that confer resistance to staurosporine encode an AP-1-like transcription factor and a protein kinase related to the mammalian ERK1/MAP2 and budding yeast FUS3 and KSS1 kinases. *Genes & development*, 5(1), 60-73.
- Tucker, M., & Parker, R. (2000). Mechanisms and control of mRNA decapping in *Saccharomyces cerevisiae*. *Annual review of biochemistry*, 69(1), 571-595.
- Ukleja, M., Cuellar, J., Siwaszek, A., Kasprzak, J. M., Czarnocki-Cieciura, M., Bujnicki, J. M., ... & Valpuesta, J. M. (2016). The architecture of the *Schizosaccharomyces pombe* CCR4-NOT complex. *Nature communications*, 7, 10433.
- Uritani, M., Hidaka, H., Hotta, Y., Ueno, M., Ushimaru, T., & Toda, T. (2006). Fission yeast Tor2 links nitrogen signals to cell proliferation and acts downstream of the Rheb GTPase. *Genes to Cells*, 11(12), 1367-1379.
- Valbuena, N., & Moreno, S. (2010). TOR and PKA pathways synergize at the level of the Ste11 transcription factor to prevent mating and meiosis in fission yeast. *PloS one*, 5(7), e11514.
- van Heeckeren, W. J., Dorris, D. R., & Struhl, K. (1998). The mating-type proteins of fission yeast induce meiosis by directly activating *mei3* transcription. *Molecular and cellular biology*, 18(12), 7317-7326.
- Wang, C., Zhu, Y., Bao, H., Jiang, Y., Xu, C., Wu, J., & Shi, Y. (2015). A novel RNA-binding mode of the YTH domain reveals the mechanism for recognition of determinant of selective removal by Mmi1. *Nucleic acids research*, 44(2), 969-982.
- Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., ... & Ren, B. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, 505(7481), 117.
- Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., ... & He, C. (2015). N6-methyladenosine modulates messenger RNA translation efficiency. *Cell*, 161(6), 1388-1399.

- Watanabe, Y., & Yamamoto, M. (1994). *S. pombe* mei2<sup>+</sup> encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. *Cell*, 78(3), 487-498.
- Watanabe, Y., Shinozaki-Yabana, S., Chikashige, Y., Hiraoka, Y., & Yamamoto, M. (1997). Phosphorylation of RNA-binding protein controls cell cycle switch from mitotic to meiotic in fission yeast. *Nature*, 386(6621), 187.
- Wei, L. H., Song, P., Wang, Y., Lu, Z., Tang, Q., Yu, Q., ... & Jia, G. (2018). The m6A reader ECT2 controls trichome morphology by affecting mRNA stability in Arabidopsis. *The Plant Cell*, 30(5), 968-985.
- Weisman, R., Roitburg, I., Schonbrun, M., Harari, R., & Kupiec, M. (2007). Opposite effects of tor1 and tor2 on nitrogen starvation responses in fission yeast. *Genetics*, 175(3), 1153-1162.
- Willer, M., Hoffmann, L., Styrkársdóttir, U., Egel, R., Davey, J., & Nielsen, O. (1995). Two-step activation of meiosis by the mat1 locus in *Schizosaccharomyces pombe*. *Molecular and cellular biology*, 15(9), 4964-4970.
- Wojtas, M. N., Pandey, R. R., Mendel, M., Homolka, D., Sachidanandam, R., & Pillai, R. S. (2017). Regulation of m6A transcripts by the 3→5 RNA helicase YTHDC2 is essential for a successful meiotic program in the mammalian germline. *Molecular cell*, 68(2), 374-387.
- Wu, B., Xu, J., Su, S., Liu, H., Gan, J., & Ma, J. (2017). Structural insights into the specific recognition of DSR by the YTH domain containing protein Mmi1. *Biochemical and biophysical research communications*, 491(2), 310-316.
- Xiao, W., Adhikari, S., Dahal, U., Chen, Y. S., Hao, Y. J., Sun, B. F., ... & Wang, X. (2016). Nuclear m6A reader YTHDC1 regulates mRNA splicing. *Molecular cell*, 61(4), 507-519.
- Xie, G., Vo, T. V., Thillainadesan, G., Holla, S., Zhang, B., Jiang, Y., ... & Shi, Y. (2019). A conserved dimer interface connects ERH and YTH family proteins to promote gene silencing. *Nature communications*, 10(1), 251.
- Xu, C., Liu, K., Ahmed, H., Loppnau, P., Schapira, M., & Min, J. (2015). Structural basis for the discriminative recognition of N6-methyladenosine RNA by the human YT521-B

homology domain family of proteins. *Journal of Biological Chemistry*, 290(41), 24902-24913.

Xu, C., Wang, X., Liu, K., Roundtree, I. A., Tempel, W., Li, Y., ... & Min, J. (2014). Structural basis for selective binding of m<sup>6</sup>A RNA by the YTHDC1 YTH domain. *Nature chemical biology*, 10(11), 927.

Yamamoto, M. (2010). The selective elimination of messenger RNA underlies the mitosis–meiosis switch in fission yeast. *Proceedings of the Japan Academy, Series B*, 86(8), 788-797.

Yamanaka, S., Mehta, S., Reyes-Turcu, F. E., Zhuang, F., Fuchs, R. T., Rong, Y., ... & Grewal, S. I. (2013). RNAi triggered by specialized machinery silences developmental genes and retrotransposons. *Nature*, 493(7433), 557.

Yamanaka, S., Yamashita, A., Harigaya, Y., Iwata, R., & Yamamoto, M. (2010). Importance of polyadenylation in the selective elimination of meiotic mRNAs in growing *S. pombe* cells. *The EMBO journal*, 29(13), 2173-2181.

Yamashita, A., Shichino, Y., Tanaka, H., Hiriart, E., Touat-Todeschini, L., Vavasseur, A., ... & Yamamoto, M. (2012). Hexanucleotide motifs mediate recruitment of the RNA elimination machinery to silent meiotic genes. *Open biology*, 2(3), 120014.

Yamashita, A., Watanabe, Y., Nukina, N., & Yamamoto, M. (1998). RNA-assisted nuclear transport of the meiotic regulator Mei2p in fission yeast. *Cell*, 95(1), 115-123.

Zhang, Z., Theler, D., Kaminska, K. H., Hiller, M., de la Grange, P., Pudimat, R., ... & Stamm, S. (2010). The YTH domain is a novel RNA binding domain. *Journal of Biological Chemistry*, 285(19), 14701-14710.

Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., ... & Lu, Z. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Molecular cell*, 49(1), 18-29.

Zheng, L., Baumann, U., & Reymond, J. L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic acids research*, 32(14), e115-e115.

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S. R., & Qian, S. B. (2015). Dynamic m<sup>6</sup>A mRNA methylation directs translational control of heat shock response. *Nature*, 526(7574), 591.

Zhou, Y., Zhu, J., Schermann, G., Ohle, C., Bendrin, K., Sugioka-Sugiyama, R., ... & Fischer, T. (2015). The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome. *Nature communications*, 6, 7050.

Zofall, M., Yamanaka, S., Reyes-Turcu, F. E., Zhang, K., Rubin, C., & Grewal, S. I. (2012). RNA elimination machinery targeting meiotic mRNAs promotes facultative heterochromatin formation. *Science*, 335(6064), 96-100.



# RÉSUMÉ DE THÈSE



## **Contrôle de la dégradation des ARNm par les protéines YTH pendant la transition de la méiose vers la mitose chez les levures.**

Le cycle cellulaire est contrôlé par des processus complexes et interconnectés. Lorsqu'un gène est transcrit en ARN messagers (ou ARNm) lui-même traduit en protéines, de nombreux processus de régulation travaillent pour contrôler chaque étape de ces processus apparemment simples. Parmi ces points de contrôle, la régulation post-transcriptionnelle est importante, et la formation d'un complexe protéine-ARN peut influencer le destin cellulaire. Parmi ces protéines de liaison à l'ARN, les protéines contenant des domaines YTH n'ont été découvertes qu'à la fin des années 90. Les protéines contenant des domaines YTH sont abondantes chez les eucaryotes et absentes chez les procaryotes. Elles constituent la majorité des protéines « readers » capables de reconnaître spécifiquement la modification m<sup>6</sup>A. L'Homme possède cinq protéines YTH, YTHDF1-3, YTHDC1,2 (Hazra, D., C. Chapat, et Graille, M. (2019). Destin de l'ARNm de m<sup>6</sup>A : enchaînés au rythme par les protéines contenant de la YTH. , 10 (1), 49.), l'organisme végétal modèle *Arabidopsis thaliana* 13, tandis qu'une seule de ces protéines (Pho92 et Mmi1, respectivement) est présente chez la levure de boulanger (*Saccharomyces cerevisiae*) ou la levure fissionnaire (*Schizosaccharomyces pombe*). Bien qu'il soit évident que ces protéines contrôlent le destin cellulaire, la fonction de chaque protéine et son réseau d'interaction restent à élucider. Hormis le domaine YTH, il n'y a pas d'homologie de séquence entre ces deux protéines de levures mais leur fonction cellulaire est similaire.

Il est bien établi que Mmi1 est responsable de la dégradation des transcrits spécifiques de la méiose au cours de la croissance végétative des cellules chez la levure *S. pombe*. Mmi1 s'associe avec la protéine Erh1 pour former un complexe stable Erh1-Mmi1 (également appelé EMC). Le complexe EMC peut physiquement interagir avec la sous-unité Not1 du complexe CCR4-Not et la recruter pour la dégradation des ARNm contenant des motifs DSR (déterminant de l'élimination sélective). L'action de Mmi1 est à son tour régulée par la protéine Mei2 qui possède trois domaines RRM (pour RNA Recognition Motif). Au cours de la méiose, Mei2, avec l'aide de l'ARN long non codant (lncRNA) meiRNA, séquestre Mmi1 dans un foyer nucléaire, le rendant inactif. Ceci permet de stabiliser les ARNm codants pour des protéines méiotiques et d'assurer la continuité de la méiose. Ces trois protéines Mmi1-Erh1-Mei2 jouent donc un rôle clé dans la transition de la mitose vers la méiose.

Chez *S. cerevisiae*, Pho92 est impliquée dans la dégradation des transcrits du gène PHO4, contribuant au contrôle de la voie du métabolisme du phosphate, pendant la privation en phosphate et participe également à la dégradation des ARNm contenant les marques épitranscriptomiques de N6-méthyladénosine (m6A). Comme pour la protéine Mmi1 de *S. pombe*, la protéine Pho92 recrute le complexe CCR4-Not via une interaction physique avec Not1.

Au cours de ma thèse, j'ai tenté d'élucider le rôle de ces deux protéines à domaine YTH de deux organismes modèles eucaryotes (*S. cerevisiae* et *S. pombe*) dans la dégradation de l'ARNm et la régulation du cycle cellulaire par des approches de biochimie et de biologie structurale.

La protéine Pho92 de *S. cerevisiae* interagit physiquement avec la protéine Not1 qui joue un rôle central dans l'assemblage du complexe CCR4-Not. En combinant des approches de co-expression de divers fragments des protéines d'intérêt chez *E. coli* et de co-purification avec l'aide d'étiquettes (hexahistidine ou GST), j'ai pu déterminer les limites des domaines de ces deux protéines qui sont suffisants et nécessaires pour cette interaction. L'interaction entre ces deux protéines a été étudiée par anisotropie de fluorescence. Le complexe protéique a été purifié avec succès et des essais de cristallisation sont en cours.

Chez *S. pombe*, j'ai étudié troisième domaine RRM de la protéine Mei2, domaine qui est le plus important pour la fonction biologique de cette protéine. J'ai déterminé la structure tridimensionnelle de ce domaine par la technique de cristallographie aux rayons X. Grâce à la comparaison de cette structure à celles d'autres domaines RRM déjà connus, j'ai pu valider par des expériences de biochimie que ce domaine fixe des ARN et également identifier une séquence d'ARN ayant une forte affinité pour ce domaine. J'ai ensuite déterminé la structure tridimensionnelle du complexe entre ce domaine RRM3 et un ARN. J'ai enfin résolu la structure de la protéine Erh1, révélant une organisation en homodimère. En combinant des expériences in vivo et in vitro, avec mes collaborateurs, nous avons pu montrer que la formation de ces homodimères est cruciale pour la fonction biologique d'Erh1 mais aussi de son partenaire Mmi1.



**Titre :** Contrôle de la dégradation des ARNm par les protéines YTH lors de la transition mitose-méiose chez les levures.

**Mots clés :** Epitranscriptomique, dégradation des ARNm, méiose, complexes multi-protéiques, domaine YTH.

**Résumé :** Chez les eucaryotes, les modifications post-transcriptionnelles des ARNm offrent un niveau de régulation de l'expression des gènes qui a été découvert récemment. Différentes protéines fixent les ARN sur des séquences spécifiques et déterminent leur devenir cellulaire: localisation, stabilité, transport, traduction et fonction dans la cellule. Un nouveau domaine de fixation aux ARN appelé domaine YTH a été découvert en 2011 et détecte spécifiquement les nucléotides modifiés N6-méthyl-adenosine (m<sup>6</sup>A) sur les ARNm. Ces signaux m<sup>6</sup>A constituent la plus abondante et dynamique modification des ARNm et sont reliés à différents types de cancers et de troubles développementaux. Il s'agit d'un domaine actif de recherche et jusqu'à présent, seules quelques protéines YTH ont été caractérisées. Chez les levures *Schizosaccharomyces pombe* et *Saccharomyces cerevisiae*, les protéines YTH jouent un rôle dans la dégradation d'ARNm impliqués dans la transition de la mitose vers la méiose.

Afin de comprendre le rôle de ces protéines YTH dans ces mécanismes, nous avons montré que la protéine YTH Pho92 de *S. cerevisiae* interagit avec Not1 pour accélérer la dégradation des substrats ARNm en recrutant le complexe Ccr4-Not. La levure *S. pombe* possède une protéine YTH atypique, Mmi1 qui dégrade les ARNm méiotiques dans les cellules en croissance mitotique. Nous avons montré que Mmi1 interagit avec Erh1 et déterminé la structure cristalline de Erh1, qui est nécessaire pour la dégradation des ARNm dépendante de Mmi1. Lors de la méiose, la protéine Mei2 séquestre Mmi1. Nous avons déterminé la structure cristalline du domaine fonctionnel de Mei2 et avons identifié une séquence d'ARN fixant ce domaine de Mei2 par ITC. La structure du complexe entre ce domaine et cette séquence d'ARN a été déterminée permettant de comprendre le mode de reconnaissance de l'ARN par Mei2.

**Title :** Insights into the control of mRNA decay by YTH proteins during the transition from mitosis to meiosis in yeasts

**Keywords :** Epitranscriptomics, mRNA decay, meiosis, multi-protein complexes, YTH domain

**Abstract:** In eukaryotes, post-transcriptional modifications of mRNA add an extra layer of regulation of gene expression. Different proteins interact with RNA based on sequence specificity and determine the cellular fate of the RNA, its localization, stability, transport, translation and function in cell. The novel RNA binding domain, discovered in 2011 is YTH domain. Majority of the YTH domains are readers of m<sup>6</sup>A mark on mRNA. This m<sup>6</sup>A mark is the most abundant dynamic modification of mRNA and was found to be linked with different types of cancer and developmental disorders. This is an active field of research and so far only a few YTH domain has been characterized. In the two yeasts *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*, the YTH domain proteins are involved in mRNA decay during mitosis to meiosis transition. In an attempt to decipher the underlying molecular mechanism, we showed that Pho92 of *S. cerevisiae* interacts with Not1 to accelerate the degradation of the RNA substrates by recruiting the Ccr4-Not complex. *S. pombe* has an atypical YTH domain protein, Mmi1, which degrades meiotic RNAs in mitotically growing cells. We found that Mmi1 interacts with Erh1 and determined the crystal structure of Erh1, which is required for Mmi1 dependent mRNA decay. During meiosis, Mei2, sequesters Mmi1. We have determined the crystal structure of the functional domain of Mei2. The RNA substrate of this Mei2 domain was determined by ITC and the structure of Mei2-RNA was solved, giving an indication about the binding motif of RNA to Mei2.