



HAL
open science

Etudes fonctionnelles sur le composant de la voie des piRNA TDRD1

Nikolaos Mathioudakis

► **To cite this version:**

Nikolaos Mathioudakis. Etudes fonctionnelles sur le composant de la voie des piRNA TDRD1. Autre [q-bio.OT]. Université de Grenoble, 2012. Français. NNT : 2012GRENV081 . tel-00907417

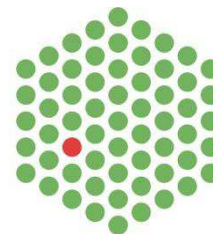
HAL Id: tel-00907417

<https://theses.hal.science/tel-00907417>

Submitted on 21 Nov 2013

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.



THESIS/THÈSE

To obtain the title of/Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Discipline/Spécialité : **Structural Biology - Nanobiology/Biologie Structurale et Nanobiologie**

Arrêté ministériel : 7 août 2006

Presented by/Présentée par

Nikolas Mathioudakis

Thesis supervisor/Thèse dirigée par **Stephen Cusack**

Thesis prepared at/Thèse préparée au sein du
European Molecular Biology Laboratory (EMBL), Grenoble Outstation

In/dans l'École Doctorale de Chimie et Sciences du vivant

Structural and functional studies on the piRNA pathway component TDRD1

Études structurales et fonctionnelles sur le composant TDRD1 de la voie des piRNA

Public defense/Thèse soutenue publiquement le **25/09/2012**
Jury members/devant le jury composé de :

Prof. Michael Sattler

Rapporteur

Dr. Andres Ramos

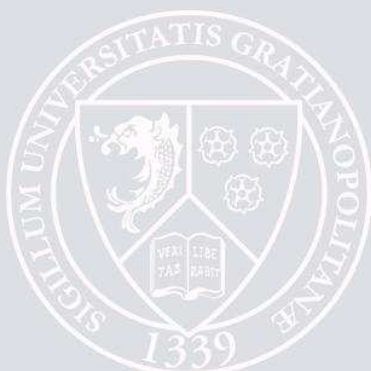
Rapporteur

Dr. Carlo Petosa

Examineur

Prof. Winfried Weissenhorn

Président



Summary

Piwi-interacting RNAs (piRNAs) are small non-coding RNAs expressed in the germ line of animals. They associate with Argonaute proteins of the PIWI subfamily, forming ribonucleoprotein complexes that are involved in maintaining genome integrity. The N-terminal region of some PIWI proteins contains symmetrically dimethylated arginines. This symmetrical dimethylation is thought to be responsible for the recruitment of Tudor domain-containing proteins (TDRDs), which might serve as platforms mediating interactions between various proteins in the piRNA pathway. We measured the binding affinity of the four individual extended Tudor domains (TDs) of murine TDRD1 protein for three different methyl-arginine containing peptides from murine PIWI protein Mili. The results show a preference of TD2 and TD3 for consecutive Mili peptides whereas TD4 and TD1 have respectively lower and very weak affinity for any peptide. These observations were confirmed by pull-down experiments with endogenous PIWI and PIWI-associated proteins. The affinity of TD1 for methyl-arginine peptides can be restored by a single point mutation in the aromatic cage back to the consensus sequence. The crystal structure of TD3 bound to a methylated Mili peptide shows an unexpected, non-canonical orientation of the bound peptide and of the methylated arginine side-chain in the aromatic cage. Finally, the SAXS model of the four tandem Tudor domains of TDRD1 reveals a flexible, elongated shape of the protein. Overall the results show that TDRD1 can accommodate different peptides from different proteins and therefore act as a scaffold protein in the piRNA pathway.

FK506-binding protein 6 (FKBP6) was recently identified as a novel piRNA pathway associated factor. It is comprised of an isomerase FK domain and a tetratricopeptide domain (TPR) and loss of *Fkbp6* results in transposons de-repression and male sterility in mouse. The TPR domain is implicated in interaction with the chaperone Hsp90 and the FK domain is an inactive isomerase that has evolved into a structural module. We performed preliminary biochemical experiments that identify the N-terminal MYND domain of TDRD1 as the interaction partner of the FK domain of FKBP6. Overall, we propose that TDRD1 protein is a molecular platform that recognizes multiple methylation marks of Mili and recruits FKBP6 for the promotion of a complex formation indispensable for the proper function of the piRNA pathway.

Résumé

Les ARN interagissant avec Piwi (ARNpi) sont des petits ARN non-codants qui sont exprimés dans la lignée germinale des animaux. Ils interagissent avec les protéines de la branche Piwi de la famille des Argonautes en formant des complexes des ribonucléoprotéines impliqués dans le maintien de l'intégrité du génome. La région N-terminale de quelques protéines Piwi contiennent symétriquement des arginines diméthylées. Il est considéré que ce statut symétrique de la diméthylation est responsable du recrutement des protéines possédant des domaines Tudor (TDRDs). Ces domaines peuvent avoir un rôle comme plateforme pour médier les interactions entre les protéines de la voie de l'ARNpi. Nous avons mesuré individuellement l'affinité de liaison des quatre domaines étendus Tudor (TD) de la protéine murine TDRD1 pour les trois différents peptides de la protéine murine Mili qui contiennent de la méthyl-arginine. Les résultats montrent une préférence des TD2 et TD3 pour les peptides consécutifs Mili alors que TD4 et TD1 ont une affinité plus basse et plus faible respectivement pour tous les peptides. Ces observations ont été confirmées par des expériences pull-down en utilisant des protéines Piwi endogènes et des protéines interagissant avec Piwi. L'affinité de TD1 pour les peptides qui contiennent de la méthyl-arginine peut être restaurée par une seule mutation ponctuelle dans la cage aromatique pour revenir à la séquence consensus. La structure de cristal de la protéine TD3 liée au peptide méthyle Mili montre une orientation inattendue de la peptide de liaison et de la chaîne latérale de l'arginine méthyle dans la cage aromatique. Finalement, le modèle SAXS des quatre domaines tandem Tudor de TDRD1 révèle une forme de la protéine flexible et allongée. Globalement, les résultats montrent que la protéine TDRD1 peut accommoder des différents peptides des différentes protéines et ainsi de fonctionner comme une protéine d'échafaudage dans la voie de l'ARNpi.

La protéine FKBP6 (FK506 Binding Protein) a été récemment identifiée comme un nouveau facteur interagissant dans la voie de l'ARNpi. FKBP6 est constituée d'un domaine d'isomérase FK et d'un domaine de tétratricopeptide (TPR). Une perte de la *Fkbp6* conduit à la dérépression des transposons et à la stérilité masculine des souris. Le domaine TPR est impliqué dans l'interaction avec la protéine chaperonne Hsp90 et le domaine FK est une isomérase inactive qui a évolué à un module structurale. En effectuant des expériences biochimiques préliminaires nous avons identifié la région N-terminale du domaine MYND de TDRD1 comme le partenaire d'interaction du domaine FK de la FKBP6. Nous proposons que la protéine TDRD1 est une plateforme moléculaire qui reconnaît des marques de méthylation de MILI et elle recrute FKBP6 pour promouvoir la formation d'un complexe indispensable pour la fonction de la voie de l'ARNpi.

TABLE OF CONTENTS

1 – INTRODUCTION	10
1.1 - Small non-coding RNAs	13
1.1.1 - The Argonaute family of proteins.....	14
1.1.2 - MicroRNAs (miRNAs)	16
1.1.2.1 – Biogenesis.....	17
1.1.2.2 – Function.....	18
1.1.2.3 – Biological importance.....	21
1.1.3 – Small interfering RNAs (siRNAs).....	21
1.1.3.1 – Biogenesis.....	21
1.1.3.2 – Function.....	23
1.1.3.3 – Biological importance.....	24
1.2 – Piwi-interacting RNAs (piRNAs)	25
1.2.1 – General features.....	25
1.2.2 – Origin and expression pattern	26
1.2.3 – Biogenesis.....	29
1.2.3.1 – Primary pathway.....	29
1.2.3.2 – Secondary biogenesis (ping-pong model)	31
1.2.4 – Regulation by PIWI-interacting proteins	34
1.2.4.1 – Tudor domain proteins.....	34
1.2.4.2 – Additional PIWI-associated proteins	40
1.2.5 – Sub-cellular localization of piRNA pathway components.....	41
1.2.6 – Function.....	45

1.2.6.1 – Post-transcriptional transposon silencing	45
1.2.6.2 – Transcriptional transposon silencing	47
1.2.6.3 – Silencing of protein-coding genes	48
1.2.7 – Summary	48
1.3– Post-translational modification of proteins	49
1.3.1 – Methylation as a protein modification	49
1.3.2 – “Royal domain superfamily of proteins”	50
1.3.2.1 – Tudor domains as methyl-arginine/lysine readers	51
1.4 – Aim of the thesis.....	57
2 – MATERIALS AND METHODS.....	58
2.1 – Genetic engineering methods	59
2.2 – Protein expression and purification	62
2.2.1 – Recombinant protein expression in <i>E.coli</i>	62
2.2.2 – Expression of Se-Methionine-substituted proteins.....	65
2.2.3 – Protein purification	65
2.3 – Biochemical methods	68
2.3.1 – Western blot	68
2.3.2 – Blotting for N-terminal sequencing.....	68
2.3.3 – Limited proteolysis.....	69
2.3.4 - Pull-down assay using recombinant proteins.....	69
2.3.5 – Pull-down assay using <i>in vitro</i> translation system.....	70

2.3.6 – Pull-down assays with endogenous proteins	71
2.3.7 – Isothermal Titration Calorimetry	72
2.4 – Small Angle X-ray Scattering (SAXS)	73
2.4.1 Sample preparation, SAXS data collection and analysis	73
2.5 – X-ray crystallography	74
2.5.1 – Protein crystallization and crystal freezing	74
2.5.2 – Data collection	75
2.5.3 – Data processing	75
2.5.4 – Refinement and structure analysis	76
3 – RESULTS	77
3.1 – Structural and functional studies on TDRD1 - a molecular scaffold in the piRNA pathway	80
3.1.1 – Expression, purification and crystallization trials of tandem eTudor domains of TDRD1	80
3.1.2 – Identification and crystallization of single eTudor domains	83
3.1.3 – TDRD1 is involved in the piRNA pathway and interacts with methylated arginines at the N-terminus of Mili	92
3.1.4 – Recognition of sDMA peptides by single eTudor domains of TDRD1	94
3.1.5 – Crystallization trials of eTudor domains-peptide complexes	101
3.1.6 – Crystal structure of the TD3- R45me2 peptide complex	102
3.1.6.1 – Crystallization and data analysis	102
3.1.6.2 – Description of the TD3-R45me2 structure	104
3.1.6.3 – Recognition of the sDMA by the aromatic cage	106
3.1.6.4 – Orientation of the R45me2 peptide in the TD3 domain structure.	107

3.1.7 – Structure of the unliganded TDRD1 TD3.....	112
3.1.7.1 – Crystallization and structure determination	112
3.1.7.2 – Structure description.....	112
3.1.8 – Conserved residues in eTudor domains.....	113
3.1.9 – ‘Active’ and ‘inactive’ eTudor domains.....	116
3.1.9.1 – Characterization of ‘active’ – ‘inactive’ eTudor domains of TDRD1	116
3.1.9.2 – Interaction of endogenous murine proteins with TDRD1 eTudor- domains	120
3.1.10 – SAXS analysis of the tandem eTudor domains of TDRD1	124
3.2 – Characterization of the MYND domain of TDRD1 as a protein-protein interaction module	128
3.2.1 – Introduction	128
3.2.1.1 – FK506-binding proteins (FKBPs)	128
3.2.1.2 – FKBP6 is a novel piRNA pathway-associated factor	129
3.2.1.3 – FKBP6 as a co-chaperone.....	131
3.2.1.4 – The FK domain of FKBP6 is inactive as an isomerase	132
3.2.1.5 – Description of MYND domains	134
3.2.2 – Results	136
3.2.2.1 – Expression and purification of FKBP6 and TDRD1’s N-terminal fragments	136
3.2.2.2 – The MYND domain of TDRD1 interacts with the FK domain of FKBP6	139
4 – DISCUSSION AND FUTURE PERSPECTIVES.....	141
5 - REFERENCES.....	155
6 - ACKNOWLEDGEMENTS	183

1 – INTRODUCTION

Summary of chapter Introduction

Small non-coding RNAs associate with Argonaute proteins to form the RNA-induced silencing complex (RISC), which mediate transcriptional or post-transcriptional silencing of their nucleic acid targets. Based on their sequence conservation, Argonautes can be partitioned into the Ago and PIWI clades. Ago members associate with ~21 nucleotide microRNAs (miRNAs) or small interfering RNAs (siRNAs) and are responsible for the control of gene expression in all cell types. On the other hand, the PIWIs are restricted to the animal gonads where they associate with ~24-31 nucleotides Piwi-interacting RNAs (piRNAs). Together, they are implicated in controlling transposon mobility in worms, flies, fishes, amphibians and mammals. Loss of PIWI proteins or other factors acting in the pathway leads to the activation of transposons, resulting in DNA damage, and finally causing sterility of the individual.

Ago and PIWI clade members differ in their N-terminal domains. PIWI proteins (Mili, Miwi and Miwi2 in mouse) uniquely carry multiple arginine-glycine (RG) dipeptides. Several of these arginines are post-translationally modified by methylation, which is catalyzed by the methyltransferase PRMT5. These modifications act as affinity tags for members of the Tudor domain protein family, which belongs to the “royal family” of modules that can specifically recognize methylated ligands to enable protein-protein interactions. Tudor domain-containing (TDRD) proteins contain one or more extended Tudor domains (eTud) and their conserved aromatic cage binds symmetrically dimethylated arginines (sDMAs) of the PIWI proteins in *Drosophila* and in mouse. All these suggest the potential to generate a dynamic network of interactions and assemblies that carry unique functions. The importance of TDRD proteins for germline development is highlighted by the sterility observed in their mutants, which is often accompanied by transposons de-repression and mislocalization.

TDRD1 is a multi-domain protein with an N-terminal MYND zinc-finger domain followed by four tandem eTud domains and is reported to interact with the sDMAs at the N-terminus of Mili. The scope of this thesis is to investigate the structural aspects of TDRD1-MILI interaction and how eTud domains recognize methylated arginines. Finally, preliminary characterization of the MYND domain as a protein-protein interaction module in the piRNA pathway is also given.

Résumé du chapitre “Introduction”

Les petits ARN non-codants interagissent avec les protéines Argonautes afin de former les complexes RISC (RNA-induced silencing complex) qui interviennent au silençage transcriptionnel et post-transcriptionnel des acides nucléiques. Les Argonautes peuvent être classifiées selon leur séquence de conservation aux protéines Ago et PIWI. Les Ago interagissent avec des microARNs (miARN) de ~21 nucléotides ou avec des petits ARN interférents (pARNi) et sont responsables pour le contrôle de l’expression des gènes dans tous les types cellulaires. En revanche, les protéines Piwi sont présentes seulement dans les gonades animales ou elles interagissent avec les ARN interagissants avec Piwi (ARNpi). Ensemble, elles sont impliquées dans la mobilité contrôlée des transposons dans les vers, mouches, poissons, amphibiens and mammifères. Perte de la fonction des PIWI ou des autres facteurs qui participent à ce voie conduit à l’activation des transposons qui provoque des dommages de l’ADN et finalement la stérilité.

Les membres Ago et PIWI diffèrent à leur N-domaines. Les protéines PIWI (Mili, Miwi et Miwi2 chez la souris) sont les seules qui ont des dipeptides arginine-glycine (RG). Il a été montré que quelques-unes de ces arginines sont modifiées au niveau post-translational par méthylation catalysée par la méthyltransferase PRMT5. Ces modifications peuvent être des marqueurs d’affinité pour les membres de la famille des protéines Tudor. Ces membres font partie de la ‘famille royale’ des modules qui peuvent spécifiquement reconnaître des ligands méthylés pour permettre des interactions protéine-protéine. Les TDRD (Tudor-domain-containing proteins) peuvent avoir un ou plus domaines Tudor étendus (eTud) et leur cage aromatique conservé est lié symétriquement aux arginines diméthylées (sDMAs) des protéines Piwi chez la Drosophila et la souris. Toutes ces informations suggèrent le potentiel de générer un réseau dynamique d’interactions et assemblées avec des fonctions uniques. L’importance des protéines TDRD dans le développement de la ligne germinale est indiquée par la stérilité observée chez leurs mutants qui est souvent accompagnée par une de-répression et une mauvaise localisation des transposons. TDRD1 est une protéine multidomaine avec un domaine N-terminal des doigts de zinc MYND suivi par quatre domaines tandem eTud. Elle est considérée d’interagir avec des sDMAs du N-terminal des MILI. Le but de cette thèse est d’explorer les aspects structuraux de l’interaction TDRD1-MILI et comment les domaines Tudor reconnaissent des arginines méthylées. Finalement, une caractérisation préliminaire du domaine MYND comme module d’interaction protéine-protéine dans la voie de l’ARNpi est aussi décrite.

1.1 - Small non-coding RNAs

Research over the last decades has seen an increasing importance of RNAs as regulatory molecules. Even if a big part of the human genome is transcribed into RNA, only 2 % encodes proteins. Most transcripts are non-coding RNAs (ncRNAs) that have diverse roles in many cellular processes, including transcriptional regulation, chromosome replication, RNA processing, RNA modification and messenger RNA stability and translation. A certain class of these RNAs has defining features, including their short length (~20-30 nucleotides) and their association with members of the Argonaute family of proteins. Beyond their common properties, these small RNAs exert many diverse schemes of gene regulation, ranging from heterochromatin formation to mRNA destabilization and translational control, by showing sequence complementarity to RNA target molecules.

During the past two decades, the number of small ncRNAs in a variety of organisms has increased substantially. The first small non-coding RNA was discovered in the nematode worm *Caenorhabditis elegans* after genetic screens (Lee et al., 1993; Wightman et al., 1993). This small ncRNA, *lin-4*, was found to have complementary sequence to the 3' untranslated region (3' UTR) of *lin-14* mRNA, suggesting that *lin-4* regulates a messenger RNA in a RNA-RNA dependent manner. Few years later, in 1998, Fire and Mello discovered in *C.elegans* the phenomenon of RNA interference (RNAi) (Fire et al., 1998). In RNAi, a double-stranded RNA (dsRNA) is involved, and may originate from viral infection (exogenously), inverted-repeat transgenes or from transcription of nuclear genes (endogenously). These dsRNAs are processed into single-stranded RNAs (ssRNAs) and can ultimately base pair to the mRNA target, thereby preventing its translation into protein. RNA interference has been found in plants, fungi, and a variety of animals, including the *C. elegans*, *Drosophila melanogaster* zebrafish and mouse (Cogoni and Macino, 1999; Kennerdell and Carthew, 1998; Sanchez Alvarado and Newmark, 1999; Waterhouse et al., 1998). Thus, these findings set the foundation of understanding gene silencing by small RNAs in a variety of organisms.

Small ncRNAs bind to Argonaute proteins and are guided to their target through distinct and complex pathways, as described later. There are at least three types of small RNAs encoded,

classified according to their size, biogenesis mechanism and type of Argonaute protein they are associated with: microRNAs (miRNAs), small interfering RNAs (siRNAs) and PIWI-interacting RNAs (piRNAs). The widely studied miRNAs and siRNAs bind to the AGO subfamily of Argonaute proteins, while the recently discovered and less understood piRNAs bind to the PIWI subfamily of Argonautes.

The scope of this thesis is the description of the structure and function of Argonaute-associated proteins in the piRNA pathway. A brief introduction of the miRNA and siRNA pathways as well as their effector molecules, the Argonaute proteins, is given below.

1.1.1 - The Argonaute family of proteins

The regulation of gene expression by small ncRNAs is mediated by Argonaute proteins, which lead these RNAs to their complementary target. The resultant protein-RNA complex along with other associated proteins form a minimal core of the effector complex termed RNA-Induced-Silence-Complex (RISC) (Hammond et al., 2000). A number of Argonaute proteins bound to the small silencing RNA (known as guide strand) silence target mRNAs through their endonucleolytic activity, a process called Slicing (Hutvagner and Simard, 2008). Structural studies on bacterial, archaeal and recently human Argonaute proteins in complex with RNA and RNA-DNA duplexes, as well as individual domains from metazoan have shed light on the activity and function of both AGO and PIWI subfamily proteins.

Argonautes have a molecular weight of approximately 100 kDa and have four distinct domains: the highly heterogeneous N-terminal domain, the PIWI-Argonaute-Zwille (PAZ) domain, the MID and the PIWI domains ([fig. 1.1.A](#)) (Jinek and Doudna, 2009). Crystallographic studies have revealed that MID and PIWI domains form one lobe and N-terminal and PAZ domains constitute the other. The PAZ domain is an RNA-binding domain and is responsible for binding the 3' ends of single-stranded RNAs, suggesting that it anchors the small RNA involved in the silencing ([fig.1.1.C](#)) (Lingel et al., 2003; Song and Joshua-Tor, 2006; Yan et al., 2003). The N-terminal domain is postulated to be required for small RNA duplex unwinding, without being involved in the later steps of loading and target cleavage (Kwak and Tomari, 2012). The 5' phosphate of the

guide RNA is buried in a pocket between the MID and PIWI domains and is bound to a magnesium ion (fig.1.1.C), a feature not observed in the human Argonaute 2 (Schirle and MacRae, 2012) (fig.1.1.D). The four residues that coordinate the magnesium ion and the 5' phosphate are highly conserved amongst Argonaute proteins and mutations to any of the four amino acids impair the slicing activity of the protein (Ma et al., 2005). The PIWI domain has a similar fold to the bacterial RNase H, which is an endoribonuclease that cleaves RNA-DNA hybrids (Nowotny et al., 2005). PIWI is the catalytically active domain of the protein, with an active site formed by three conserved residues—two aspartates and a histidine—known as the “DDH” motif. These amino acids form the “Slicer” activity of an Argonaute, and mutations in any of the three residues damage the endonucleolytic process *in vitro* and *in vivo*. (Liu et al., 2004; Rivas et al., 2005). Overall, the recently solved crystal structure of the human Argonaute 2 shows a similar fold to prokaryotic Argonaute structures (Schirle and MacRae, 2012). Individual domains superimpose reasonably well, with the eukaryotic structure having additional secondary structure elements and extended loops, suggesting for conserved functions in silencing mechanisms (fig.1.1.D).

Argonaute proteins are the key proteins common in the three small RNA pathways and the interactions between the small RNA and the protein are mediated by the sugar-phosphate backbone; as a result, the bases of the guide RNA are free to base pair with many diverse RNA targets (Parker et al., 2004; Wang et al., 2009a). Many studies focusing on the function and variety of Argonaute proteins have shown that not all Argonaute proteins have Slicer activity. In humans, there are four Argonaute proteins (AGO1, AGO2, AGO3 and AGO4) and only AGO2 is catalytically active (Liu et al., 2004; Meister et al., 2004). AGO1 and AGO4 do not have the “DDH” motif, but AGO3 has a complete Asp-Asp-His motif and yet is inactive *in vitro*. Nevertheless, *D. melanogaster* Piwi protein (belonging to the PIWI subfamily) has a “DDL” motif and proved to have a slicer function (Saito et al., 2006). Taken altogether, additional factors to the catalytic residues contribute to the Slicer activity of an Argonaute *in vivo*.

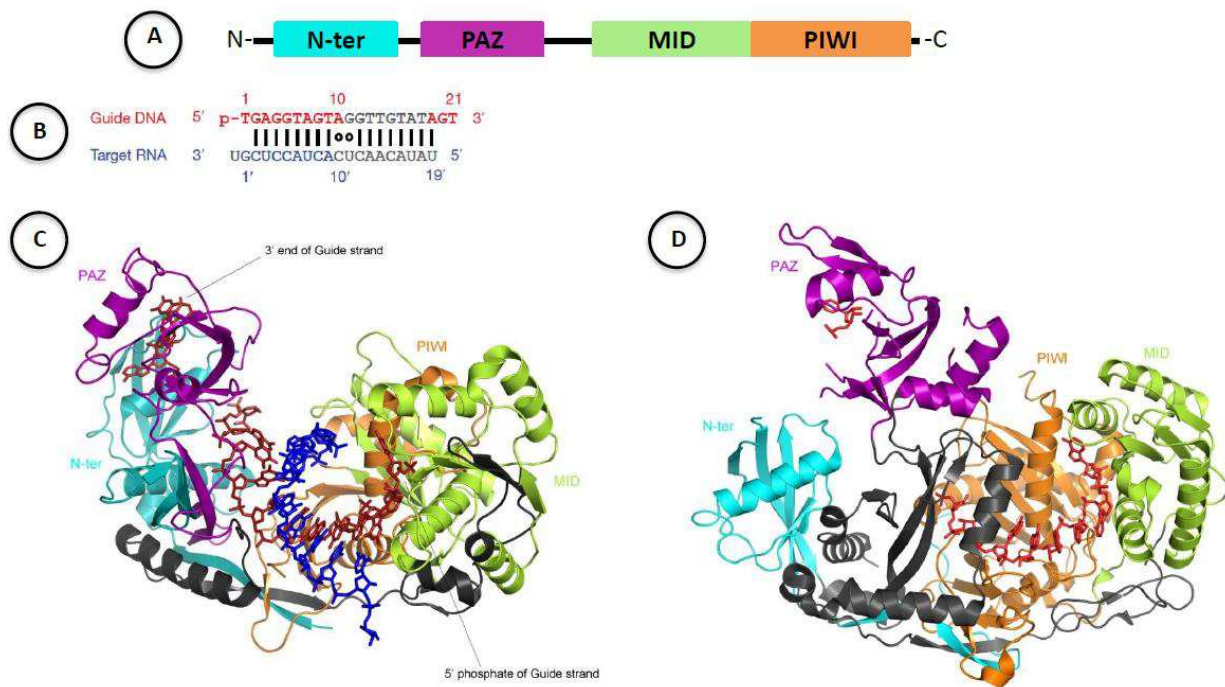


Figure 1.1: Argonaute proteins are the effector molecules of small RNA pathways. **A.** Domain organization of a typical Argonaute protein. **B.** Sequence of the DNA-RNA duplex. Grey nucleotides are disordered in the structure. **C.** Crystal structure of *T. thermophilus* Ago bound to a 5'-phosphorylated 21-nucleotide guide DNA and 20-nucleotide target RNA (**B**) at 3.0 Å resolution (PDB 3F73). The mismatches of the DNA-RNA duplex at positions 10-11 prevent cleavage, thus facilitating the detailed examination of the complex. Protein domains: N-terminal domain (cyan), PAZ domain (purple), MID domain (limon) and PIWI domain (orange). Disordered regions corresponding to linkers are displayed in black. The guide DNA (red) and the target RNA (blue) are shown in a stick-model representation. Bound to the PAZ domain, the 3' end of the guide DNA is shown, while the 5' phosphate is bound to a cleft between the MID and PIWI domains. (reproduced from Wang et al., 2008). **D.** Crystal structure of the human Argonaute 2 at 2.3 Å (PDB 4E13). The bound RNA (red stick model) likely corresponds to cellular RNA (reproduced from Schirle and MacRae, 2012).

1.1.2 - MicroRNAs (miRNAs)

MicroRNAs (miRNAs) are the most abundant species of ncRNAs in mammalian cells and they have fundamental roles in regulating genes in the post-transcriptional level (Bartel, 2009; Carthew and Sontheimer, 2009). These ~22 nucleotide long RNAs are reported to fine-tune almost the half of protein-coding genes in metazoans and plants and are involved in every cellular process known so far. A brief description of their biogenesis, mode of function and biological importance will be given.

1.1.2.1 – Biogenesis

miRNA genes are transcribed in the nucleus by RNA polymerase II (Cai et al., 2004; Lee et al., 2002; Lee et al., 2004a) and the primary transcripts are capped and polyadenylated and called primary miRNAs (pri-miRNAs) (Kim et al., 2005). Pri-miRNAs are long double-stranded RNA molecules of ~80 base pairs (bp), having a stem-loop structure (fig.1.2), which are further processed by the Rnase III endonuclease Drosha and its mammalian dsRNA-binding domain (dsRBD) partner, DGCR8 (Pasha in flies and *C. elegans*) (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Landthaler et al., 2004; Lee et al., 2002; Lee et al., 2004b). The endonucleolytic activity results in the cleavage of the stem of pri-miRNAs and the release of the pre-miRNAs, which have a hairpin-like structure (Lee et al., 2002). The following step is the export of the pre-miRNAs to the cytoplasm. This step is mediated by the nuclear export protein Exportin 5 (EXP5), which binds the miRNA hairpin duplex along with the GTP- bound form of the cofactor Ran, and releases it in the cytoplasm after hydrolysis of GTP (fig.1.2) (Bohnsack et al., 2004; Lund et al., 2004; Yi et al., 2003; Zeng and Cullen, 2004).

Once being in the cytoplasm, pre-miRNAs are cleaved further by another endonuclease, termed Dicer, releasing ~22-nt miRNA duplexes, containing mismatches and having a two-nucleotide overhang at their 3' end left by Drosha (Bernstein et al., 2001; Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001; Knight and Bass, 2001). Dicer is a conserved protein amongst eukaryotes, and can have multiple homologues with distinct functions (Lee et al., 2004b; Xie et al., 2004). It associates with dsRNA-binding proteins, like Drosha; however, in plants, the protein DCL1 and its dsRBD partner HYL 1 fulfill the roles of both Drosha and Dicer, by cleaving pri-miRNAs to miRNA duplexes in one step (Papp et al., 2003; Park et al., 2002; Reinhart et al., 2002; Xie et al., 2004). Human Dicer interacts with TRBP (TAR RNA binding protein) and PACT (Chendrimada et al., 2005; Haase et al., 2005; Lee et al., 2006), whereas Dicer 1 in flies associates with Loquacious (LQQS), which encompasses three dsRBDs (Forstemann et al., 2007; Jiang et al., 2005; Saito et al., 2005).

In the next step, the miRNA duplex is loaded onto an Argonaute protein, which along with Dicer and TRBP form the effector RNA-induced silencing complex (RISC). Once RISC is assembled, one

strand is bound to Ago (guide strand) and the other (passenger strand or miRNA*) is released and degraded (Du and Zamore, 2005; Khvorova et al., 2003; Schwarz et al., 2003). Strand selection is based on thermodynamic stability of the duplex ends, with the strand having a less stably-paired 5' end being selected as a guide (Khvorova et al., 2003; Schwarz et al., 2003). In those cases where the thermodynamic stability is the same in the two ends, both species get loaded with approximately even frequencies. Ultimately, RISC loaded with the mature miRNA can bind target mRNAs and cause either translational repression or endonucleolytic cleavage (fig.1.2).

1.1.2.2 - Function

In the miRNA pathway, the final step is dictated by both the type of Argonaute protein and the extent of complementarity between the miRNA and the target mRNA (Haase et al., 2010; Hutvagner and Zamore, 2002; Liu et al., 2004; Meister et al., 2004). In the case of full complementarity to the target, the result is the endonucleolytic cleavage of the mRNA by the Argonaute protein (fig.1.2) (Davis et al., 2005; Mansfield et al., 2004; Pfeffer et al., 2004; Sullivan et al., 2005; Yekta et al., 2004). For target cleavage, Watson-Crick pairing is required between the guide miRNA and the target mRNA at two regions: the seed region, which is located 2-7 nt as counted from the 5' end of the guide strand and the cleavage region, 10-11 nt, also from the 5' end of the miRNA (fig.1.1.B, fig.1.1.C) (Meister et al., 2004; Tomari and Zamore, 2005). The small size of the seed explains how a wide variety of gene targets can be regulated by miRNAs (Baek et al., 2008; Selbach et al., 2008). In flies and mammals few miRNAs are reported to lead to cleavage of their target, whereas in plants this is the universal mechanism, with only few exceptions (Brodersen et al., 2008; Llave et al., 2002; Rhoades et al., 2002; Tang et al., 2003).

The imperfect complementarity, on the other hand, leads to repression of translation or promotion of mRNA deadenylation and decay (fig.1.2). miRNAs within the RISC complex bind imperfectly to the 3' untranslated Region (UTR) of mRNAs, usually in multiple copies. They can pair with their target only within the seed region, which mediates the primary recognition of the target strand. (Brennecke et al., 2005; Lai et al., 2004; Lewis et al., 2005; Lewis et al., 2003;

Stark et al., 2005; Xie et al., 2005). miRNAs and their targets are usually mismatched at miRNA positions 10-12, preventing endonucleolytic cleavage, and the pairing is reinforced by additional matches beyond the seed region (Grimson et al., 2007; Lewis et al., 2005).

Mechanistic details of the translational repression of target mRNAs are subject to ongoing debate. Models involving repression at the initiation and post-initiation steps of translation have been suggested over the years, involving either decapping of the 5' cap, deadenylation of the 3' poly-(A) tail, and ribosome drop-off of the messenger RNA (Brodersen et al., 2008; Fabian et al., 2010; Lytle et al., 2007; Petersen et al., 2006; Pillai, 2005). The function of mRNA deadenylation, however, is better understood and established (Chekulaeva and Filipowicz, 2009; Eulalio et al., 2009; Fabian et al., 2010). This molecular mechanism involves the key glycine-tryptophan protein of 182 kDa (GW182), which interacts with Argonaute proteins and acts downstream of them (fig.1.2). The N-terminal part of GW182 (GW repeats) interacts with Ago and most likely with its PIWI domain, which contain tandem tryptophan binding pockets (Schirle and MacRae, 2012). The C-terminal part of GW182 associates with the poly(A)-binding protein (PABP) and recruits the factors CCR4 and CAF1 that are responsible for target mRNA deadenylation, as well as the decapping enzymes DCP1 and DCP2, resulting in mRNA decay (Behm-Ansmant et al., 2006; Eulalio et al., 2009; Fabian et al., 2010; Rehwinkel et al., 2005).

Products of the translational repression machinery of messenger RNAs are found in discrete cytoplasmic loci, termed processing bodies (P-bodies). P-bodies are enriched in mature miRNAs, repressed mRNAs and various proteins, including Argonaute and GW182 proteins and factors involved in repression, deadenylation, decapping and degradation (Eulalio et al., 2007; Franks and Lykke-Andersen, 2008). These cytoplasmic granules have a dynamic structure and serve in both storage and decay of repressed mRNAs.

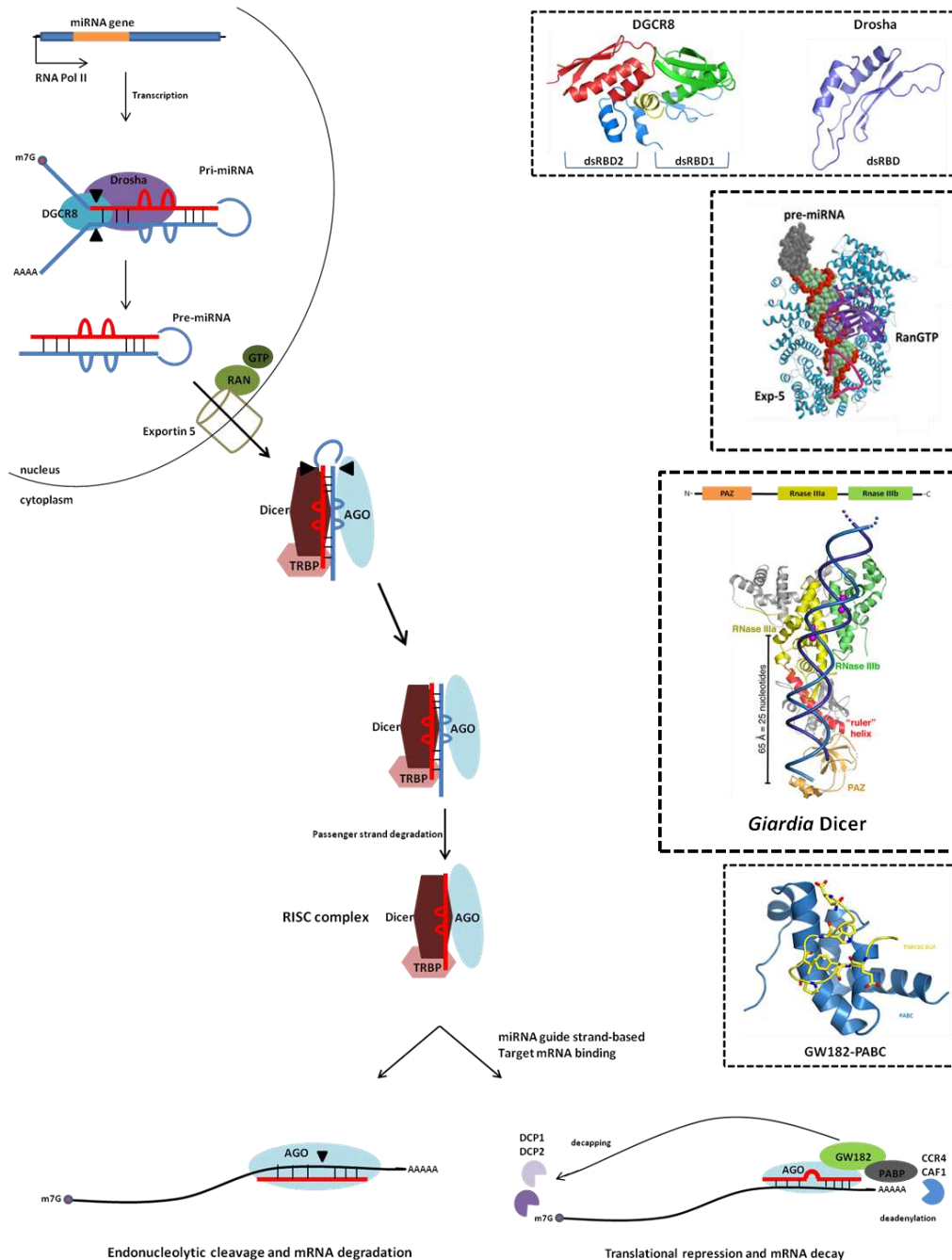


Figure 1.2: Micro-RNA biogenesis and function. Depicted are the steps of the miRNA biogenesis, loading of the mature miRNA to Argonaute, formation of the RISC complex and mRNA-targeted functions. In dashed boxes, known structures of the miRNA-involved proteins are shown: Crystal structure of the human DGCR8 core, comprised of two dsRBDs (PDB 2YT4) (Sohn et al., 2007). Solution structure of the Drosha double-stranded RNA-binding domain (PDB 2KHX) (Mueller et al., 2010). 2.9 Å structure of pre-miRNA nuclear export machinery (pre-miRNA, Exp-5, and guanine triphosphate (GTP) bound form of Ran (PDB 3A6P) (Okada et al., 2009). Crystal structure of the full-length *Giardia* Dicer at 3.3 Å, modelled with dsRNA (PDB 2FFL). Human Dicer includes additionally a helicase domain, a domain of unknown function and a dsRBD domain (Macrae et al., 2006). Crystal structure of a conserved motif in the human GW182 paralog TNRC6C (yellow) in complex with the C-terminal domain of PABC (blue) (PDB 2X04) (Jinek et al., 2010).

1.1.2.3 - Biological importance

microRNAs are key regulators of diverse biological processes in plants and animals and are involved in almost all cellular pathways under physiological or stress conditions. The importance of miRNA-mediated regulation is reflected by the fact that mutant animals for miRNA components are either non-viable or show severe developmental defects, while the dysregulation of the pathway leads to a variety of diseases, among which cancer and neurodegeneration. (Bartel, 2004; Stefani and Slack, 2008; Yekta et al., 2008).

The function of miRNAs can be found in specific tissues at specific developmental time windows (Landgraf et al., 2007). The research in the RNA silencing field has answered many questions concerning pluripotency, tumorigenesis, apoptosis and cell identity while new therapeutic agents are synthesized for regulating gene expression, using the knowledge coming from small RNAs.

1.1.3 – siRNAs

Small-interfering RNAs are the effector molecules of RNAi and are widely distributed in both phylogenetic and physiological terms. siRNAs are ~21-nt long and act not only as regulators of genes, like miRNAs, but also as gatekeepers of genome integrity in response to foreign nucleic acids such as viruses. A summary of their biogenesis, function and biological significance is following.

1.1.3.1 Biogenesis

siRNAs arise from long double-stranded RNA molecules that are cleaved into small RNA duplexes. The cleavage is performed by the ribonuclease Dicer (Bernstein et al., 2001) and the duplexes are bearing a 5' phosphate and a 3' hydroxyl group and they show perfect complementarity except for 2 nt overhangs at their 3' end ([fig.1.3](#)) (Elbashir et al., 2001a; Elbashir et al., 2001b; Zamore et al., 2000). Mammals and *C.elegans* have a single Dicer responsible for both miRNA and siRNA pathways (Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001; Knight and Bass, 2001), whereas *Drosophila* species have two Dicers: DCR-1 responsible for miRNA cleavage and DCR-2 for siRNA duplexes production (Lee et al., 2004b),

with this distinction probably reflecting the source of the ncRNAs entering the cell. Once siRNA duplexes are formed, the Argonaute protein is recruited, which cleaves the passenger strand. (Kim et al., 2006; Leuschner et al., 2006; Matranga et al., 2005; Miyoshi et al., 2005; Rand et al., 2005). The thermodynamic stabilities of the 5' ends of the duplex determine which strand is the guide, a mechanism that is sensed by the dsRNA-binding protein R2D2, the partner of DCR-2 in *Drosophila melanogaster* (Aza-Blanc et al., 2003; Khvorova et al., 2003; Liu et al., 2003; Schwarz et al., 2003; Tomari et al., 2004). The Argonaute protein with its bound guide RNA and the R2D2-DCR-2 complex constitute the mature RISC. In flies and plants, the siRNA molecule bound to Ago is 2'-O-methylated at its 3' end by the methyltransferase HEN1, completing RISC assembly (fig.1.3) (Horwich et al., 2007; Pelisson et al., 2007).

siRNAs that trigger RNAi mechanism inside the cell can derive either from an endogenous source (endo-siRNAs) or from exogenous agents (exo-siRNAs). Endo-siRNAs usually derive from centromeres, transposons transcripts, sense-antisense transcript pairs and long-stem-loop structures (fig.1.3) (Babiarz et al., 2008; Chung et al., 2008; Czech et al., 2008; Ghildiyal et al., 2008; Okamura et al., 2008) and are also present in mouse oocytes (Tam et al., 2008; Watanabe et al., 2008). The endo-siRNA pathway in worms and plants is more complex, since they are using RNA-dependent RNA polymerases (RdRPs) for the generation of endo-siRNAs (fig.1.3) (Pak and Fire, 2007; Sijen et al., 2007). RdRPs are enzymes that transcribe ssRNAs from an RNA template, giving rise to dsRNAs.

Exo-siRNAs that induce the siRNA pathway usually come from ectopically introduced long dsRNAs, inverted-repeat transgenes or siRNA duplexes (fig.1.3) (Liu et al., 2004). In *Drosophila melanogaster*, RNAi directs a defend mechanism against viruses that produce dsRNA during infection (Wang and Ligoxygakis, 2006), while mammals might not use RNAi because they have developed a protein-based immune system to respond to virus attack (Kunzi and Pitha, 2005). Plants present a diversity of small RNA types and proteins that are involved in silencing mechanisms and this might reflect the need of an elaborate system to respond to environmental conditions like biotic and abiotic stress. Exogenous sources of siRNAs are single stranded and are converted to dsRNA by the RdRP enzyme RDR6 (Qu et al., 2005).

1.1.3.2 - Function

siRNAs like miRNAs act mostly in a post-transcriptional mechanism. The siRNA guide strand directs RISC to the mRNA target and through base complementarity promotes its degradation (fig.1.3.B). The cleavage is catalyzed by the PIWI domain of the Argonaute protein. This slicer activity of the Ago is very precise and cleaves in a position 10 and 11, counting from the 5' end, generating products with a 5'-monophosphate and a 3'-hydroxyl termini (Tomari and Zamore, 2005). The 3' end of the cleaved target is a substrate for oligouridylation, which recruits exonucleases to complete the degradative process (Shen and Goodman, 2004). Once the target mRNA is degraded, the RISC complex is recycled and free to cleave additional targets.

In some cases, there is not perfect base pairing and mismatches between siRNA and target suppress cleavage; in these cases, imperfect complementarity leads to translational repression or degradation of the target, in a similar manner to miRNA pathway (fig.1.3.A) (Tomari and Zamore, 2005).

siRNAs are not restricted to post-transcriptional repression of the target mRNA; in *S. pombe*, it was shown that they are able to induce heterochromatin formation, and this mechanism is also reported in animals and plants (fig.1.3.C) (Chan, 2008; Moazed, 2009). In fission yeast, the siRNAs lead the RNA-induced transcriptional silencing complex (RITS) to centromeric repeats. siRNAs are reported to base pair with nascent transcripts facilitated by direct interaction between RITS and RNA polymerase II (Djupedal et al., 2005). The mechanism involves specific methylation of lysine 9 on histone H3 (H3K9) by histone methyltransferases, which is promoted by RITS. This in turn leads to the recruitment of the chromodomain protein Swi6, which promotes chromatin compaction (Lippman and Martienssen, 2004). Transcriptional silencing is enhanced by the action of the RNA-dependent RNA polymerase complex (RDRC), which generates secondary siRNAs (Sugiyama et al., 2005).

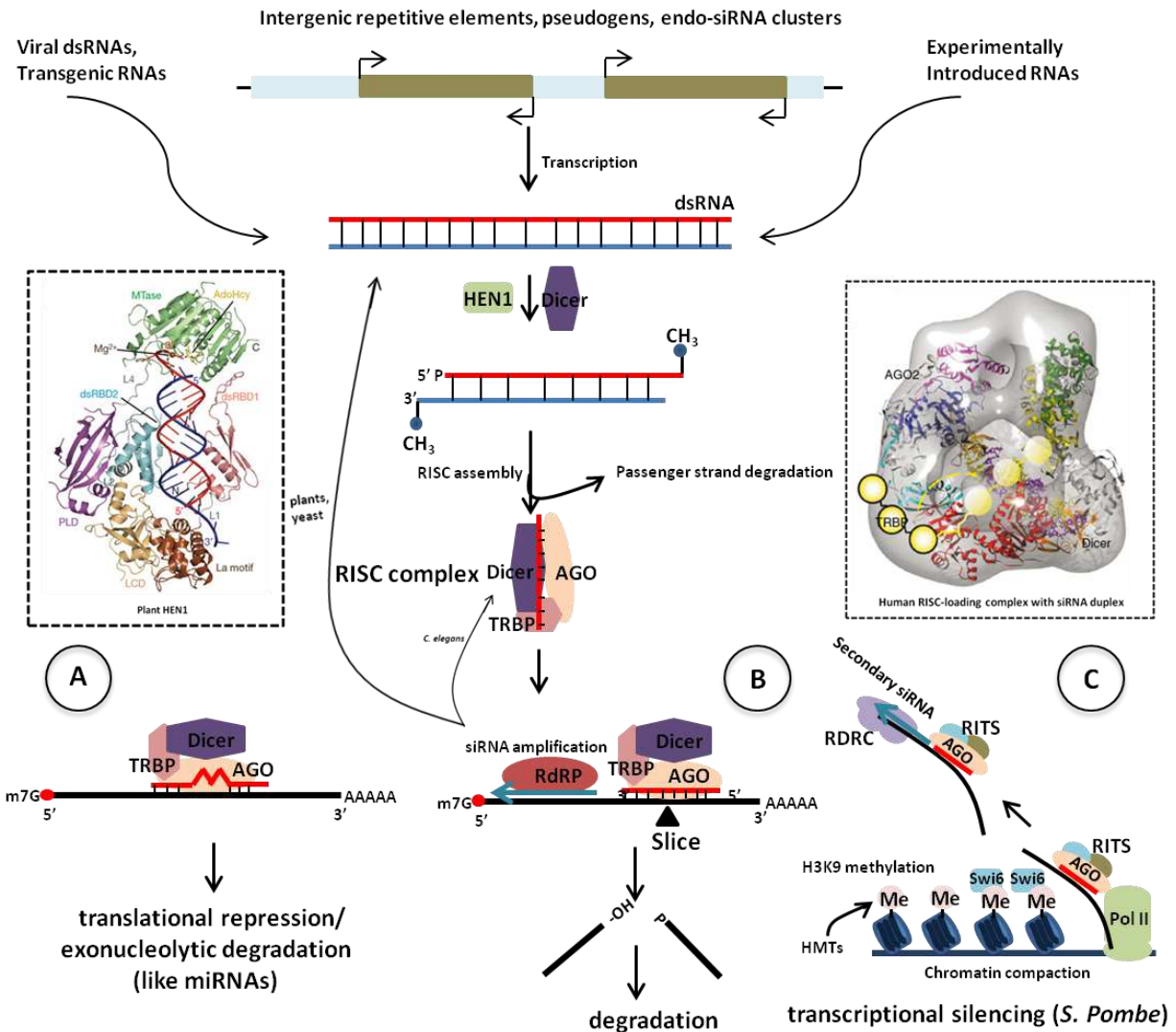


Figure 1.3: siRNA biogenesis and function. Depicted are the sources of dsRNAs that trigger the commencement of the siRNA pathway, along with the loading of the siRNA to the RISC complex and the diverse gene silencing functions: **A.** translational repression, **B.** degradation, **C.** transcriptional silencing. Secondary siRNAs are produced by RdRPs, which are loaded directly to RISC in *C.elegans* or further processed by Dicer in plants and yeast. In dashed boxes: Left: 3.1 Å structure of full-length HEN1 methyltransferase from *Arabidopsis* in complex with a 22-nt small RNA duplex (PDB 3HTX) (Huang et al., 2009). Right: EM structure of the human RISC-loading complex (RLC). The flexible TRBP is modeled to translocate upon transfer of the newly diced siRNA duplexes onto AGO2's PAZ domain (orange), in order to facilitate their transfer and correct strand selection (Wang et al., 2009a).

1.1.3.3 - Biological importance

siRNAs through the mechanism of RNAi have become an everyday experimental tool in laboratory research. The discovery of their biogenesis and function has made siRNAs excellent candidates for possible therapeutic applications and challenges. The ability to knock-down any

gene of interest has given rise to RNAi trials in clinical research for various diseases. However, there are many challenges, especially in mammals, since the effectiveness in different cell types can vary. Finally, a large number of RNAi screens have become very useful in terms of discovering new functions of genes in many biological pathways inside the cell.

1.1 - Piwi-interacting RNAs (piRNAs)

1.2.1 - General features

piRNAs are the most recently discovered class of small ncRNAs and they bind to the PIWI subclade of Argonaute proteins. They are clearly distinct of the other small RNA classes because of their longer size (24-30 nt) and their expression pattern, since they are found in the germ line of animals. Furthermore, in the piRNA pathway, it was shown that Dicer enzyme is not required for piRNAs' production, while they are 2'-O-methylated at their 3' end termini, unlike miRNAs but similar to siRNAs in flies (Horwich et al., 2007; Kirino and Mourelatos, 2007; Ohara et al., 2007; Pelisson et al., 2007; Saito et al., 2006; Saito et al., 2007; Vagin et al., 2006). They were firstly discovered in *Drosophila*, where it was shown that they are associated with silencing of repetitive elements and were firstly named "repeat associated small interfering RNAs", or rasiRNAs (Aravin et al., 2003; Aravin et al., 2001). Piwi-interacting RNAs are widely found throughout the animal kingdom, in human, fish, mouse, rat, worm and even in simpler animals, like sea anemone (*Nematostella vectensis*) and sponges (*Amphimedon queenslandica*), but not in plants or fungi (Grimson et al., 2008).

The PIWI clade of Argonaute proteins is comprised of Piwi, Aubergine (Aub) and AGO3 in flies and Mili, Miwi and Miwi2 in mice (Aravin et al., 2006; Cox et al., 1998; Cox et al., 2000; Girard et al., 2006; Grivna et al., 2006; Nishida et al., 2007; Saito et al., 2006). PIWI proteins associate with piRNAs and function in silencing transposons and repetitive genomic sequences in a transcriptional and post-transcriptional level. Transposons are mobile genetic elements that are able to multiply within the genome by transposition to new sites (Slotkin and Martienssen, 2007). As a result, their movement can cause deleterious effect to the organisms, leading to genomic structural rearrangements, such as deletions, duplications and inversions. Thus,

organisms have developed an effective defense mechanism via the piRNA pathway against these genome intruders and as a result a faithful transmission of genetic material from one generation to the other is maintained. The functional importance of this pathway was revealed with genetic studies, showing that PIWI proteins mutants lead to fertility defects in mouse, fish and flies (Carmell et al., 2007; Deng and Lin, 2002; Houwing et al., 2008; Houwing et al., 2007; Kuramochi-Miyagawa et al., 2004; Lai et al., 2004; Li et al., 2009; Lin and Spradling, 1997; Schmidt et al., 1999).

1.2.2 - Origin and expression pattern

In *Drosophila melanogaster*, PIWI proteins show different sub-cellular localization and expression patterns, suggesting that they have distinct roles. The protein Piwi is the only one found in the nucleus (Cox et al., 2000) where it was found to silence Transposable Elements (TEs) (Saito et al., 2010). In ovaries, Piwi is detected both in somatic and germ cells (Brennecke et al., 2007; Cox et al., 2000; Nishida et al., 2007; Saito et al., 2006), while in testes it is found mainly in somatic niche cells, which associate with germline stem cells. Aub and AGO3 are exclusively found in the cytoplasm in the germ cells of both males and females, but they are absent from somatic cells (fig.1.4).

piRNAs in flies originate from TEs and genomic repetitive elements (Brennecke et al., 2007; Gunawardane et al., 2007; Saito et al., 2006; Yin and Lin, 2007) and arise from long single stranded precursors that are processed via an unknown mechanism (Vagin et al., 2006). Furthermore, their context is highly variable and is more heterogeneous than miRNAs. Mapping of piRNAs to the *Drosophila* genome revealed that they come from particular loci, termed piRNA clusters. These clusters include transposons sequences and transposons remnants and are mainly located in pericentromeric and subtelomeric areas of heterochromatic regions. They can be divided into two groups: i) dual strand-clusters, producing piRNAs from both strands, named sense and antisense and ii) uni-strand clusters, that are transcribed only in one direction, giving rise to piRNAs that are antisense to their target transcripts. The latter are represented by the *flamenco* (*flam*) locus on the X chromosome, which is a master regulator of the expression of *gypsy*, *idefix* and ZAM elements that form retroviral particles (fig.1.4) (Desset

et al., 2008; Pelisson et al., 1994). The *flam* piRNA locus is found only in the ovarian somatic cells.

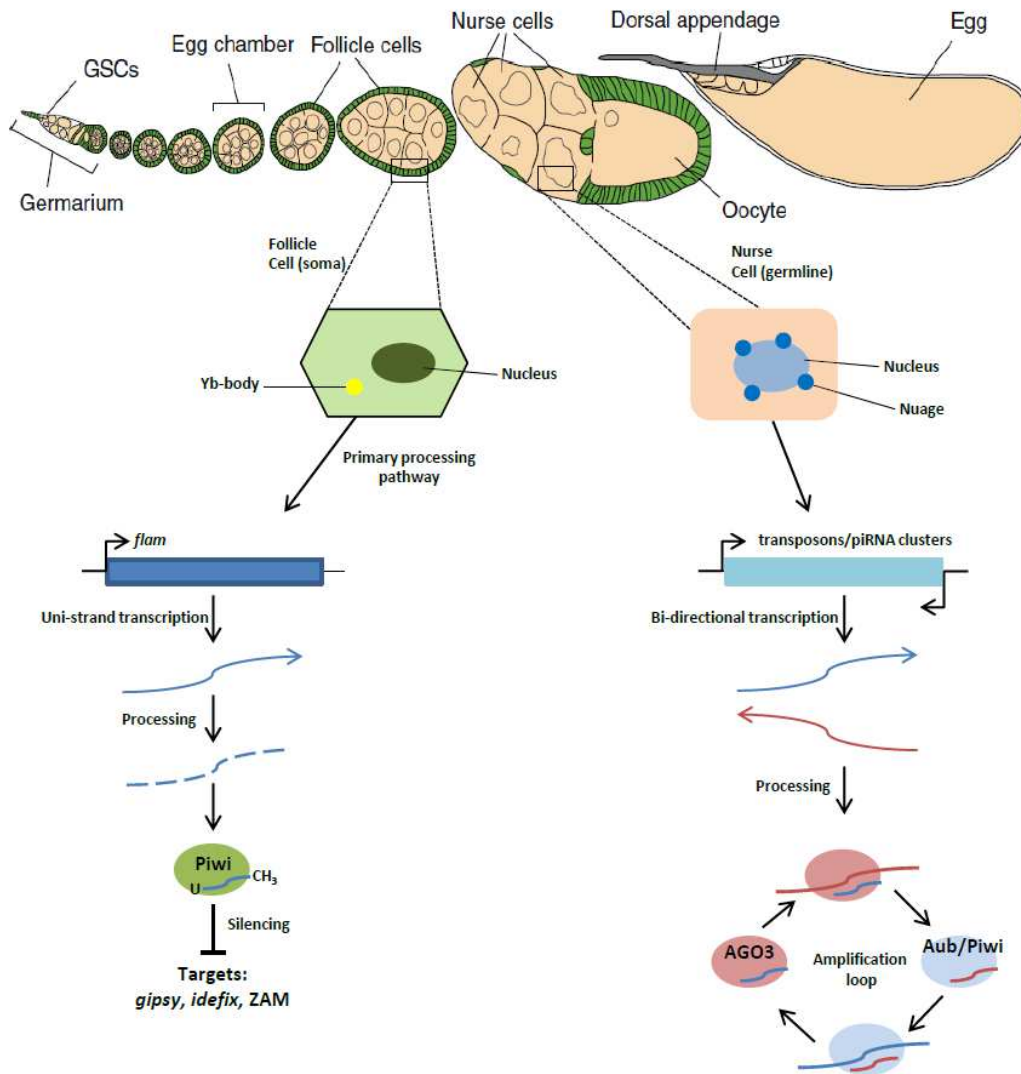


Figure 1.4: piRNA pathway and function in flies. Top: *Drosophila melanogaster's* oogenesis stages (adapted from Senti and Brennecke, 2010). In somatic cells, where the Piwi protein is involved, the primary processing of piRNAs takes place and involves the silencing of transposable elements, like *gipsy*, *idexif* and *zam*.. In germline cells, AUB and/or Piwi with AGO3 participate in the amplification loop/ping-pong cycle where they cleave transposon transcripts using their slicer activity.

In mouse, the three PIWI proteins (Mili, Miwi and Miwi2) are expressed during spermatogenesis where in complex with piRNAs function in transposon silencing. Loss Mili and Miwi2 leads to activation of TEs and spermatogenic arrest leading to sterility, while Miwi knock-

out mice are blocked in spermatogenesis but show no transposon derepression (Aravin et al., 2007b; Carmell et al., 2007; Deng and Lin, 2002; Kabsch, 1993; Kuramochi-Miyagawa et al., 2004; Kuramochi-Miyagawa et al., 2001). Murine PIWI proteins have distinct and overlapping expression patterns throughout spermatogenesis (fig.1.5). Mili is detected in both females and males, whereas Miwi and Miwi2 are found exclusively in male germ cells (Kabsch, 1993; Kuramochi-Miyagawa et al., 2001). Expression of Mili starts from primordial germ cells at embryonic day (E) 12.5 and reaches until the round spermatid stage, being present throughout adult spermatogenesis (Kuramochi-Miyagawa et al., 2004). Miwi2 is present in the developmental window between embryonic gonocytes at E15.5 until 3 days after birth, prior to meiosis (Aravin et al., 2008; Kuramochi-Miyagawa et al., 2008). Miwi's expression is restricted to meiotic and postmeiotic spermatogenesis, reaching the round spermatid stage (Deng and Lin, 2002; Kuramochi-Miyagawa et al., 2001). During germ cell differentiation, epigenetic regulation takes place, which includes global DNA demethylation in primordial germ cells (PGCs), where transposons sequences can become highly active (fig.1.5). The erasure of methylation marks is followed by a *de novo* methylation establishment in male (in spermatogonia/gonocytes) and female germ cells (Mochizuki and Matsui, 2010). As described later, piRNA pathway first mediates transposon silencing at the *de novo* methylation stage.

piRNAs, during spermatogenesis, can be divided into two classes: i) pre-pachytene piRNAs, that are expressed before meiotic pachytene, and ii) pachytene piRNAs, that are present in the pachytene stage (fig.1.5). (Aravin et al., 2007b; Girard et al., 2006; Grivna et al., 2006; Lau et al., 2006; Watanabe et al., 2008). Pre-pachytene piRNAs are derived from repeat and transposons rich clusters (Aravin et al., 2006), similar to *Drosophila* piRNAs and are mostly transcribed in both directions (Aravin et al., 2008). This class of piRNAs interact with Mili and Miwi2 (Aravin et al., 2008; Aravin et al., 2007b; Kuramochi-Miyagawa et al., 2008) and is involved in transposons silencing. During postnatal spermatogenesis at the onset of meiosis, Mili is the only PIWI protein expressed, binding to transposons-derived piRNAs as well (Aravin et al., 2007b). Pachytene piRNAs, on the other hand, derive mainly from non-TE, non-annotated sequences and they bind to Mili and Miwi (Aravin et al., 2008; Aravin et al., 2007b; Girard et al., 2006).

Intergenic unannotated sequences are up to 200 kilobases long and their role remains elusive (fig.1.7).

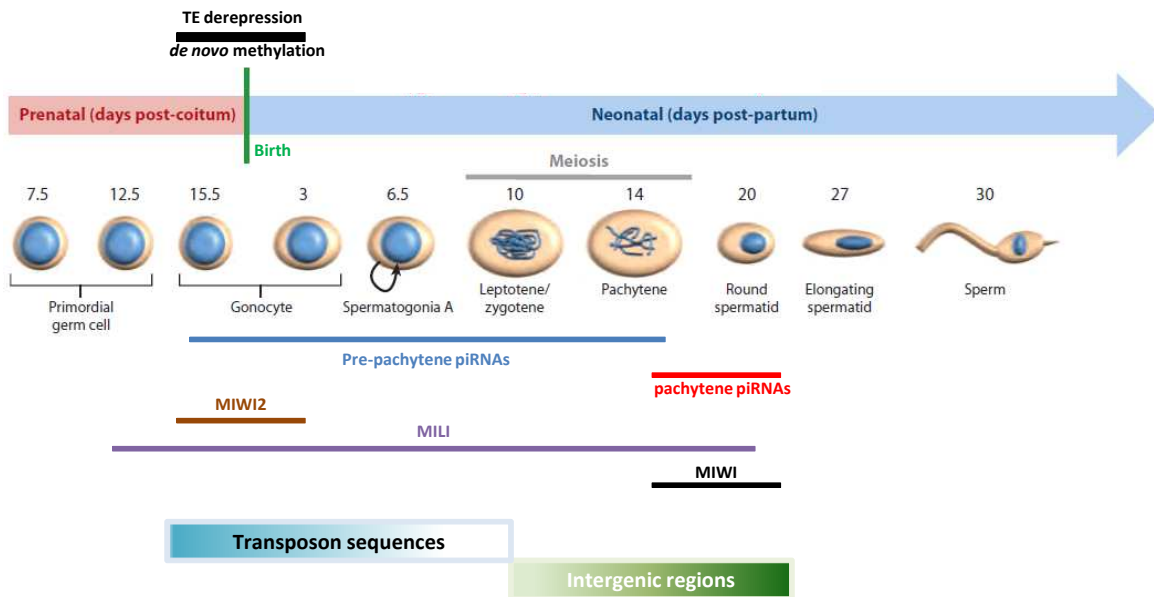


Figure 1.5: Spermatogenesis in mouse and expression profile of PIWI proteins. Depicted are the different stages of spermatogenesis in mouse (modified from Thomson et al., 2009). Transposon de-repression and their *de novo* methylation via the piRNA pathway takes place just before and after birth. The developmental windows of pre-pachytene and pachytene piRNAs are shown, along with the three murine PIWI proteins. piRNAs at the beginning of spermatogenesis until meiosis are originating from transposable sequences, that are reduced when approaching meiosis (dark blue: rich in TEs, light blue: poor in TEs). At meiosis and in the pachytene stage, piRNAs are arising from intergenic regions and become abundant (light green: poor in intergenic regions, dark green: rich in intergenic regions). piRNAs at this stage and their mode of action are not yet understood.

1.2.3 - Biogenesis

1.2.3.1 - Primary pathway

The linear transmission of the precursor transcript to complexes with PIWI proteins is called primary piRNA biogenesis (Brennecke et al., 2007) and the mechanistic details of this pathway remain unknown. In flies, primary biogenesis is believed to be the only processing mechanism in somatic support cells and the Argonaute protein Piwi is involved. Studies in ovarian somatic cell line (OSC) (Lau et al., 2009; Niki et al., 2006), which contains only somatic cells of fly ovaries strongly suggest that single-strand long transcripts are produced from piRNA clusters and are

substrates for Piwi protein, which in turn silence TE transcripts. The majority of these piRNAs is derived from the *flamenco* locus and is suggested that *flam*-piRNAs are exclusively produced through the primary pathway. piRNAs also originate from mRNAs, preferentially from the 3' UTRs (Robine et al., 2009). An example of this is the *Drosophila* gene *traffic jam*, which is translated and produces piRNAs from its 3' UTR. The long transcripts are trimmed to mature piRNAs via an unknown mechanism, since Dicer, the processing enzyme in miRNA and siRNA pathways is absent in piRNAs. The location of the primary biogenesis is still debated, even though accumulating evidence suggests that it takes place in the cytoplasm of somatic and germline cells (Saito et al., 2009).

Primary biogenesis is also found in mouse spermatogenesis and is believed to be the only processing mechanism in pachytene piRNAs. However, we still lack information about how it functions. In both organisms, it is believed that the long cluster transcript is a substrate for an unknown endonuclease, that gives rise to smaller piRNA, termed pre-piRNAs (fig.1.7). It seems likely that pre-piRNAs are then selectively loaded to PIWI proteins, which via their MID domain recognize the 5' end of the piRNAs. Sequencing analyses has shown that piRNAs show a strong bias for uridine (U) at their 5' end, but differ at their 3' ends (Ghildiyal and Zamore, 2009; Ma et al., 2005; Ruby et al., 2006). Interestingly, the crystal structure of the human Ago2 MID domain showed that it preferentially accommodates 1U at the 5' end via a rigid loop that provides base-specific recognition (Frank et al., 2010) and suggests for a similar mechanism for the PIWI argonautes. PIWI proteins are unlikely to use their slicer activity for primary biogenesis, as it was shown not to be required (Saito et al., 2009). At a final step, the 3' end of the pre-piRNA is trimmed by an unknown exonuclease and then methylated by the RNA methyltransferase Hen1, which protects piRNAs from degradation (fig.1.7) (Horwich et al., 2007; Kamminga et al., 2010; Kawaoka et al., 2011; Kirino and Mourelatos, 2007; Saito et al., 2007). In addition, mature piRNAs vary in size. For instance, Mili binds piRNAs ~26 nt long, while Miwi binds longer RNAs (~30 nt) and in most cases, the differences resides in the 3' end (Aravin et al., 2006). Thus, the footprint of the PIWI binding the 5' end of the pre-piRNAs is determining piRNAs' length, and this explains the different size of piRNAs bound to different PIWI proteins (Brennecke et al., 2007).

1.2.3.2 - Secondary biogenesis (ping-pong model)

The secondary pathway was initially described in *Drosophila*, where different PIWI proteins were found to bind transcripts from piRNA clusters that have different orientation (Brennecke et al., 2007; Gunawardane et al., 2007; Saito et al., 2006; Yin and Lin, 2007). In particular, in *Drosophila* ovaries, the proteins Piwi and Aub were found to have a strong preference for uridine (1-U) at their 5' end and their bound piRNAs are mainly antisense to active TE transcripts. The AGO3 protein, on the other hand, was binding piRNAs coming from the sense transcripts of TEs and showed a strong bias for an adenosine (A) at the tenth nucleotide from their 5' end (10-A). Interestingly, the first ten nucleotides of the antisense piRNAs overlap with the sense piRNAs. This observation is referred to as the ping-pong model, and suggests that Aub (and Piwi)-associated piRNAs base pair with AGO3-associated piRNAs through their first 10 nucleotides (fig.1.6.A) (Brennecke et al., 2007; Gunawardane et al., 2007). This pathway makes use of the slicer activity of the PIWI proteins, since Aub associated with an antisense transcript, cleaves sense transposon transcripts, generating the 5' end of a secondary piRNA. An unknown nuclease would then trim the 3' end to the size of the mature piRNA that in turn binds to AGO3. Following that, AGO3 can target an antisense transcript and through a slicer-mediated cleavage generate the same initial Aub-associated antisense piRNA (fig.1.6.A). The endonucleolytic activity of the three PIWI proteins was shown *in vitro*, where they cleave complementary target RNAs between nucleotides 10 and 11 of their guide-bound piRNAs (Gunawardane et al., 2007), and this probably explains why piRNAs are generated in a Dicer-independent manner (Vagin et al., 2006). Thus, this mechanism has a dual role in the piRNA pathway; firstly, it degrades transposon mRNAs and secondly generates new piRNAs through an amplification loop (Saito et al., 2010).

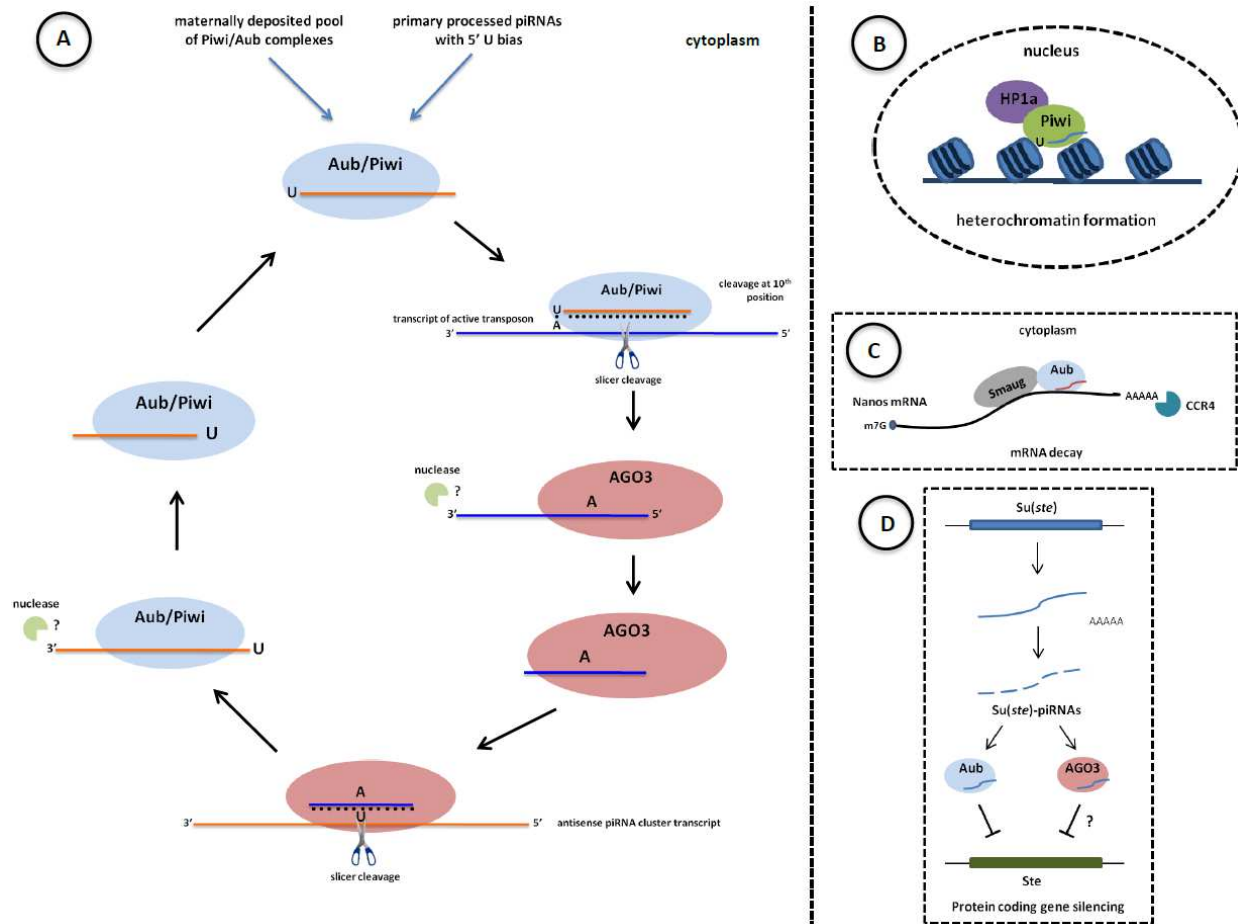


Figure 1.6: Function of the piRNA pathway in *Drosophila melanogaster*. **A.** Ping-pong amplification. piRNA molecules entering the amplification cycle can derive either maternally or from the primary processing pathway. **B.** Piwi protein is reported to have a role in heterochromatin formation along with HP1a. **C.** Another proposed mechanism is the mRNA decay of Nanos, after its recognition by the Aub-piRNA complex. **D.** *Ste* gene is silenced by PIWI-piRNA complexes (adapted from Siomi et al., 2011).

In mice, the piRNA amplification loop is probably also present and is detected at the pre-pachytene stage of spermatogenesis (Aravin et al., 2008; Aravin et al., 2007b), while pachytene piRNAs binding to Mili and Miwi have an unknown function (fig.1.7). The ping-pong cycle is directed by Mili and Miwi2 where Mili associates with sense 10-A piRNAs (generated by the primary processing mechanism), cleaves primary piRNAs through its slicer activity, generating secondary piRNAs that enter Miwi2 (Aravin et al., 2008). Miwi2 loaded with antisense to transposons transcripts piRNAs is then transported to the nucleus where it promotes DNA methylation of transposons elements (fig.1.7.A) (Aravin and Bourc'his, 2008; Aravin et al., 2008). Genetic evidence has proven that mice mutants lacking Mili lead to Miwi2 being

unloaded with piRNAs and mislocalized in the cytoplasm (Aravin et al., 2008). Nevertheless, there are evidence that the transmission of piRNAs from Mili to Miwi2 can be linear, since in mouse *Miwi2*^{DAH} mutant (point mutant in the catalytic DDH motif), the spermatogenesis had no defects, suggesting that Miwi2 may not be a slicer (De Fazio et al., 2011). In addition, in a *Miwi2* knock-out mouse, Mili is normally loaded with piRNAs (Kuramochi-Miyagawa et al., 2010). Finally, it is suggested that at a certain period of spermatogenesis there is potentially a piRNA amplification loop involving only Mili, since it was found to have a tendency for binding 1-U and 10-A piRNAs (Aravin et al., 2007b).

The ping-pong cycle in flies is restricted to the germ cells, where all three proteins are involved with Piwi to a less extent. There is supporting evidence that primary processing can be coupled with the amplification pathway, where primary produced piRNAs can become substrates for the ping-pong cycle (fig.1.6.A) (Li et al., 2009). These two pathways are occurring in many vertebrates, including frogs and zebrafish apart from mouse (Aravin et al., 2007b; Ghildiyal and Zamore, 2009; Houwing et al., 2007; Malone and Hannon, 2009).

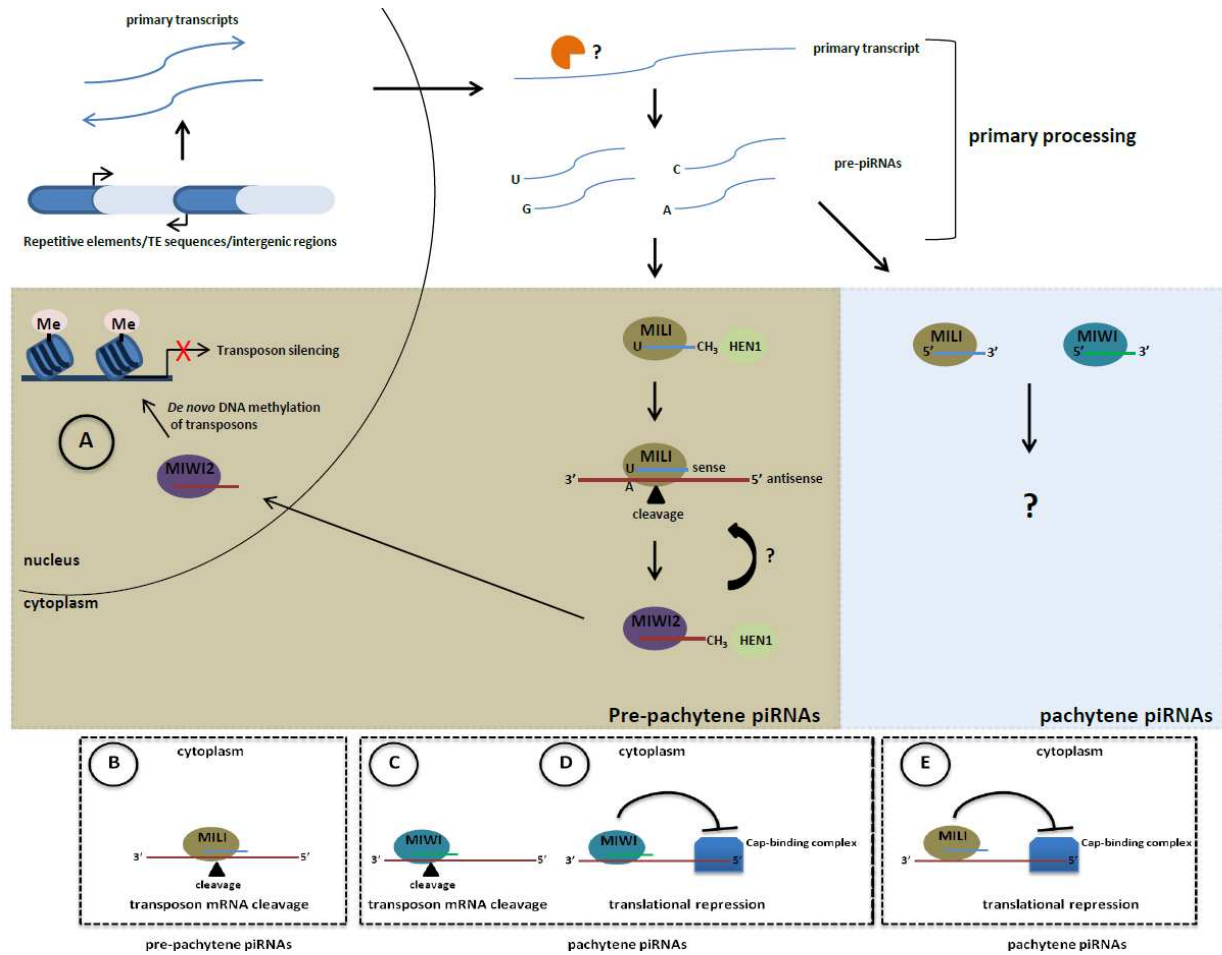


Figure 1.7: piRNA biogenesis and function in mouse. piRNA clusters (transposon sequences, non-annotated regions, repetitive elements) give rise to primary transcripts that are processed via the primary processing pathway in the cytoplasm. In pachytene piRNAs, Mili and Miwi with their bound piRNAs are involved, while in pre-pachytene piRNAs, Mili and Miwi2 participate in a ping-pong or linear transmission of sense and anti-sense piRNAs. **A.** Miwi2 is finally transported to the nucleus where it is involved in transcriptional silencing of transposons. **B.** Mili, through its slicer activity, is able to cleave transposons mRNAs. **C, D, E.** Proposed functions in the pachytene stage of spermatogenesis. Miwi can cleave transposons transcripts via its slicer activity, while along with Mili are also proposed to be involved in translational repression (modified from Pillai and Chuma, 2012)).

1.2.4 - Regulation by PIWI-interacting proteins

1.2.4.1 - Tudor domain proteins

PIWI proteins are the catalytic molecules in the piRNA pathway where they can cleave complementary target loci using their Slicer activity and generate secondary piRNA molecules. However, genetic studies in *Drosophila* and mouse have shown that several other proteins are

required in order the piRNA pathway to be efficient. This means that PIWI proteins form large effector complexes, similar to RISC complexes in miRNA and siRNA pathways (Ghildiyal and Zamore, 2009). Immunoprecipitation experiments using a specific antibody (Y12) recognizing symmetrically dimethylated arginines, identified PIWI proteins with their bound piRNAs in mouse testes and *Xenopus laevis* oocytes (Kirino et al., 2009a). Additionally, proteomic studies on Argonaute proteins revealed that the PIWI clade is post-translationally modified at the N-terminus at specific arginines residues, found in arginine-glycine (RG) or arginine-alanine (RA) repeats. These arginines are symmetrically dimethylated (symmetrically dimethylated arginines-sDMAs) or monomethylated (MMAs) (Chen et al., 2009; Kirino et al., 2009a; Nishida et al., 2009; Vagin et al., 2009). This modification is exclusively found on PIWI proteins, suggesting that arginine modification is a molecular signature of the piRNA pathway. Biochemical studies in flies and mice on proteins co-purifying with PIWIs showed that they associate with the protein arginine N-methyltransferase 5 (PRMT5) in complex with the methylosome protein 50 (MEP50) (Kirino et al., 2009a; Nishida et al., 2009; Vagin et al., 2009). PRMT5 is known to symmetrically dimethylate arginine residues in many animal species and is believed to be the responsible methyltransferase for the arginine methylation on PIWI proteins (fig.1.14.B) (Chen et al., 2009; Kirino et al., 2009a; Nishida et al., 2009; Reuter et al., 2009; Vagin et al., 2009). This modification on PIWI proteins is important for their binding capacity to piRNAs, since loss of *Drosophila* PRMT5 leads to unloaded Aubergine (Nishida et al., 2009). Furthermore, proteomic studies on PIWI-complexes revealed that proteins containing Tudor domains were also present. Tudor domains are well-studied protein-protein interaction domains known to bind sDMAs (fig.1.14.A) (Chen et al., 2009; Reuter et al., 2009; Vagin et al., 2009). Various proteins with a different number of Tudor domains were found in complexes with all three mouse and *Drosophila* PIWIs, having different expression patterns during germ cell development (fig.1.8) (Chen et al., 2009; Nishida et al., 2009; Reuter et al., 2009; Shoji et al., 2009; Vasileva et al., 2009). The association between Tudor proteins and the sDMAs of PIWIs was confirmed in a heterologous cell culture system as well, where the presence of dimethylated arginines was indispensable for this interaction (Vagin et al., 2009). Thus, PRMT5 and Tudor domains act synergistically as regulators in the piRNA pathway.

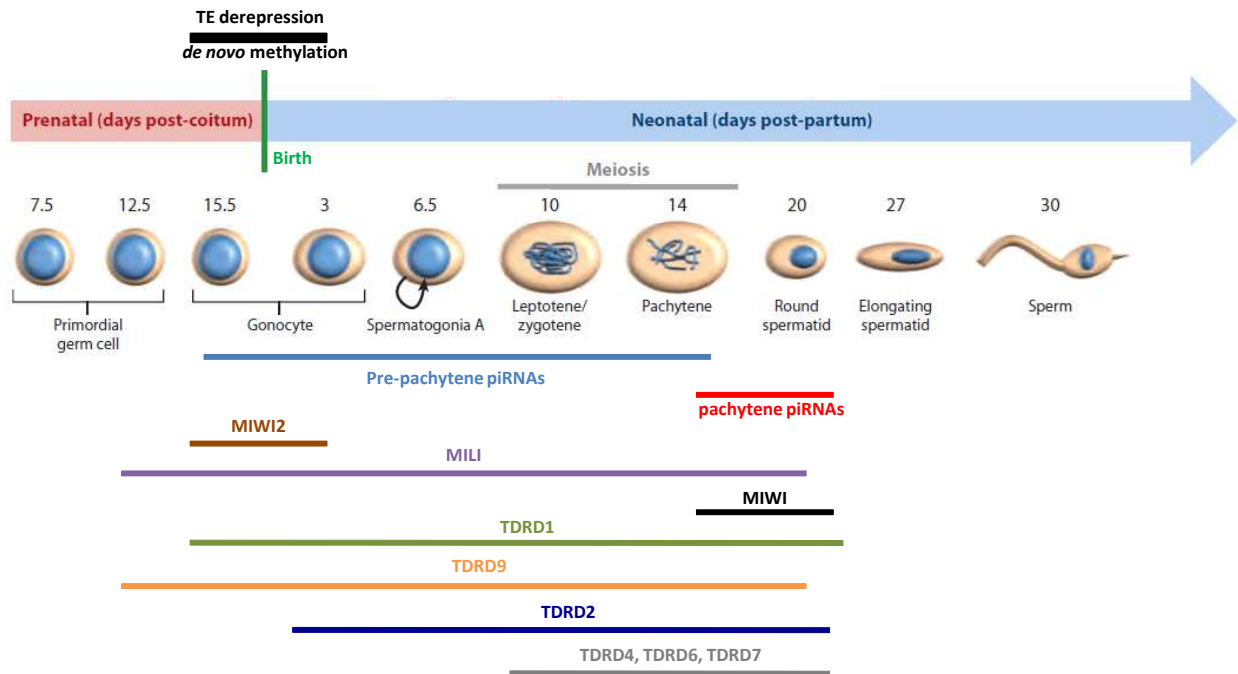


Figure 1.8: Expression pattern of TDRD proteins during spermatogenesis. Tudor domain-containing proteins (TDRDs) have non-redundant roles in the piRNA pathway and this is depicted by their expression profile. While some Tudor domain proteins have overlapping expression patterns, some are expressed only at restricted developmental windows.

In the fly piRNA pathway, there have been reported more than twelve proteins having Tudor domains and their exact role is still a mystery. The number of Tudor domains can widely vary from one to eleven, and most of the Tudor-containing proteins are essential, since they are implicated in transposon silencing and normal progression of germ cell differentiation and are summarized in [fig.1.9](#). Some of these proteins affect piRNA population and distinct PIWI's localization, and others are involved either in the primary processing or in the amplification cycle. For instance, the protein Tudor (TUD) encompasses eleven Tudor domains and is necessary for gametogenesis, germ cell formation and female fertility (Arkov et al., 2006; Boswell and Mahowald, 1985; Thomson and Lasko, 2004). TUD associates with Aub and AGO3 in an sDMA-dependent manner, and this interaction promotes piRNA-PIWI association (Nishida et al., 2009) and proper localization of PIWI proteins in certain perinuclear organelles (Kirino et al., 2009b; Nishida et al., 2009). In addition, information about the methylarginine dependence of Tudor domains is also missing, since there are examples of Tudor domain-

proteins interacting with PIWI proteins in a sDMA-independent manner (Patil and Kai, 2010). Thus, proteins with Tudor domains are important for the piRNA pathway, but their direct interaction with PIWI proteins and their sDMA-Tudor interaction are still undetermined.

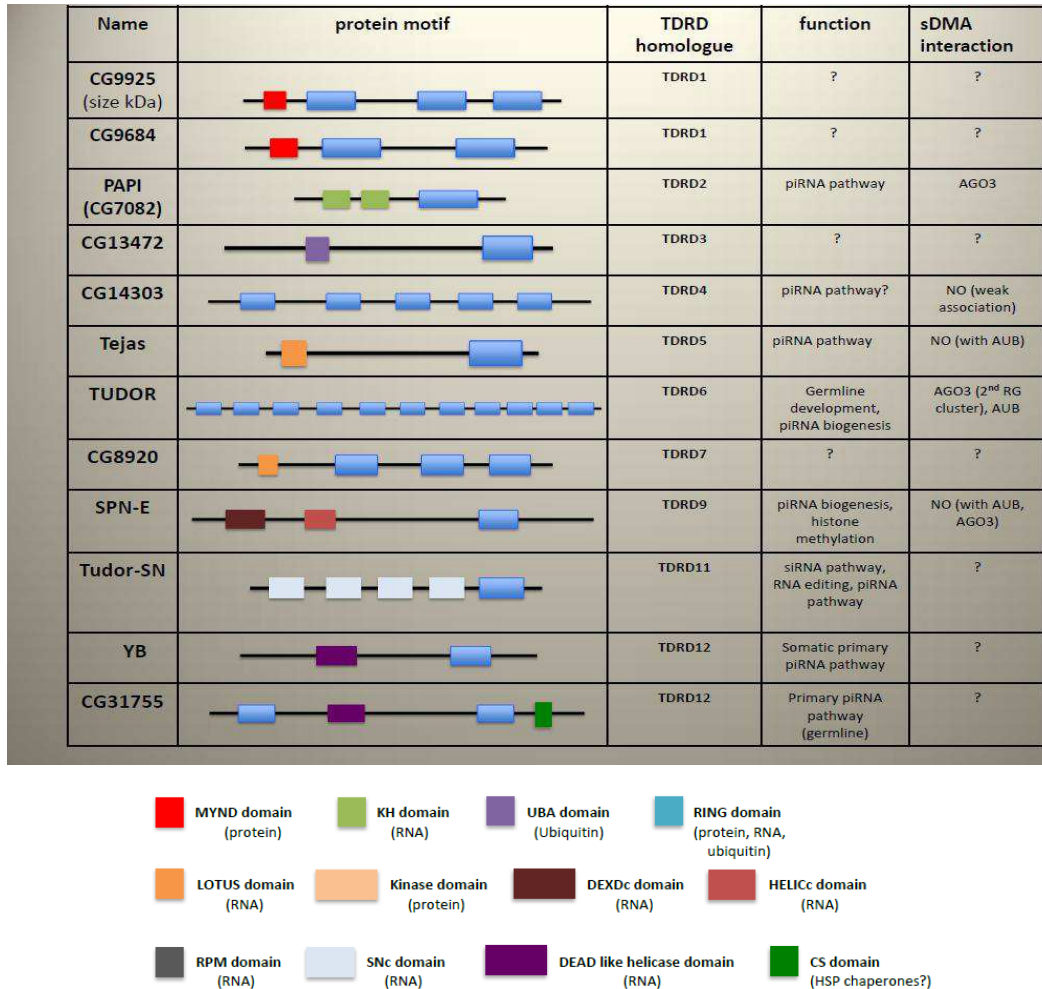


Figure 1.9: Tudor domain-containing proteins in *Drosophila*, involved in the piRNA pathway. This table summarizes the proteins encompassing Tudor domains found in *Drosophila melanogaster*. Their domain organization, homologues from the mouse TDRD family, function, and dimethylated arginine dependence is depicted. Question marks denote for unknown or not-reported. In flies, more than 20 proteins implicated in the piRNA pathway are found; however, for reason of comparison, only the TDRD homologues are depicted. Tudor domains are colored in blue. The remaining protein domains are also listed. In parentheses the potential interactor for each domain is shown.

In mouse, the first protein found to interact with the PIWI protein Mili in an RNA-independent manner is Tudor domain-containing protein 1 (TDRD1), a germline-specific protein containing four Tudor domains (Chuma et al., 2003; Chuma et al., 2006; Kojima et al., 2009; Reuter et al.,

2009; Vagin et al., 2009; Wang et al., 2009b). TDRD1, which is the protein studied during this thesis, belongs to the sub-family of Tudor domain containing proteins (TDRDs, described later) and has the same expression pattern as Mili, interacting with its N-terminus via dimethylated arginines (fig.1.8) (Kojima et al., 2009; Reuter et al., 2009; Vagin et al., 2009; Wang et al., 2009b). Loss-of function mutations on *Tdrd1* exhibit male-specific sterility and early spermatogenic arrest, a phenotype similar to that of Mili and Miwi2 mutants (Carmell et al., 2007; Chuma et al., 2006; Kuramochi-Miyagawa et al., 2004). More specifically, *Tdrd1* mutants show defects in meiosis and de-repression of the LINE1 class of retrotransposons, a consequence of failure in *de novo* methylation of TEs (Reuter et al., 2009; Shoji et al., 2009). *Tdrd1* mutants also affect piRNA population and localization of PIWI proteins; *Tdrd1*-deficient mice demonstrate that although piRNAs are still present, the generation of secondary piRNAs through the ping-pong cycle is impaired (Vagin et al., 2009) while another study showed that in *Tdrd1* mutant mice, Mili accepts small RNAs originating from ribosomal RNAs and protein-coding RNAs through an unknown mechanism (Reuter et al., 2009). Finally, the functional implication of TDRD1 in the amplification loop is strengthened by the observation that when TDRD1 is absent, Miwi2 stays in the cytoplasm, devoid of piRNAs (Aravin and Bourc'his, 2008; Reuter et al., 2009). Thus, TDRD1 has a central role in the amplification of piRNAs as well as the repression of TEs.

In the murine piRNA pathway, there are also other functionally important TDRDs with their mechanistic role still being elusive. There are twelve TDRD proteins identified so far and the number of their Tudor domains can vary, similar to *Drosophila* (fig.1.10). During spermatogenesis, the expression pattern of TDRD proteins is wide and either overlapping or distinct, ranging from fetal prospermatogonia, primary spermatocytes to round spermatids and spermatogonia (fig.1.8). Another TDRD member, TDRD9, interacts with MIWI2 (Shoji et al., 2009), suggesting for non-redundant but specific TDRD-PIWI interactions. Information whether TDRD9-MIWI2 interaction is sDMA-dependent is still missing. *Tdrd9* mutant mice exhibit the same phenotype as the *tdrd1* ones, indicative for their essential roles in the epigenetic regulation by piRNAs. In addition, mice lacking TDRD4, TDRD6 and TDRD7 exhibit spermatogenic arrest but post-meiotically, after the formation of round spermatids, resembling

the phenotype of Miwi-knockout mice (Deng and Lin, 2002). Recent studies identified the proteins TDRD5 and TDRD7 being involved in the piRNA pathway (Hosokawa et al., 2007; Tanaka et al., 2011; Yabuta et al., 2011). *Tdrd5* mutants show both meiotic and post-meiotic germ cell arrest, overexpression of retrotransposons and mislocalisation of MIWI2. Information about the piRNA profile is still missing in this mutant. Mice lacking TDRD7 show de-repressed LINE1 transposons but had no defects in the piRNA pathway, PIWI's localization or DNA methylation. Taken altogether, TDRD proteins in the murine piRNA pathway have probably distinct and non-redundant functions during spermatogenesis and as mentioned later, distinct sub-cellular localization (Hosokawa et al., 2007).

Tudor domain-containing proteins in *Drosophila* and the TDRD family in mouse contain other functional domains and their role is also unknown (fig.1.9, 1.10). For Instance, the fly protein PAPI has one Tudor domain and two KH domains, known to be involved in RNA binding. Furthermore, *Tdrd1* has a N-terminal MYND domain (Myeloid-nervy-DEAF1) that is reported to be involved in protein-protein interactions, while TDRD9 carries a helicase domain. Concomitantly, these proteins might have additional functional roles in the piRNA pathway. The presence of multiple methylarginines in PIWI proteins and the vast number of Tudor domains suggest that Tudor domains may serve as a molecular platform, by binding different methylated arginines from different PIWI proteins and thus enhancing the piRNA amplification and contributing to the assembly of ribonucleoprotein complexes. A more detailed description of the scaffold theory will be given in the Results and Discussion chapters.

Name (size)	protein motif	PIWI interaction	function	Localization	sDMA interaction
TDRD1 (103 kDa)		MILI, MIWI, MIWI2?	piRNA pathway, spermatogenesis	Nuage (pi-), CB	MILI, MIWI
TDRD2 (TDRKH) (62 kDa)		MILI, MIWI, MIWI2	Implicated in spermatogenesis	Nuage, CB	MIWI (1 st N-terminal Rgclyster)
TDRD3 (72 kDa)		?	Transcription regulation	?	Histone tails
TDRD4 (RNF17) (185 kDa)		MIWI	Nuage formation, spermatogenesis	Nuage, CB	?
TDRD5 (105 kDa)		?	piRNA pathway, nuage formation, spermatogenesis	Nuage, CB	?
TDRD6 (237 kDa)		MIWI, MILI	Nuage formation, spermatogenesis	Nuage, CB	?
TDRD7 (125 kDa)		MIWI	Spermatogenesis, eye development	Nuage, CB	?
TDRD8 (STK31) (115 kDa)		MIWI	Implicated in spermatogenesis	?	?
TDRD9 (156 kDa)		MIWI, MILI(?), MIWI2	piRNA pathway (biogenesis), spermatogenesis	Nuage (piP), CB, nucleus	?, MIWI(?)
TDRD10 (27 kDa)		?	?	?	?
TDRD11 (SND1) (102 kDa)		MIWI	siRNA pathway, RNA editing, piRNA pathway	?(CB?)	MIWI
TDRD12 (137 kDa)		?	ES cell function	?	?

Figure 1.10: TDRD protein family. This table summarizes the TDRD proteins found in mouse and implicated in the piRNA pathway. Their domain organization, interaction with PIWI proteins, function, localization and dimethylated arginine dependence is depicted. Question marks denote for unknown or not-reported. Except for TDRD3, all the TDRD proteins were found to be involved in the piRNA pathway. Tudor domains are colored in blue and are all predicted to have an extended Tudor domain fold (see section 1.3.2.1). The remaining protein domains are listed in Figure 1.9. Please note the smaller size of the Tudor domain of TDRD3, denoting that it is comprised only of the prototypic Tudor domain (~60 amino acids) (see section 1.3.2.1), as its recent crystal structure revealed (PDB 3S6W) (Liu et al., 2012).

1.2.4.2 - Additional PIWI-associated proteins

The biochemical purification of PIWI complexes identified additional to Tudor domains factors implicated in the piRNA pathway. Zucchini, a member of the phospholipase D family is involved in the primary pathway of piRNAs in flies, while in mouse affects the overall levels of piRNAs (Choi et al., 2006; Haase et al., 2010; Huang et al., 2011; Olivieri et al., 2010; Watanabe et al., 2011). Zucchini is suspected to be a nuclease in the biogenesis of piRNAs, while another putative nuclease, Squash (SQU), is thought to be involved in the silencing step (Haase et al., 2010). Other factors affecting piRNA pathway include Maelstrom (MAEL) and GASZ. Maelstrom is believed to affect piRNA biogenesis both in flies and mice (Aravin et al., 2009; Lim and Kai,

2007; Soper et al., 2008), while GASZ is important for Mili's stability and is implicated in the biogenesis of pachytene piRNAs (Ma et al., 2009).

There are also a number of helicases reported to participate in the piRNA pathway, with their exact role not determined yet. MOV10L1 and its fly orthologue Armitage (Armi) are required for primary piRNA biogenesis (fig.1.11) (Frost et al., 2010; Haase et al., 2010; Olivieri et al., 2010; Saito et al., 2010; Zheng et al., 2010). Both proteins are believed to be responsible for loading primary piRNA precursors to PIWI proteins. However, the role of their helicase domains is unknown. The Vasa helicase in *Drosophila* and its mouse orthologue Mouse Vasa Homologue (MVH/DDX4) are reported helicases and when absent, secondary piRNA biogenesis is eliminated (Kuramochi-Miyagawa et al., 2010; Malone and Hannon, 2009) but their mechanistic role is still missing. Interestingly, the crystal structure of Vasa, which is also involved in *Drosophila melanogaster's* development (Hay et al., 1988; Lasko and Ashburner, 1988) revealed that is implicated in RNA duplex unwinding (Sengoku et al., 2006). An interesting feature of both proteins is that they contain symmetrical and asymmetrical dimethylated arginines (aDMAs) at their N-terminus (Kirino et al., 2010), thus being implicated in the molecular platform theory along with PIWIs and Tudor domain proteins (fig.1.11, fig.1.12).

1.2.5 - Sub-cellular localization of piRNA pathway components

PIWI proteins and their associated factors are located both in the cytoplasm and the nucleoplasm. Their localization is present in distinct cytoplasmic structures in the germline cells of many species. Piwi protein, as noted before, is found in the nucleus; however, increasing evidence suggest that the biogenesis and loading of piRNAs occurs in cytoplasmic granules (Aravin et al., 2008; Aravin et al., 2009; Olivieri et al., 2010; Saito et al., 2009; Saito et al., 2010). These granules are found in close proximity to the nuclear membrane, in order to recognize, receive and process piRNA molecules both in flies and mice (Aravin et al., 2009; Kojima et al., 2009; Kotaja et al., 2006; Ma et al., 2009; Nagao et al., 2010; Nishida et al., 2009; Nishida et al., 2007; Patil and Kai, 2010; Saito et al., 2010; Shoji et al., 2009; Toyooka et al., 2000).

In *Drosophila*, PIWI proteins and their loaded piRNAs are located in electron-dense, perinuclear structures, termed nuages (Fukuda et al., 1975). Nuage is found in the nurse cells of the developing germline in the ovaries, and many of its components including PIWIs are localized at the posterior end of the oocyte (fig.1.4) (Anne and Mechler, 2005; Kirino et al., 2009b; Nishida et al., 2007). Inside the nuage, there are many proteins and particular RNAs indispensable for germline development that give rise to the pole plasm, the structure responsible for the formation of primordial germ cells in the developing embryo (Eddy, 1974).

piRNA – associated factors that are present in the nuage include among others SPN-E, Vasa, TUD and MAEL, and mutations in these proteins can alter piRNA profile or cause mislocalization of the rest of the proteins (Lim and Kai, 2007; Patil and Kai, 2010). It is proposed that these cytoplasmic granules are the sites for piRNA production and recognition of target transcripts (Lim and Kai, 2007).

In the somatic follicle cells in flies, where Piwi protein is present and Aub and AGO3 are absent, piRNA pathway is associated with granules called Yb-bodies (fig.1.4) (Olivieri et al., 2010; Qi et al., 2011; Saito et al., 2010). In these adjacent to the nucleus bodies, primary piRNA biogenesis is likely to take place and many factors implicated in the primary processing pathway, like Armi, Zucchini are located there, while factors involved in the amplification loop (Vasa, TUD, SPN-E) are missing. Another factor, protein FS(1)YB is probably the core component of YB's, since when it is missing, Piwi is unloaded with piRNAs and absent from nucleus.

In mice, spermatogenesis consists of many developmental stages, however, PIWI proteins are found in certain granules similar to nuage, termed either 'intermitochondrial cement' or 'chromatoid bodies', depending on the particular spermatogenesis stage (Aravin et al., 2008; Aravin et al., 2009; Kotaja et al., 2006; Kuramochi-Miyagawa et al., 2001; Ma et al., 2009; Vasileva et al., 2009). More specifically, Mili and Miwi2 are found in distinct granules; Mili's granules contain components exclusively found in the piRNA pathway and are named 'pi-bodies' (al-Mukhtar and Webb, 1971) and were originally described as the intermitochondrial cement. pi-bodies contain additionally TDRD1, MVH and GASZ (fig.1.11) (al-Mukhtar and Webb, 1971; Kuramochi-Miyagawa et al., 2010; Reuter et al., 2009). MIWI2's granules, on the other

hand, contain the components of the processing bodies (P bodies, see section 1.1.2.2) and thus are termed 'piP-bodies' (fig.1.11). Furthermore, piP bodies contain the protein TDRD9 and MAEL. These cytoplasmic granules are found in close proximity, and functional cross-talk between them possibly involves exchange of piRNAs through the ping-pong cycle (fig.1.11). Disruption of pi-bodies de-condenses piP-bodies and prevents piRNA loading to MIWI2, which remains cytoplasmic, while there is de-repression of transposons and loss of *de novo* methylation. Thus, both bodies have indispensable roles for the proper function of the piRNA pathway while the molecular processes in pi-bodies are essential for piP-body formation (Aravin et al., 2009).

After birth, where Miwi2's expression stops and Miwi appears during meiosis, the formation of the chromatoid bodies (CBs) takes place (fig.1.12). In these granules, where 'P bodies' components are also present, Mili, Miwi, MVH, TDRD1 as well as other TDRD proteins are located there (Kojima et al., 2009; Kotaja et al., 2006; Vasileva et al., 2009). Pachytene piRNAs in complex with Mili and Miwi probably function in the CBs, even though their exact role and function awaits further investigation. Overall, cytoplasmic granules both in flies and mice appear to be the sites for piRNA processing, amplification and targeting. The presence of multiple Tudor-domain proteins and PIWI proteins suggest that there is a protein interaction network through the recognition of methylated arginines (Rme2s), important for the stability and the quality control of piRNAs (fig.1.11, fig.1.12).

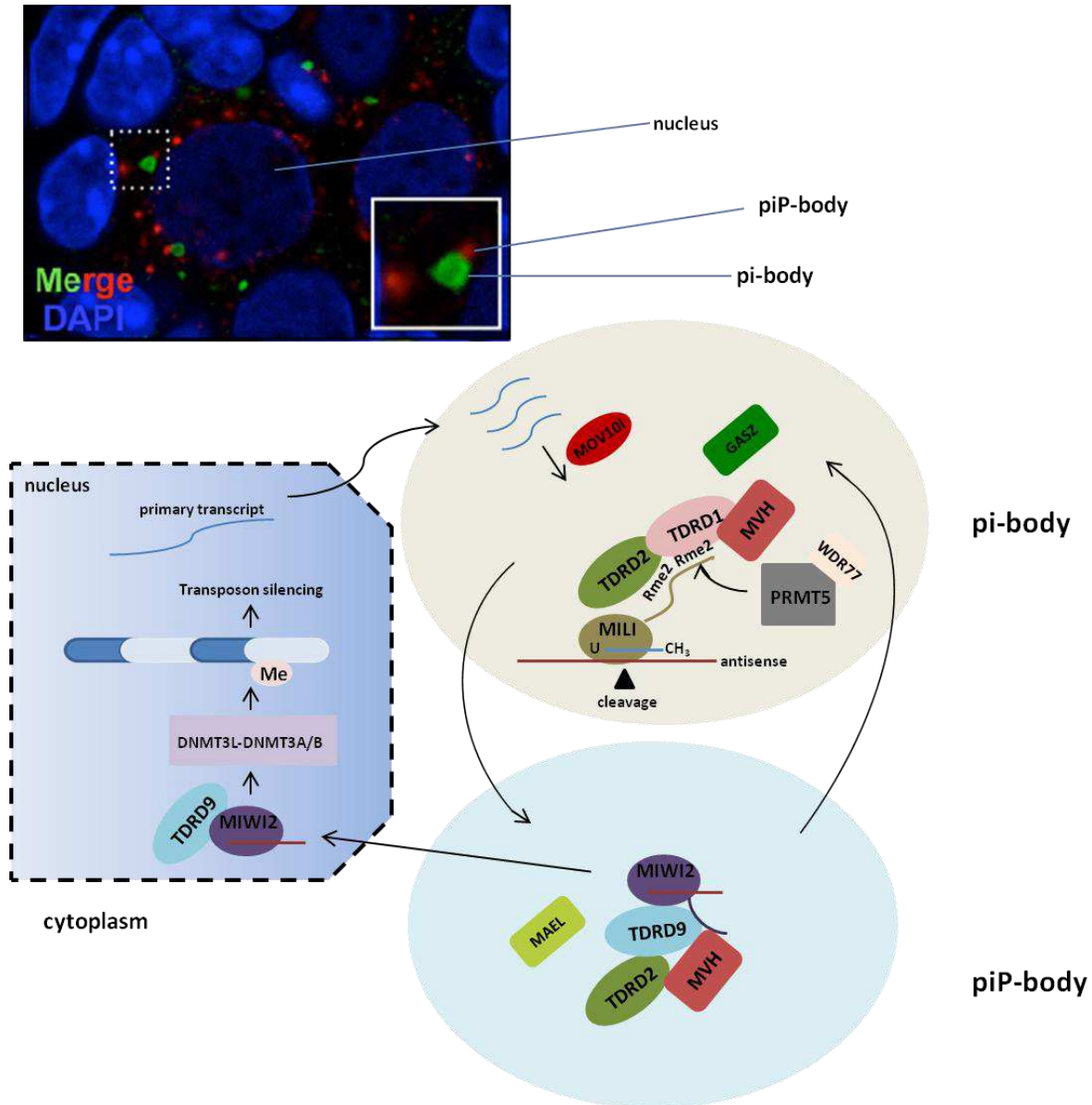


Figure 1.11: Compartmentalization of piRNA pathway-associated factors during the pre-pachytene stage of spermatogenesis. Depicted are the two bodies, where different PIWI-associated factors are localized. Localization experiments in prenatal E 18.5 gonocytes for TDRD1 (green) and MAEL (red) clearly show that pi-body and piP-body are distinct and in close proximity (Aravin et al., 2009). Black arrows indicate that an interchange of piRNA-associated factors takes place between the two bodies. The methylation status of MIW12's N-terminus is not yet determined.

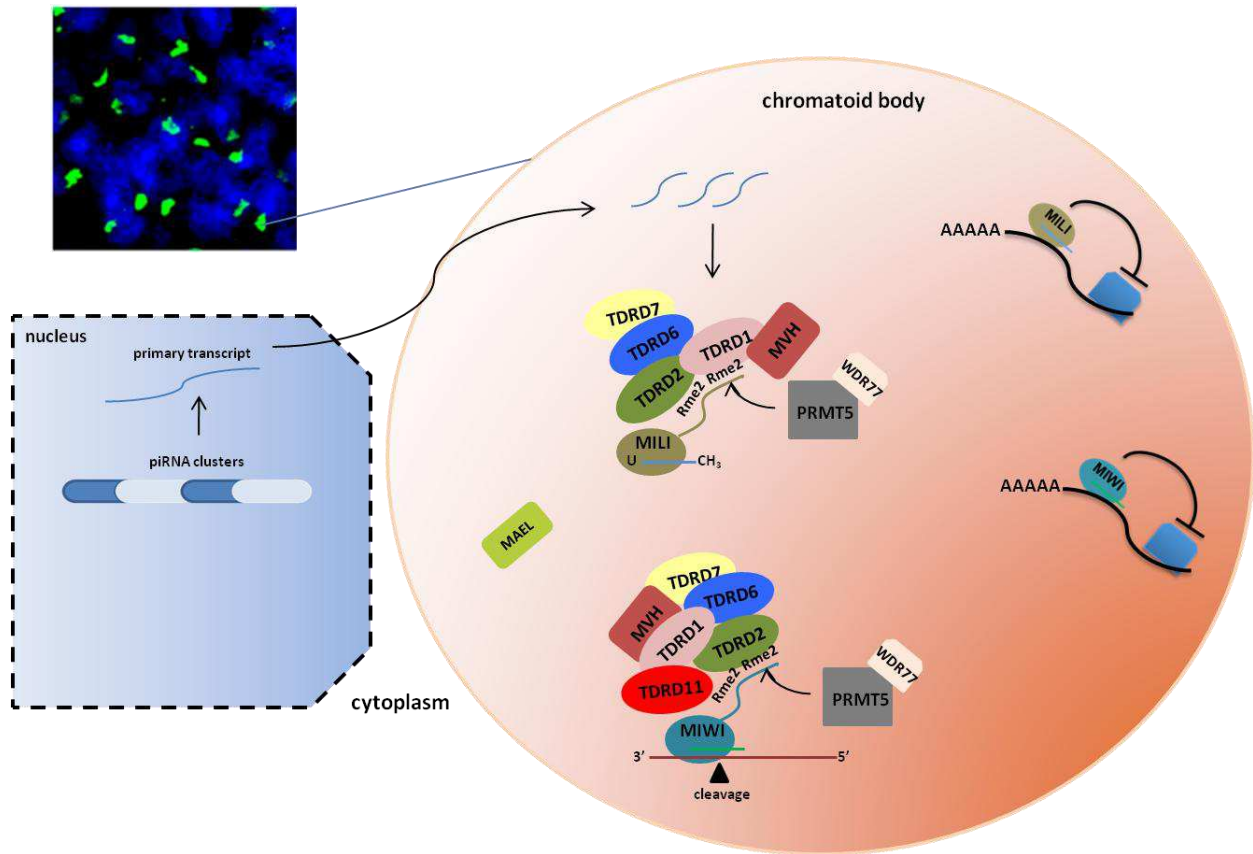


Figure 1.12. Compartmentalization of piRNA pathway-associated factors during the pachytene stage of spermatogenesis. In pachytene stage, piRNAs and their interacting partners are accumulating in chromatoid bodies (CBs) that are adjacent to the nucleus. Localization experiments for Mili in spermatocytes show the chromatoid body's location (Vagin et al., 2009). IN CBs, many TDRD proteins are found, with their exact role still to be missing. The proposed functions for Mili and Miwi in translational repression are also shown (see also [figure 1.7](#)).

1.2.6 - Function

1.2.6.1 - Post-transcriptional transposon silencing

PIWI proteins form ribonucleoprotein complexes with piRNA molecules and they are involved in a surveillance mechanism where they target transposons transcripts and suppress them. This silencing mechanism is most likely to make use of the endonucleolytic activity of the Argonaute

proteins, similar to miRNAs and siRNAs. In *Drosophila*, the three PIWI proteins exhibit slicer activity *in vitro* (Gunawardane et al., 2007; Saito et al., 2006). More specifically, in germline cells Aub (and Piwi) and AGO3 involved in the amplification loop mediate RNA-mediated cleavage to their target RNAs and thus destroy TE mRNA transcripts. In somatic cells, where Piwi protein is found, a mechanism of target-RNA cleavage is also believed to take place (fig.1.4, fig.1.6.A) (Saito et al., 2006).

Another model proposed in flies involves targeting of messenger RNAs and induction of mRNA decay. Aub-piRNA complexes can function like animal miRNAs, where they were found to silence the mRNA of the embryonic posterior morphogen *nanos* (Rouget et al., 2010) by recruiting the deadenylation enzyme CCR4 and the RNA-binding protein Smaug (fig.1.6.C).

In mouse, all three PIWI proteins are involved in transposons silencing, however, only Miwi and Mili were shown to possess slicer activity (fig.1.7.B, fig.1.7.C) (De Fazio et al., 2011; Reuter et al., 2011). In pre-pachytene piRNAs, Mili's role is possibly to target and destroy transposons mRNAs by endonucleolytic cleavage and also fuel Miwi2 with secondary piRNAs (De Fazio et al., 2011). In pachytene piRNAs, where no signatures of ping-pong cycle are found and proteins Mili and Miwi are expressed, it is believed that both proteins use their slicer activity to silence their targets through a linear pathway. The origin of pachytene piRNAs as discussed above corresponds to mostly unannotated regions and to a lower extent to transposons sequences, and the latter are thought to be targeted by Miwi for destruction (Reuter et al., 2011). However, the function of the majority of pachytene piRNAs remains unknown, as well as the implication of PIWI proteins.

Translational control of mRNA levels is debated in the murine piRNA pathway. There has been reported that Miwi and Mili associate with the mRNA cap-binding complex, however, there is no evidence reporting the mechanism of target selection or how piRNAs are involved (fig.1.7.D, fig.1.7.E) (Grivna et al., 2006; Pillai and Chuma, 2012; Unhavaithaya et al., 2009). Other studies support the notion that Miwi is not regulating mRNA levels (Reuter et al., 2011; Unhavaithaya et al., 2009).

1.2.6.2 - Transcriptional transposon silencing

In flies, the only protein reported to participate in silencing transposons in a transcriptional level is Piwi. In OSC cells, a slicer-deficient mutant of Piwi was found to be loaded with mature piRNAs and rescue transposons silencing (Saito et al., 2010) suggesting that it can induce repression of transposons through effects on chromatin (Thomson and Lin, 2009). Interestingly, some studies implicate Piwi in influencing position-effect variegation in somatic cells, by directing Heterochromatin protein 1A (HP1A) to heterochromatin (fig.1.6.B) (Brower-Toland et al., 2007; Pal-Bhadra et al., 2004). Other studies, however, showed that HP1 recruitment does not depend on Piwi protein (Moshkovich and Lei, 2010). Thus, the exact mechanism for Piwi's function on chromatin level is still a mystery.

In mice, in the pre-pachytene stage, all the epigenetic marks on the gametic DNA material are erased and a new cycle of DNA methylation is taking place, thus providing a stable and reversible mark for transcriptional repression. As a result, transposons remain silenced and genome integrity is maintained. In particular, methyl groups are added on cytosines within CG dinucleotides by the *de novo* methyltransferases Dnmt3A and Dnmt3L, after chromatin rearrangements on histone tails (fig.1.11) (Kaneda et al., 2004; Kato et al., 2007; Ooi et al., 2007). The involvement of piRNAs can be a potential trigger for this methylation mechanism, since Miwi2 in complex with anti-sense piRNAs is transferred to the nucleus and is believed to mediate *de novo* methylation of transposons' promoter regions. (Aravin and Bourc'his, 2008; Aravin et al., 2007a; Kuramochi-Miyagawa et al., 2008). Two observations strengthen this mechanism: i) Slicer-deficient Miwi2 has normal spermatogenesis and no transposon activation (De Fazio et al, 2011) and ii) Dnmt3L^{-/-} male mice are sterile and lose TE DNA methylation in testes, a phenotype almost identical for *Mili* and *Miwi2* mutants (Aravin et al., 2007b; Carmell et al., 2007). Also, in *dnmt3L* mutants, piRNAs are still produced, which means that piRNA pathway is upstream of DNA methylation (Aravin and Bourc'his, 2008). However, the biochemical and molecular aspects of this mechanism are unknown. One possibility is that Miwi2-piRNAs base pair with a nascent transcript and sequentially recruit the methylation factors, similar to the RITS complex in *S.pombe* (see section 1.1.3.2, fig.1.3.C).

1.2.6.3 - Silencing of protein-coding genes

piRNAs in *Drosophila melanogaster*, were also found to target non-TEs. These targets are protein coding genes, like the *Stellate*(Ste) genes in flies testis (Aravin et al., 2001). On Y chromosome, the Su(ste) locus give rise to piRNAs that bind Aubergine and AGO3 and silence Ste genes (fig.1.6.D) (Nagao et al., 2010; Nishida et al., 2007). If these piRNAs are absent, Ste locus is transcribed and produces STE crystals that cause male infertility. Other examples of protein-coding gene silencing involve *Fascilin 3* (Fas3) in the ovaries, which is repressed by piRNAs that are bound on Piwi protein and they are derived from the 3' UTR of the *traffic jam's* transcript (Li et al., 2003; Saito et al., 2009). Finally, the essential maternal gene *vasa* in *Drosophila's* testis is down-regulated by Aub and AGO3- bound piRNAs ((Nishida et al., 2007). The mother deposits these piRNAs and evidence suggests that in the piRNA pathway, maternally deposited piRNAs and PIWI proteins protect the developing embryo from transposon activation (Brennecke et al., 2008). In the phenomenon, called hybrid dysgenesis (Kidwell and Sang, 1986), male flies carrying transposon elements when crossed with females that do not bear these sequences give sterile offspring due to transposon activation, while in the reciprocal crossing the progeny are fertile. Thus, the protection is conferred by the piRNA pathway, which is transmitted by the mothers (Brennecke et al., 2008) (fig.1.6.A).

1.2.7 - Summary

Transposable elements can cause severe mutations in many organisms and defense mechanisms have been developed, in order to prevent transposon insertion in protein-coding gene regions. Both piRNAs and endo-siRNAs are diverse in a sequence manner; however they are able to associate with and silence transposons sequences, distinguishing them from protein-coding transcripts. While piRNAs function mainly in the germline, endo-siRNAs can protect the somatic cells from the harmful invasion of TEs, suggesting that transposon surveillance is achieved through a cross-talk between piRNA and endo-siRNA mechanisms.

1.3 - Post-translational modification of proteins

Proteins that form the base of the behavior of every cell are produced by selective gene expression and are expanded by alternative RNA-splicing. However, the cell is not in a static state and has to adapt to dynamic external and internal environmental conditions, such as growth factors, ligands on the surface, cell-cycle checkpoints, DNA damage, etc. The response to these cues is usually mediated by reversible covalent modifications on proteins. As a result, protein molecules can create new binding sites for interacting molecules and synergize to regulate cell behavior. These post-translational modifications include phosphorylation, methylation of arginine and lysine residues, acetylation, ubiquitinylation or sumoylation of lysine residues, and prolyl hydroxylation (Yang, 2005). The following chapter will deal with methylation as a post-translational modification and the protein domains that recognize it, giving emphasis on Tudor domains in a structural and functional fashion.

1.3.1 - Methylation as a protein modification

Over the years it has become established that methyl groups are major controlling elements in protein function. This modification generates various sets of chemical interactions that are involved in a repertoire of regulatory mechanisms.

The addition of a methyl group can take place on an arginine or lysine residue. Methylated lysine is commonly found on histone tails and is involved in chromatin remodeling, gene expression or gene silencing. Lysine residue can be found mono- di- and tri-methylated state and be a recognition target for a set of protein domains (see section [1.3.2](#)).

Arginine methylation is one of the most widespread modifications found on proteins in mammalian cells (Najbauer et al., 1993; Paik et al., 1980). Arginine is in a methylated state in histones or in proteins, where, apart from epigenetic regulation, is a substrate for protein modules that regulate silencing of gene elements, trigger signaling pathways and the formation of ribonucleoproteins (RNPs). Arginine is commonly found in three methylation states. Addition of two methyl groups on one of the terminal nitrogen atoms of the guanidino group results in an asymmetric dimethylated arginine (aDMA). The symmetrical dimethylated arginine (sDMA)

is formed when the two methyl groups are placed on each of the two terminal guanidine nitrogens and the monomethylated arginine (MMA) derivative rises when on methyl group is added to the terminal nitrogen atom ([fig.1.14.B](#)) (Bedford and Clarke, 2009)

Methylation of arginine residues is formed by specific enzymes, called protein arginine methyltransferases (PRMTs). There are eight mammalian PRMTs and are classified into two groups: i) type I, including PRMT1-4, PRMT6 and PRMT8 that catalyze the formation of aDMAs, and ii) type II, including PRMT5 and PRMT7, that generate sDMAs. Monomethylation is generated by both types of PRMTs as an intermediate.

In methyllysine signaling, there have been identified demethylases that specifically remove methyl groups and together with methyltransferases orchestrate a dynamic regulation of protein-protein interaction network. On the other hand, protein arginine demethylation is still under investigation; recent work has shown that a protein of the jumanji family, JMJD6, is directly demethylating arginine residues from histones (Chang et al., 2007). Thus, it is still a debate if arginine methylation is a stable modification or not.

1.3.2 – “Royal domain superfamily of proteins”

Extensive database analysis and a global search for sequence and structural similarities led to the establishment of the ‘Royal domain superfamily’ of proteins (Maurer-Stroh et al., 2003). These proteins are readers of methylation marks on lysine or arginines and have a common barrel-like protein fold comprised of twisted antiparallel β -strands, able to accommodate methyl-modified residues, and probably originate from a common ancestor (Maurer-Stroh et al., 2003; Taverna et al., 2007). The binding pocket, termed aromatic cage, is composed of two to four aromatic residues and they stabilize their methylated ligands through hydrophobic and electrostatic interactions ([fig.1.13](#), [fig.1.14.A](#)) (Taverna et al., 2007).

The members of the Royal family are the Chromo, MBT, PWWP, plant Agenet and some Tudor domains and they are implicated mainly in histone modification and chromatin remodeling mechanisms ([fig.1.13](#)). The chromatin-binding (Chromo) domain is part of the Heterochromatin protein 1 (HP1) and is usually found to be part of proteins that are associated with chromatin,

since these domains are binding methylated lysines on histones (Ball et al., 1997; Eissenberg, 2001; Nielsen et al., 2002). However, recently it was revealed that the chromodomain of the chloroplast signal recognition particle (cpSRP) can also accommodate non-methylated arginines (Holdermann et al., 2012). Malignant-Brain-Tumor (MBT) protein domains are involved in the transcriptional repression of developmental control genes in *Drosophila*, such as *Hox* genes, and they can bind to mono- or dimethylated lysines on histones H3 and H4 (Trojer et al., 2007). The PWWP domain possesses a conserved Pro-Trp-Trp-Pro motif and is found on nuclear proteins, interacting with proteins that specifically recognize histone modifications (Stec et al., 2000). Moreover, recently it was identified that PWWP domains can also recognize trimethylated lysines on histone H3 tails (Vezzoli et al., 2010). The plant Agenet domains were identified using a sequence and structural homology approach and are suspected to be the plant homologues of Tudor domains (Maurer-Stroh et al., 2003). However, their exact role and function remains unknown.

1.3.2.1 – Tudor domains as methyl-arginine/lysine readers

Tudor domains are the broader class of the Royal superfamily and consist of a large number of proteins. They have varying functions; however, they share a common 50 amino acid motif that serves as a protein-protein interaction module, known as prototypic Tudor domain. The Tudor domain was first identified in *Drosophila*, where the *Tud* gene belongs to maternally expressed genes in the posterior pole and its absence results in lethality and sterility of the embryos (Boswell and Mahowald, 1985). The protein encoded (as described in section [1.2.4.1](#)) has eleven Tudor domains and was first reported to be implicated in RNA binding (Ponting, 1997). Subsequently, sequence alignments of individual domains identified Tudor domains in a variety of organisms, including fungi, plants and animals (Chen et al., 2011).

The mammalian genome encodes for more than 30 proteins encompassing one or multiple copies of Tudor domains and they can bind methylated arginines and lysines. They are implicated in many cellular processes, including RNA metabolism, germ cell development, transposons silencing, DNA damage response and chromatin remodeling and histone modifications. The two latter involve methyllysine Tudor binders, a common feature of the

Royal family of proteins. In particular, Tudor domains targeting methylated lysines are usually close or adjacent to chromatin-binding domains, such as PHD and Chromo domains. Two well-studied examples include the proteins JMJD2A and 53BP1 ([fig.1.13](#)). JMJD2A is a member of the JmjC histone demethylase superfamily and in complex with the protein Rb, function as transcriptional repressors (Gray et al., 2005; Zhang et al., 2005). This ~110 kDa size protein encodes at its C-terminus two tandem Tudor domains, meaning that one Tudor domain is followed by another one. The crystal structure revealed an elongated Tudor domain fold assembly, where the two Tudor domains are forming a saddle-shaped structure, with each domain resembling the prototypic Tudor domain (Huang et al., 2006). This hybrid tandem Tudor domain has the ability to bind two different peptides harboring a tri-methylated lysine, H3K4me3 and H4K20me3 with equal affinity. Both peptides are interacting with the second Tudor domain in different orientations along the protein surface and the methyllysine is accommodated in the aromatic cage formed by aromatic residues. The p53-binding protein 1 (53BP1) is a central mediator of the DNA damage checkpoint (Charier et al., 2004) and its two tandem prototypic Tudor domains are binding the dimethylated lysine 20 on histone H4 (H4K20me2). The individual β -barrel Tudor domains are connected by a 6-residue linker and the histone peptide binds to one of the Tudor domains where the dimethyllysine is caged by 4 aromatic residues (Botuyan et al., 2006).

Because of their structural and sequence similarity, other Tudor domains can also be considered as part of the Royal family of proteins. For instance, the Fragile X mental Retardation Protein (FMRP) and its autosomal paralogues, FXR1 (Fragile X mental Retardation Syndrome-related protein 1) and FXR2, contain tandem tudor domains of ~60 amino acids each that share β -strand folding and are reported to bind methylated lysine partners (Adams-Cioaba et al., 2010; Ramos et al., 2006). Finally, the crystal structure of the human spindlin1 protein, which is involved in cell cycle regulation, contains three repeats of prototypic Tudor domains, revealing a novel arrangement of tandem Tudor folding (Zhao et al., 2007). Thus, the Royal superfamily of proteins can include a variety of different Tudor domains, implicated in many cellular processes.

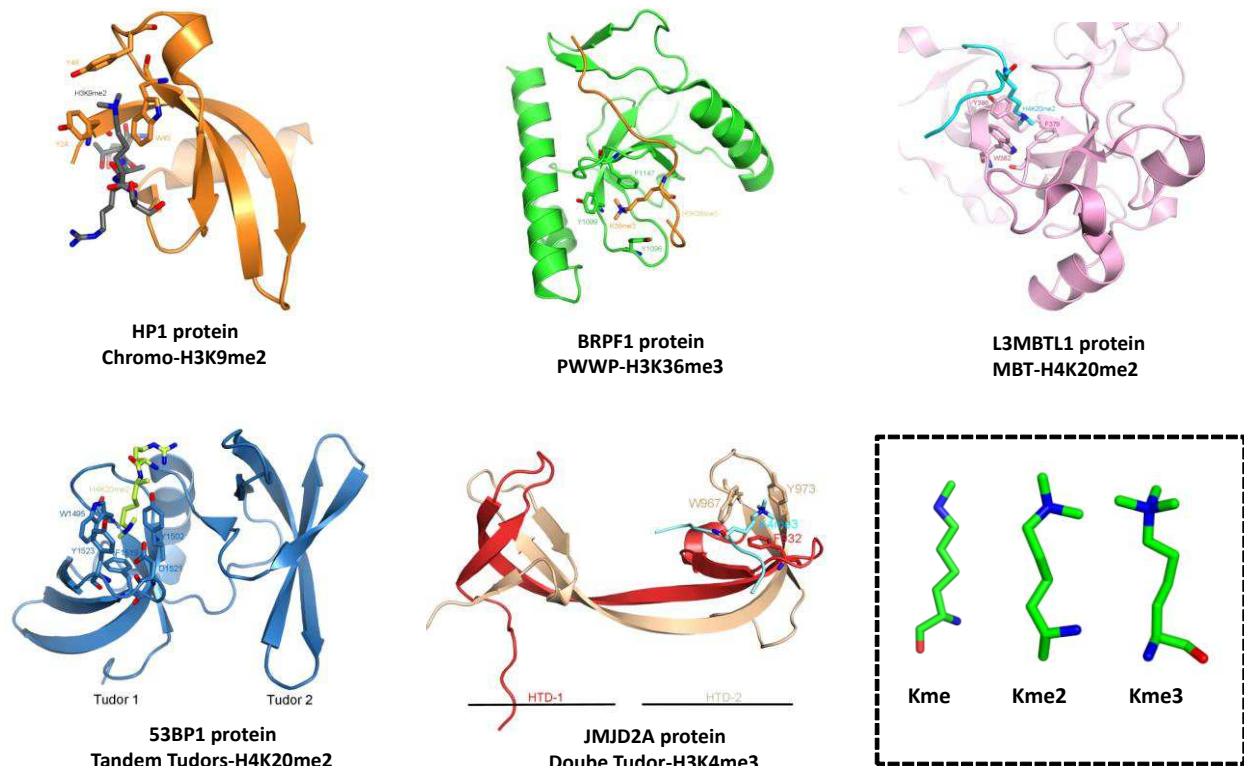


Figure 1.13: The royal superfamily domains and their methylated lysine ligands. Depicted are the crystal structures for each member of the Royal family of proteins, bound to their methylated histone targets: Chromo-H3K9me2 (PDB 1KNA) (Jacobs and Khorasanizadeh, 2002); PWWP-H3K36me2 (PDB 2X35) (Vezzoli et al., 2010); MBT-H4K20me2 (PDB 2PQW) (Min et al., 2007); 53BP1-H4K20me2 (PDB 2IG0) (Botuyan et al., 2006); JMJD2A-H3K4me3 (PDB 2GFA) (Huang et al., 2006). In a stick-model representation, the aromatic cage and the accommodated methylated lysine are depicted. In the dashed box, stick models of different post-translationally modified lysines (mono-, di-, tri-methylated) are shown.

Proteins that have Tudor domains recognizing methylated arginine residues are usually involved in RNA metabolism (Pahlich et al., 2006), although recently one Tudor family protein was found to be involved in transcriptional regulation (TDRD3 protein, see later). One extensively studied example is the SMN (Survival motor neuron) and SPF30 (motor neuron-related splicing factor 30) proteins, which are implicated in the assembly of small nuclear ribonucleoprotein complexes and pre-mRNA splicing respectively (Buhler et al., 1999; Meister et al., 2001; Pellizzoni et al., 1998; Pellizzoni et al., 2002; Rappsilber et al., 2001). Both proteins contain a prototypic Tudor domain of ~50 amino acids that recognize methylated arginines (fig.1.14.A). More specifically, a sub-group of proteins from the SM protein family (SMB, SMB', SMD1, SMD3) contain arginine-glycine (RG) clusters, which are symmetrically dimethylated by

PRMT5 (Brahms et al., 2001; Brahms et al., 2000; Friesen and Dreyfuss, 2000; Friesen et al., 2001). These proteins are taking part in the formation of small nuclear ribonucleoprotein (snRNP) particle assembly, and SMN protein, through methylated arginine recognition by its Tudor domain, facilitates the formation of this ribonucleoprotein core. After SM-SMN interaction, maturation of snRNP takes place, which is critical for pre-mRNA splicing. The correct mRNA splicing depends strongly on the methylation status of SM proteins and on Tudor-RG clusters interaction, since mutation of critical for the binding residues in the Tudor domain can cause improper assembly of the spliceosome and ultimately the neurodegenerative spinal muscular atrophy (SMA) disease (Brahms et al., 2001; Selenko et al., 2001).

The aromatic cage of the SMN Tudor domain was also found to accommodate aDMA ligands of the transcription factor CA150 (Cheng et al., 2007) and the structural details of the Tudor-aDMA interaction were recently revealed (Tripsianes et al., 2011). In both sDMA and aDMA recognition, SMN and SPF30 Tudor domains do not depend on adjacent to the methylarginine residues (Tripsianes et al., 2011). Another Tudor-containing protein showing high sequence homology with SMN and SPF30 proteins, known as TDRD3, was also found to interact with methylated RG-containing peptides from SM proteins (Cote and Richard, 2005). However, TDRD3 was recently involved in transcriptional activation, by recognizing aDMAs on histone tails (Yang et al., 2010). Furthermore, in a protein-domain array screening, TDRD3 was able to specifically recognize in an asymmetrically-demethylated dependent manner the C-terminal domain of RNA polymerase II, with the functional importance of this interaction to remain unclear (Sims et al., 2011). Finally, TDRD3 also appeared in the cytoplasm, interacting with the Fragile X Mental Retardation Protein (FMRP) in stress granules (Goulet et al., 2008; Linder et al., 2008). Thus, the Tudor domain of TDRD3 appears to be a universal methylarginine reader involved in various cellular pathways.

As previously discussed, the Tudor domain-containing proteins (TDRD proteins) comprise the rest and largest group of the methylarginine Tudor proteins and they are involved in the piRNA pathway. TDRD proteins have enriched expression pattern during spermatogenesis in mice and

ogenesis in *Drosophila*, where they are found to associate with the N-terminal arginine-rich tail of PIWI proteins and participate in the silencing of transposons.

Tudor domain-containing proteins in the piRNA pathway recognize symmetrically dimethylated arginines via their aromatic cage; however they are containing additional structural elements for the recognition of their methylated ligands. Crystal structures of Tudor domains from the fly protein TUD and the mammalian TDRD11 (SND1) in complex with sDMA ligands revealed that apart from the ~60 amino acid prototypic Tudor domain, they have an additional domain called SN-like domain, that is indispensable for the recognition of PIWI-derived methylated peptides ([fig.1.14.A](#)) (Liu et al., 2010a; Liu et al., 2010b). The total ~180 amino acid Tudor domain is thus termed extended Tudor domain (eTud, eTudor) and most of the TDRD proteins are predicted to have similar fold. As mentioned before, proteins Vasa and MVH, apart from sDMAs they also have aDMAs and MMAs and interaction studies showed that eTuds can also recognize them but with reduced affinity compared to sDMAs (Liu et al., 2010b).

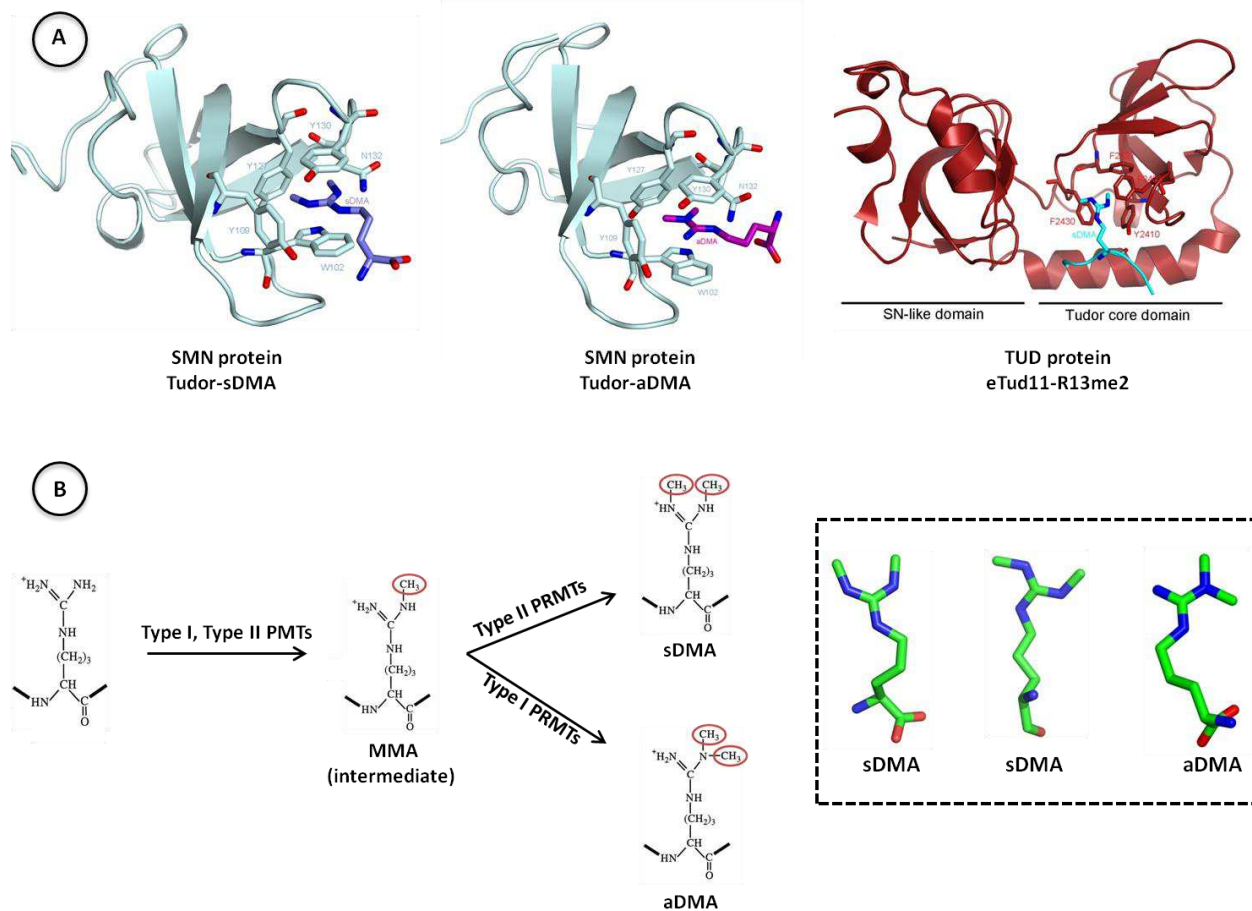


Figure 1.14: Structural features of the Tudor domains and their binding to methylated arginines. **A.** The SMN protein is able to accommodate through its aromatic cage (stick model) symmetrically (sDMA) (PDB 4A4E) and asymmetrically (aDMA) (PDB 4A4G) dimethylated arginine (Tripsianes et al., 2011). The extended Tudor domain of *drosophila* TUDOR, requires the adjacent SN-like domain for recognition of methylated peptides (PDB 3NTI) (Liu et al., 2010a). The symmetrically dimethylated arginine is also found in an aromatic residue-formed cage. **B.** The methylation steps of arginine residues by the mammalian methyltransferases PRMTs are shown. In the dashed box, stick model representation of methylated arginines in a symmetrical (two forms) and asymmetrical form.

The exact role of TDRD proteins in the piRNA pathway remains elusive; nevertheless, the multiple numbers of eTudor domains suggest that they serve as a molecular scaffold for the recruitment of methylated RG-clusters on the N-terminus tails of PIWI proteins. The piRNA pathway is dictated by an exchange of piRNAs between PIWI proteins and one possible role for Tudor-containing proteins is to bring them in close proximity in order to efficiently complete their functional role. The structural topology and the function of TDRD proteins will be further discussed in the Results & Discussion chapter.

1.4 – Aim of the thesis

The PhD project aimed to structurally and functionally characterize piRNA pathway components, and was held in close collaboration with Ramesh Pillai's group (EMBL, Grenoble outstation). At the beginning of this work, Pillai's lab identified TDRD1 protein as an essential component for transposon silencing by the piRNA pathway (Reuter et al., 2009). Furthermore, it was shown that Tudor domains of TDRD1 recognize symmetrically dimethylated arginines of the key PIWI protein Mili. Finally, in the late stages of the PhD, this study focused on the characterization of the N-terminal MYND domain of TDRD1 as an interactor of the recently identified piRNA pathway protein FKBP6. The goal of this project was to:

1. Structurally and functionally characterize the multiple and single Tudor domains of TDRD1 using X-ray Crystallography and/or biophysical methods (Small Angle X-ray Scattering).
2. Identify the way Tudor domains of TDRD1 interact with methylated arginine marks on Mili's N-terminus. Determine binding kinetics of protein-ligand recognition and investigate if all Tudor domains of TDRD1 are involved in methylarginine recognition. Finally, perform crystallization experiments of Tudor domains-methylated peptide complexes.
3. Determine how MYND domain interacts with FKBP6 protein and explore the binding specificity to its separate domains. Try to structurally characterize the potentially formed protein complex.

Overall, the aim was to understand in the atomic level how Tudor domains recognize the N-terminal region of PIWI protein Mili and shed light on the role of TDRD1 as a protein-protein interaction platform in the mammalian piRNA pathway.

2 – Materials and Methods

2.1 - Genetic engineering methods

The cDNA sequences of full-length mouse TDRD1 and mouse FKBP6 were kindly provided from Reuter et al., 2009 and Jordi Xiol (Ramesh Pillai Laboratory, EMBL, Grenoble).

Amplification of DNA constructs was performed using Polymerase Chain Reaction (PCR), following a standard protocol. The oligonucleotide primers were obtained from Invitrogen Life Technologies (Saint Aubin, France) and were designed to have a basic melting temperature (T_m) of 48-62 °C.

The PCR reaction had a final volume of 50 μ l, composed of:

Reagent	Volume (μ l)
Primer 1 (100 μ M)	0.5 μ l
Primer 2 (100 μ M)	0.5 μ l
Template DNA	50-100 ng
Pfu buffer (10X)	5 μ l
dNTPs (10 mM)	1 μ l
Pfu DNA polymerase	1 μ l
H ₂ O	Up to 50 μ l

Pfu DNA polymerase was obtained from Protein expression and purification facility, EMBL Heidelberg. Pfu buffer was from Stratagene (La Jolla, CA, USA) and the nucleotides from Fermentas (Burlington, Canada).

The PCR amplification program used was as follows:

Temperature	Time	Cycles
94 °C	1 min	1
94 °C	30 sec	32
T_m -5 °C	30 sec	
72 °C	1 min	
72 °C	10 min	1
4 °C	“for ever” (stop)	1

PCR products were then purified using the PCR purification kit (Qiagen, GmbH, Hilden, Germany) and cleaved overnight at 37 °C with the desired enzymes (mainly NcoI, NotI, New England Biolabs, Ipswich, MA, USA). Cleaved DNA fragments were purified from agarose gel using the Gel Extraction kit (Qiagen) and concentrated using a speed-vac system. Vector used were cleaved and purified in an analogous manner. For sub-cloning in a different vector, the same procedure was applied, after cleavage of the insert from the original vector.

Ligation was performed using 100 ng of vector and 1:5 molar ratio of insert in a final volume of 5 µl; the total reaction volume was 10 µl, after the addition of 5 µl of Ligation solution I (Takara Bio Inc., Otsu, Japan). The ligation was performed at 16 °C for 2-3 hours and then transformed directly to streptomycin-resistant *E.coli* TOP10 chemically competent cells (Invitrogen Ltd., Paisley, UK). Colonies grown on LB-agar plates (carrying the appropriate antibiotic resistance) were picked up to be inoculated in 5 ml LB medium (using the same antibiotic for selection) and cells were grown for 10-14 hours at 37 °C. Plasmic DNA was isolated using the Miniprep kit (Qiagen or Promega, Madison, WI, USA) and after 2 hour-digestion with the appropriate restriction enzymes, the presence of the insert was checked on an agarose gel. Positive clones were selected and sent for sequencing to Macrogen (Netherlands) to confirm the correct sequence of the insert.

For gene mutation, two different approaches were used; Quickchange site-directed mutagenesis (Stratagene) and the overlapping PCR method. All the mutated constructs were point mutants.

The Quickchange method involves gene and vector amplification simultaneously in a circular manner. For triplet mutations, primers were designed to contain the mutation site in the middle and 15-25 flanking bases at the 3' and 5' ends. The T_m of the mutated primers is preferentially high, above 75 °C. The second primer contains the reverse complementary sequence. The PCR reaction has a final volume of 50 µl and is as follows:

Reagent	Volume (μ l)
Primer 1 (100-125 ng)	1 μ l
Primer 2 (100-125 ng)	1 μ l
Template DNA	50-100 ng
Pfu buffer (10X)	5 μ l
dNTPs (10 mM)	1 μ l
Pfu Turbo DNA polymerase	1 μ l
H ₂ O	Up to 50 μ l

Pfu Turbo polymerase was obtained from Invitrogen and the PCR amplification program is as follows:

Temperature	Time	Cycles
95 °C	1 min	1
95 °C	50 sec	18
60 °C	50 sec	
68 °C	7 min	
68 °C	7 min	1
4 °C	“for ever” (stop)	1

After plasmid amplification, 1 μ l DpnI (New England Biolabs) was added for 1 hour at 37 °C to digest the parental methylated plasmid. Transformation, digestion and sequencing of the positive clones were carried out as described previously.

In the overlapping PCR method, complementary primers generate two DNA fragments, having overlapping ends, in the first PCR reaction. In a second reaction, these fragments are combined (fused) and the overlapping ends anneal, allowing the 3' overlap of each strand to serve as a primer for the 3' extension of the complementary strand (Ho et al., 1989) (fig.2.1). PCR amplification program, ligation, transformation, positive clone selection and sequencing is performed in the same way as the wild-type cloning.

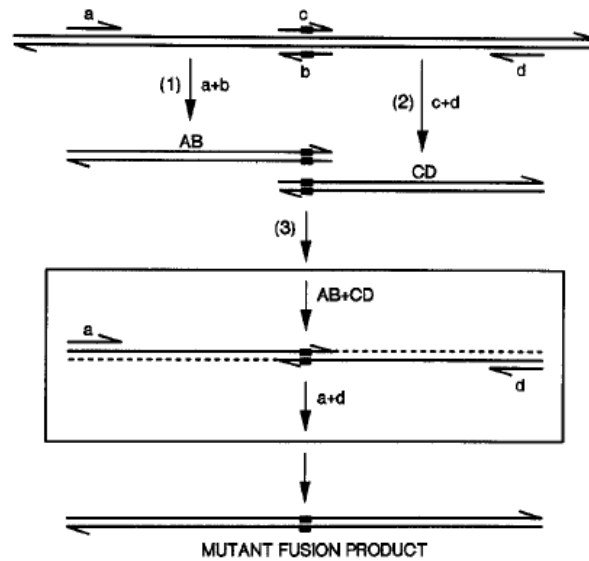


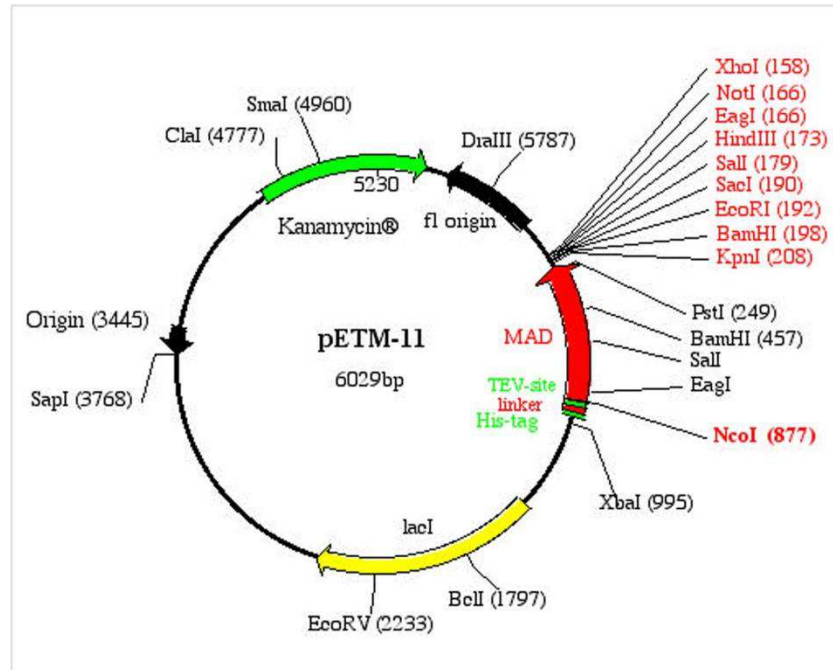
Figure 2.1 Overlapping PCR method (Ho et al., 1989)

2.2 - Protein expression and purification

2.2.1 - Recombinant protein expression in *E.coli*

For the expression of the desired cloned genes, the vectors pETM-11 and pETM-30 were used, while for co-expression experiments, pCDF-Duet-1 vector was manipulated. In almost all cases, the restriction enzymes NcoI and NotI were chosen, allowing easy sub-cloning within different vectors.

All protein constructs that were expressed in pETM11 vector were having an N-terminal His₆-tag, followed by a 15-residue linker including the sequence ENLYFQ-G that is recognized and cleaved by the TEV protease, between glutamine and glycine. After histidine-tag cleavage by TEV, the protein has at its N-terminus the sequence Gly-Ala-Met, deriving from TEV site and NcoI restriction site. In pCDF-Duet-1 vector used for co-expression of proteins, protein of interest has no tags and at its N-terminus carries the residues Met-Gly. Figures of vector maps used are depicted in [fig. 2.2](#).



T7 promoter --> lac operator XbaI
 GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAAT
 CTTAATTATGCTGAGTGATATCCCCTTAACACTCGCCTATTGTTAAGGGGAGATCTTTA

rbs His.Tag
 AATTTTGATTTAACTTTAAGAAGGAGATATACCATGAAACATCACCATCACCATCACCCC
 TTA AAACTAAATTGAAATTCTTCCTCTATATGGTACTTTGTAGTGGTAGTGGTAGTGGGG
METLysHisHisHisHisHisHisPro

TEV site NcoI
 ATGAGCGATTACGACATCCCCACTACTGAGAATCTTTATTTTCAG GCGCCATGGCGGCG
 TACTCGCTAATGCTGTAGGGGTGATGACTCTTAGAAATAAAAGTC CCGCGGTACCGCCGC
 MetSerAspTyrAspIleProThrThrGluAsnLeuTyrPheGlu|GlyAla**MET**AlaAla

GCGGTTCGGATGAAC.. **612bp**..GACAGTCACAAGGCGTGTCTTGGTCTCTAACTAGTG
 CGCCAAGCCTACTTG... **MAD**...CTGTCAGTGTCCGCACAGAACCAGAGATTGATCAC
 AlaValArgMetAsn.. **204aa**..AspSerHisLysAlaCysLeuGlyLeu***

BamHI SacI NotI
 KpnI EcoRI Sall HindIII EagI XhoI His-taq
 GTACCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGC GGCCGCACTCGAGCACCACCAC
 CATGGCCTAGGCTTAAGCTCGAGGCAGCTGTTTCGAACGCCGCGTGAGCTCGTGGTGGTG
 HisHisHis

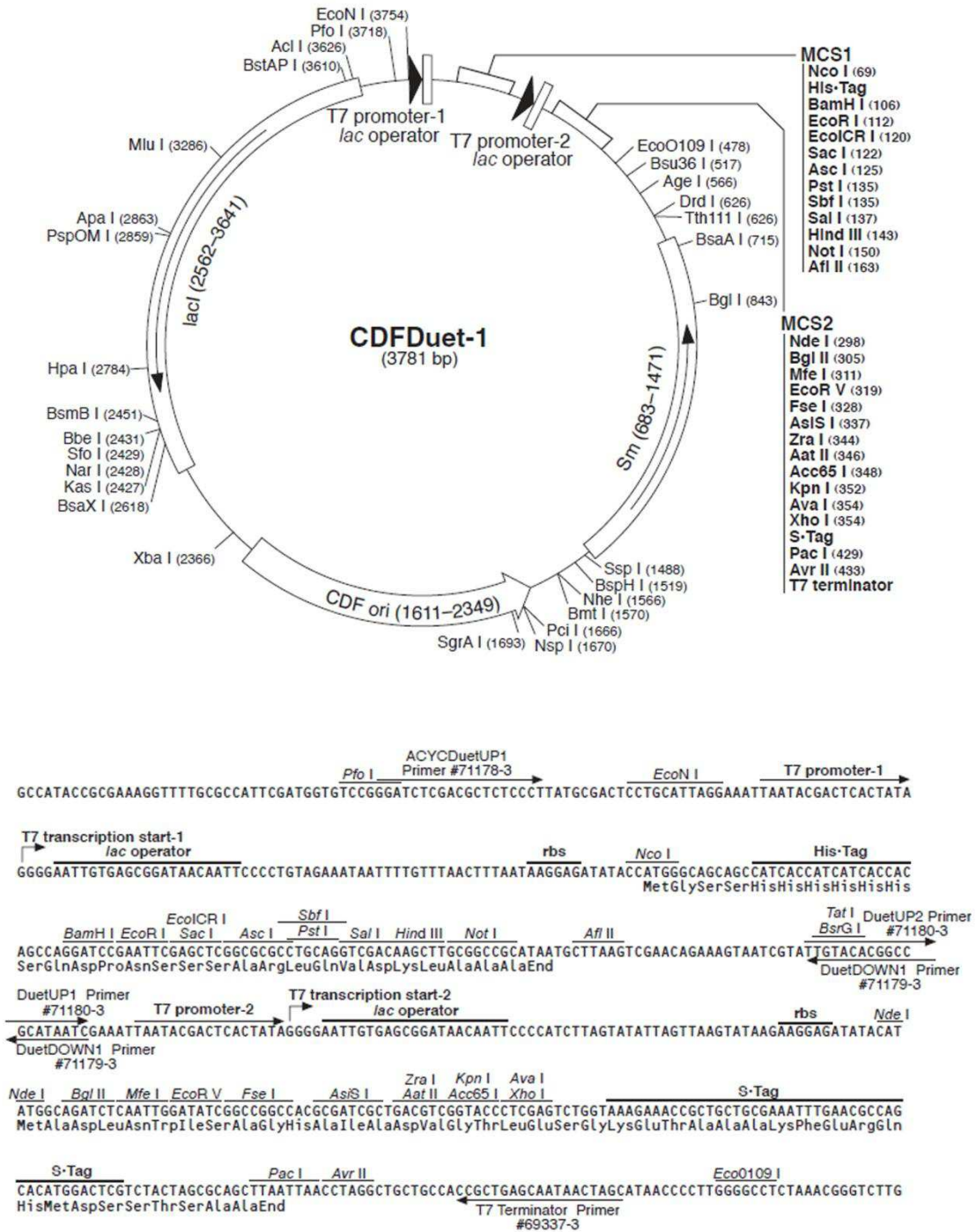


Figure 2.2 Vector maps. Shown are the features of all the vectors used in this study, including promoter, antibiotic resistance and Multi Cloning Site.

pETM11 and pCDF-Duet-1 vectors for tagged and untagged proteins respectively are compatible for co-expression and their simultaneous usage is possible because of their different antibiotic resistance. For all vectors, the expression of the gene of interest is inducible with isopropyl-beta-thio-galactoside (IPTG, Euromedex, Mundolsheim, France). All the proteins were expressed using *E.coli* BL21(DE3)-Star cells (Invitrogen) that carry a mutation in the gene coding for RNase E, thus improving the stability of mRNAs and increasing protein yield. The expression was carried out using Luria-Bertani (LB) medium and the appropriate antibiotic at a final concentration of 50 µg/ml. An overnight grown pre-culture of 10-25 ml was used to inoculate one liter of LB medium. The bacteria were left to grow at 37 °C in a shaking incubator at 150 rpm until the OD₆₀₀ reached 0.6-1.0; induction was performed after adding IPTG at a final concentration of 1mM and cells were left shaking at 20 °C for 12-14 hours.

2.2.2 - Expression of Se-Methionine-substituted proteins in *E.coli*

A 5 ml LB-antibiotic pre-culture was grown for 8 hours; the cells were spun down to eliminate the LB medium and used to inoculate 50 ml of M9 minimal medium (for 1 Liter: 10 gr Na₂HPO₄·7H₂O, 3 gr, KH₂PO₄, 0.5 gr NaCl, 1 gr NH₄Cl) that was grown overnight at 37 °C. 50 ml of M9 pre-culture were used to inoculate one Liter of M9-rich culture (M9 minimal + 1ml MgSO₄ 1M, 0.1 ml CaCl₂ 1M, 10 ml Glucose 20 %, 2ml Thiamine 1mg/ml, 50 µg/ml appropriate antibiotic) until the cells reached an OD₆₀₀ of 0.6-0.9 and the following amino acids were added: 100 mg Lysine, 100 mg Phenylalanine, 100 mg Threonine, 50 mg Isoleucine, 50 mg Leucine, 50 mg Valine, 50 mg Se-Methionine per liter of culture. After 20 min incubation at 20 °C, induction and purification were carried out as described below.

2.2.3 - Protein purification

After the induction of bacterial cell cultures, bacteria were harvested by centrifugation at 6000 g for 20 min at 4 °C. Cell paste was resuspended in a buffer containing 20 mM Tris-HCl pH 7.6, 150 mM NaCl, 20 mM imidazole, 5 mM β-mercapto-ethanol (lysis buffer) and complete protease inhibitor EDTA-free (Roche, Neuilly sur Seine, France) (1 pill for 50 ml lysis buffer). The volume of lysis buffer used was 25 ml per Liter of culture. Cell lysis was carried out by

sonication of the bacteria in a metal Becker, surrounded by ice to avoid overheating of the sample. Cells were sonicated from 1 min, 40 sec (for ~25 ml volumes) up to 6 minutes (for 150 ml volumes) in total, using bursts of 10 seconds and pauses for cooling in between for 45-50 seconds. The cell lysate was centrifuged for 30 minutes at 20.000 rpm at 4 °C in a Beckman JA-20 rotor.

Purification of single or co-expressed proteins was carried out by immobilized metal ion affinity chromatography (IMAC) using chelating Sepharose Fast Flow resin (GE Healthcare Bio Sciences AB, Sweden) charged with Ni²⁺ ions following manufacturer's instructions and equilibrated with lysis buffer. The volume of resin used for purification varied depending on the amount of protein to be purified. Theoretical capacity for this resin is 5 mg of protein per ml of resin (according to manufacturer's instructions). Usually, in order to get the maximum amount of soluble protein, 8 ml of resin was used. Ni²⁺ columns were ran by gravity flow, where the supernatant coming from the centrifugation and the following buffers were charged on the column, without resuspending the resin. After supernatant flow, the column was washed with 8 column volumes of lysis buffer, followed by 10 column volumes of wash buffer, containing 20 mM Tris-HCl pH 7.6, 150 mM NaCl, 50 mM imidazole and 5 mM β-mercapto-ethanol. The protein was eluted using lysis buffer containing 500 mM imidazole. Monitoring of the elution volume (presence of the desired protein) was carried out using the colorimetric method; that is, using the Bradford solution (Biorad, Hercules, CA, USA) to track the change of color from blue (protein present) to brown (protein absent). The elution buffer was added until the color turned brown, in order to obtain the maximum of the soluble protein. Fractions of the flow-through, washing and elution were analyzed on 10-16 % SDS-PAGE gels, depending on the size of the purified protein.

The N-terminal His-tag was removed using TEV protease (27 kDa catalytic domain of the Nuclear Inclusion a protein encoded by the tobacco etch virus, TEV), obtained from the Protein Expression and Purification Facility, EMBL Heidelberg. TEV protease was added in the ratio of 1 mg for 100 mg of the eluted protein and the sample was dialyzed in a 3.5 or 6-8 kDa cut-off Spectra/Por dialysis bag (Spectrum Laboratories, San Pedro, CA, USA) against 2 liters of a buffer

containing 20 mM Tris-HCl pH 7.6, 150 mM NaCl and 5 mM β -mercapto-ethanol, in order to remove completely the imidazole. 2 mM EDTA was added in the dialysis bag, since TEV protease requires chelating agents in order to remove divalent cations that inhibit TEV activity. As a result, dialysis for a buffer exchange and His-tag cleavage take place in the same time, at 4 °C for 12-14 hours.

The dialyzed sample was loaded again on the same volume of Ni^{2+} resin that was equilibrated before with 10 column volumes of dialysis buffer. Since the protein now contains no His-tag it passes through the resin beads without being bound. Extensive washing with dialysis buffer was followed in order to recover the rest of the desired protein. Finally, elution buffer was added to the column. In this step, remnants of the uncut protein, histidine-rich protein contaminants and the TEV protease (it contains a non-removable His-tag) are eluted. The colorimetric method described before was also used to monitor the amount of proteins passing through the column. The fractions collected during the second Ni^{2+} purification were also analyzed on SDS-PAGE gels.

Protein containing fractions were concentrated to ~2-10 mg by centrifugation using Amicon concentrators (Millipore, Billerica, MA, USA) with a suitable molecular weight cut-off. The final purification step includes a size exclusion chromatography (SEC) step. For proteins smaller than ~40 kDa a Superdex 75 10/300 GL column (GE Healthcare Bio Sciences AB) was used, while for bigger proteins a Superdex 200 10/300 GL column was applied. The columns were run on an AKTA FPLC chromatography system (GE Healthcare Bio Sciences AB) and the injection volume was varying from 100 μl to 1 ml. The buffer used for SEC was 20 mM Tris-HCl pH 7.6, 150 mM NaCl and 1 mM DTT. In many cases the protein was very pure and was eluting as a single peak, having the desired molecular weight (MW). In other cases, protein was eluting in the void volume, meaning that the protein forms soluble aggregates. Fractions containing the protein of interest were pulled together and concentrated for crystallization trials or for biochemical experiments, using the Amicon concentrators.

In order to improve the solubility of some protein constructs or eliminate the contamination from nucleic acids, purification (as described above) in high salt (500 mM NaCl) took place. High salt was kept to all purification steps.

2.3 – Biochemical methods

2.3.1 - Western blotting

In a western blot experiments, the protein of desire is detected by means of antibodies. An SDS-PAGE gel is performed and the blot is transferred from the gel onto a PVDF (Polyvinylidenfluorid) sheet. Before blotting, PVDF sheet is soaked in 100% ethanol and in western blot buffer (for 1 L: 1 gr SDS, 3.03 gr Tris, 14.41 gr Glycin, 10 % ethanol). The transfer device is then built (from negative to positive electrode), containing 4 layers of Whatman paper, the gel, the PVDF sheet and 4 more layers of Whatman paper. The samples were blotted for 1 hour at 4 °C at 100 Volts. After the transfer, the PVDF sheet was soaked in 2.5 w/v % milk powder, 0.1 % tween-20 in 1X PBS to reduce unspecific binding of the antibody added in the next step. For anti-his-tag detection, antibody conjugated to horseradish peroxidase (HRP) was added in a dilution 1:3000 in the milk buffer described above and incubated for 1 hour. After antibody incubation, two washing steps of 15 minutes were performed before detection. For direct color-detection on the membrane, SIGMAFAST 3,3-Diaminobenzidine tablets were dissolved in 1 ml water following the protocol and the membrane was stained. The staining was stopped by addition of water.

2.3.2 - Blotting for N-terminal sequencing

For N-terminal sequencing, the protein was transferred from an SDS-PAGE gel onto a PVDF membrane in a 3-(cyclohexylamino)-(propanesulfonic acid) (CAPS) buffer blotting technique. Before blotting, the PVDF sheet was shortly soaked in 100 % methanol and after equilibrated in transfer buffer (10 mM CAPS pH 11, 10 % methanol). In the same time, the SDS gel was washed in water and equilibrated in the same transfer buffer. The protein was blotted as described for the western blot and the protein was detected on the membrane with addition of 0.1 % Coomassie blue R-250 in 40 % methanol and 1 % acetic acid for 1 minute. Excess colorant was removed by repeated washes in 50 % methanol or water. Finally the gel was rinsed with water. Bands of interest were excised from the dried membrane and analyzed by N-terminal sequencing in the Institutte Biologie Structurale, Grenoble (Dr. Jean-Pierre Andrieu).

2.3.3 - Limited proteolysis

Limited proteolysis can determine fragments of a protein or protein complex which are resistant to proteolytic digestion. These fragments are compact globular domains (since inaccessible to the protease) and are likely to crystallize. For limited proteolysis experiments, trypsin was used. Trypsin cleaves peptide chains at the carboxy side of the amino acids lysine or arginine, except when either is followed by proline. Protein solution (0.5-1 mg/ml) was mixed with trypsin in a ratio of 1:1000 and 1:500. Samples were taken at several time points (5, 10, 30, 60 and 120 minutes) and the reaction was stopped by mixing with SDS loading buffer (2X concentration) and heating to 95 °C for 5 minutes. For large-scale purifications of trypsinated protein, the reaction was stopped by the addition of protease inhibitor (Complete, Roche) and the mixture was immediately purified by size exclusion chromatography to remove residual trypsin, or very small degraded protein fragments. In a final step for the identification of the proteolyzed protein products, the sample was sent for Mass Spectrometry (MS) analysis at the Institute Biologie Structurale, Grenoble (Dr. Luca Signori) or the Proteomics Core Facility, EMBL Heidelberg.

2.3.4 - Pull-down assays using recombinant proteins

To test the interaction between two proteins, pull-down assays were performed with one protein having an N-terminal His-tag and the other protein being cleaved. In all the cases proteins were expressed and purified according to the protocol described above. His-tag proteins after the first Ni²⁺ column, were dialyzed in a buffer containing no imidazole, similar to the buffer used in the SEC purification step (20 mM Tris-HCl pH 7.6, 150 mM NaCl, 1 mM DTT). The two proteins of ~100 µg each and Ni-NTA resin (Qiagen) (10 µl beads for each sample) were incubated in a rotating wheel at 4 °C for 3 hours. Before incubation, Ni-NTA beads were washed extensively with the same buffer. After incubation, protein-resin mixture was washed many times with the buffer described above by centrifugation, and a second washing step with 20 mM Tris-HCl pH 7.6, 150 mM NaCl, 50 mM imidazole and 1 mM DTT was applied. As a result, all the unspecifically bound proteins are washed away and the His-tag protein was bound to the resin, along with the protein that interacts with. In a final step, 15 µl of loading buffer (4X

concentration) was added to the beads and the samples were boiled for 5 min at 95 °C. Fractions of the washing steps and the boiled beads (eluates) were analyzed in a 10-16 % SDS-PAGE gel.

2.3.5 - Interaction studies using *in vitro* translation system

Interaction mappings between recombinant and *in vitro* translated proteins were performed using pull-down assay. *In vitro* translation was carried out with TnT Coupled Reticulocyte Lysate System (Promega), which allows coupling of transcription and translation in the same reaction and protein labeling with ³⁵S-Methionine (Perkin-Elmer). 100 µg of recombinant protein (expressed in *E.coli* with N-terminal His-tag, purified as described above) were incubated for 2 hours with one fifth of an *in vitro* translation reaction in a lysis buffer containing 20 mM Tris pH 8, 300 mM NaCl, 0.01 % NP-40 and 20 mM imidazole. The same buffer was used for equilibrating 20 µl of Ni-NTA resin beads (Qiagen). The beads were blocked with 1 % BSA for 30 minutes and incubated with the previously mixed proteins for 2 hours. After thorough washing (lysis buffer with 50 mM imidazole), the beads-protein mixture was transferred to spin columns (20 µm filter, Macherey-Nagel) for further washing by gravity flow. Elution was performed in 500 mM imidazole by spinning the columns for 1 minute, followed by analysis on 10 % SDS-PAGE. The gel was fixed in 45 % methanol and 10 % acetic acid, dried and exposed to a PhosphorImager screen to visualize the ³⁵S-labelled proteins.

For interaction studies with an *in vitro* translated protein of choice and a peptide, the protein was labeled and translated in the same way as described above. One fifth of the *in vitro* translation reaction was incubated for 1 hour on ice with a peptide conjugated to biotin at its C-terminus (Peptide Special Laboratories, Heidelberg, Germany). In parallel, streptavidin-coupled magnetic beads (Promega) were washed with buffer containing PBS, 0.05 % NP-40 using a magnetic rack and blocked with 1 % BSA for 30 minutes. Protein-peptide mixture was added to the magnetic beads and several rounds of washing followed. In this experiment, the biotin-containing peptide will interact strongly with the streptavidin of the magnetic beads and would attract the *in vitro* translated protein of choice. In the last step, 10 µl of SDS-loading buffer (2X concentration) was added to the beads and the samples were boiled at 95 °C for 5 minutes.

Analysis on 10-12 % SDS-PAGE, fixation, dry and exposure of the gel was performed as described above.

2.3.6 - Pull-down assays with endogenous proteins

Mouse adult testis were homogenized in a douncer, in a buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM MgCl₂, 1 mM DTT, 0,5 % Na-deoxycholate, 1% Triton X-100, Complete (EDTA-free) Protease Inhibitor Cocktail (Roche), 50 µg/ml tRNA, 2 mM Vanadyl RC (Sigma), 10 % Glycerol and sterilized water up to the final volume of 1 ml. The lysate was cleared by centrifugation at 9200 g, 4 °C for 5 minutes. 100 µg of recombinant His-tag protein (purified in *E.coli* as described above) were incubated with the lysate for 2 hours at 4 °C. In the meantime, Ni-NTA (Qiagen) resin (10 µl beads for each sample) was washed several times in a wash buffer containing 10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM MgCl₂, 0.0.1 % NP-40 and blocked with 1 % BSA for 30 minutes. After that, beads and protein-lysate mixture were incubated at 4 °C for 3 hours. Extensive washing by centrifugation with the wash buffer followed, and then samples were further washed on filter-containing columns (Macherey-Nagel, Duren, Germany) with wash buffer 1 (20 mM Tris-HCl pH 8, 150 mM NaCl, 5 mM β-mercaprotethanol) and wash buffer 2 (20 mM Tris-HCl pH 8, 150 mM NaCl, 50 mM imidazole, 5 mM β-mercaprotethanol) by gravity flow. In a final step, 50 µl of elution buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 500 mM imidazole, 5 mM β-mercaprotethanol) was added to the beads on the columns and eluates were collected by centrifugation. 12.5 µl of SDS-loading buffer (4X concentration) was added to the samples, that were boiled at 95 °C for 5 minutes. Samples were analyzed on 10 % SDS-PAGE gel and prepared for western blotting. For that, the SDS-PAGE was transferred into a nitrocellulose membrane using semidry blotting equipment (Bio-rad). The membrane was blocked with 5 % w/v milk in a buffer containing PBS and 1 % Tween-20 (western buffer) for 1 hour and incubated for 2-3 hours with the primary antibody in 5 % milk western buffer. Dilutions of the antibody ranged from 1:100 to 1:1000 depending on the signal obtained. After extensive washes with western buffer, the membrane was incubated with a 1:10.000 dilution of anti-mouse or anti-rabbit antibodies coupled to HorseRadish Peroxidase (HRP, GE Healthcare) for 1 hour, followed by three washes. Membranes were developed using ECL+ (GE Healthcare)

and chemiluminescence films (GE Healthcare). The antibodies that were used were: Mili, Miwi (Reuter et al., 2009), MVH (Abcam), Y12 (Millipore), aSYM11 (Millipore).

2.3.7 - Isothermal Titration Calorimetry

Isothermal titration calorimetry measurements were performed in duplicates or triplicates at 25 °C, using an ITC200 Microcalorimeter (MicroCal Inc). Experiments included injection of 40-50 μ l of peptide solution (3-10 mM) into a sample containing 60-90 μ M of wild type protein or its mutants, in 25 mM Tris-HCl pH 7.6, 150 mM NaCl and 5 mM β -mercaptoethanol (or 1 mM DTT) buffer. All peptides (Peptide Specialty Laboratories GmbH) used for ITC experiments were dissolved or dialyzed into the same buffer as the protein. In a final step before each titration the pH of the dissolved peptide was adjusted to the pH of the buffer. Phenylalanine-containing peptide concentrations were estimated with absorbance spectroscopy using the extinction coefficient $\epsilon_{257} = 195 \text{ M}^{-1} \text{ cm}^{-1}$. Otherwise, peptide concentrations were estimated from the mass. A total of 26 injections were made with a spacing of 180 seconds and a reference power of 5 μ cal/s. Control experiments were performed under identical conditions to determine the heat signals that arise from injecting the peptide into the buffer. Binding isotherms were plotted and analyzed by Lavenberg-Marquardt nonlinear regression using Origin Software version 7.0 (MicroCal Inc.). The initial data point was routinely deleted. The ITC data were fit to a one-site binding model, or in the case of tandem Tudor domains, the sequential binding site model, using software provided by MicroCal (Turnbull and Daranas, 2003; Wiseman et al., 1989).

All the peptides used during this work were ordered and synthesized from Peptide Specialty Laboratories GmbH (Heidelberg, Germany). They were purified by reverse-phase HPLC and the peptide identity was confirmed by MALDI-TOF mass spectrometry. All the peptides were containing TFA (Trifluoroacetic acid) and for that reason after resuspension to the buffer of choice, pH was adjusted for biochemical and crystallographic experiments.

2.4 - Small Angle X-ray Scattering (SAXS)

Scattering techniques like SAXS, enable structural analysis of macromolecular particles in homogeneous solutions. They analyze the molecular shape, quaternary and tertiary structure of molecules and the resulting low-resolution envelopes can be combined with known high-resolution structures. In contrast to electron microscopy, the method can be used for smaller macromolecules and is therefore complementary.

For the experiment, the sample solution is exposed to the radiation and the scattered intensity detected is dependent on the scattering angle. The scattering of the buffer only is subtracted as background. The random positions and orientations of particles result in an isotropic intensity distribution which, for mono-disperse non-interacting particles, is proportional to the scattering from a single particle averaged over all orientations. Since distinct shapes (globular, disc-like, elongated) have characteristic scattering curves, a particle fitting to the resulting scattering curve is simulated. Several fits are combined to a model of the molecular envelope of the molecule.

2.4.1 - Sample preparation, SAXS data collection and analysis

Protein samples were expressed and purified as described above. In the last purification step Superdex 200 10/300 GL column was used and 5 mM DTT was added before SAXS measurement.

Data were collected at the ESRF BioSAXS station, ID14EH3 (Pernot et al., 2010). ID14-3 is a fixed energy (13.32 keV, $\lambda = 0.931 \text{ \AA}$) station optimized for solution scattering from biological macromolecules. Samples were exposed using 30 μl of protein solution loaded into a 2 mm quartz capillary mounted in vacuum using an automated robotic system, which enables the sample to pass through the beam during exposure to minimize the effect of radiation damage. A range of protein concentrations (10, 4.5 and 1.5 mg/ml in the presented results) were measured to assess and account for inter particle effects. 2D scattering images were collected using Vantec2000 detector (Bruker) 1.83 m from the sample. Data collection used for all data was 10 frames each 30 seconds in duration. Individual time frames are processed automatically

and independently by the data collection software (BsxCUBE) developed at the ESRF, yielding individual radically averaged curves of normalized intensity versus scattering angle ($s = 4\pi\sin\theta/\lambda$ in nm). Time frames are combined excluding any data points affected by aggregation induced by radiation damage to give the average scattering curve for each measurement. The scattering from the buffer alone was measured before and after each sample measurement and the average of the scattering before and after each sample is used for background subtraction using the program PRIMUS (Konarev et al., 2003) from the ATSAS package developed by EMBL Hamburg. For both the TD1-4 and TD1-2, forty *ab-initio* models were created with DAMMIF (Franke and Svergun, 2009) which were averaged and filtered using DAMAVER (Volkov and Svergun, 2003). The output of DAMAVER was refined using it as a starting model for the program DAMMIN to yield the final calculated envelope. Rigid body models were produced with CORAL (Petoukhov and Svergun, 2005), which allows the addition of missing linkers using a database of random self avoiding chains. Fits of TD1-2 and TD3-4 to TD1-2 data were made with Crysol 27 (Svergun et al., 1995) while Oligomer (Konarev et al., 2003) was used to fit the linear combination. The plot of the fits was produced with SAXS view (saxsview.sourceforge.net).

2.5 – X-ray crystallography

2.5.1 - Protein crystallization and crystal freezing

The protein sample was concentrated to three different concentrations: low (~5-8 mg/ml), middle (10-15 mg/ml) and high (20 mg/ml or higher). Crystallization trials were performed at the HTX-laboratory of the Partnership for Structural Biology (PSB) in Grenoble with a Cartesian robot, which makes 100 nl + 100 nl drops of a buffer, using the sitting drop method. The robot allows the screening of 576 different conditions using 100 µl of protein. Images of the drops were taken by a Robodesign Minstrel III (Rigaku Msc, U.K) imaging system at 20 °C or 4 °C, and a new set of images were obtained after 24 hours, 72 hours, 7 days, 16 days, 5 weeks and 10 weeks. Positive hits were reproduced and refined using either the hanging drop vapor diffusion or the sitting drop technique in a robot plate (described above) or in 6 X 4 plates, using 0.5-2 µl of protein mixed with 0.5-2 µl of precipitant. A concentration gradient matrix of the precipitant components was set in the plates in order to find the best condition for crystal growth. The

wells of the plates were sealed with cover glasses by means of high viscous silicone paste or with transparent film.

During X-ray exposure, crystals suffer from radiation damage as free radicals form. To minimize this effect, the crystals are typically cooled down at low temperatures during the experiment. Since the formation of ice crystals would destroy the protein crystal, this is prevented by a cryo-protection before flash-freezing the crystals in liquid nitrogen. For freezing, loops of the appropriate size (50-500 μM) were used to fish the crystals that afterwards were transferred for a few seconds in a solution containing the crystallization condition and a cryoprotectant; in some cases different cryoprotectants were tried out including ethylene glycol and PEG (in different molecular weights), but 20-30 % glycerol proved to be the more efficient and suitable chemical in almost all cases. Finally, crystals were then snap-frozen into liquid nitrogen at -196 °C.

2.5.2 – Data collection

Al X-ray diffraction experiments were carried out at the European Synchrotron Facility (ESRF) in Grenoble, France at ID14 EH-4, ID23 and ID29 beamlines. A data collection strategy was designed by analyzing several diffraction images with the program MOSFLM (Leslie et al., 2002).

2.5.3 – Data processing

Data were indexed and integrated with XDS (Kabsch, 1993). Indexing determines the relative orientation of the unit cell and the crystal symmetry of the crystal from collected data. Subsequently, the data were integrated by measuring the difference in intensity between the spots and the local background. Then, the data were merged and scaled with XSCALE (Kabsch, 1993). Structure factor amplitudes were calculated from merged intensities by the program TRUNCATE implemented in CCP4. A big challenge in X-ray crystallography is the phase problem. In a diffraction experiment, the scattered waves from the crystal planes contain all information about the structure. The amplitudes are measured in the experiment, but the phase information is lost. To solve the phase problem, heavy atoms can be introduced into the protein molecule, as the more electrons an atom possesses, the more it scatters X-rays. Small

differences in reflection intensity between native and derivative can be used to first locate the heavy atoms and then perform phasing (isomorphous replacement). Anomalous scattering occurs when atoms not only scatter photons but also absorb energy due to energy transition of electrons. The required energy is characteristic for the chemical element and nature of transition. At wavelengths close to an absorption edge the atomic scattering factor includes an anomalous component and the phase of the emitted photon is altered, which can be used to obtain the phases. The positions of the anomalous components were identified with SHARP (de La Fortelle et al., 1997). The second way to find phases is molecular replacement (Rossmann, 1990). In the molecular replacement method, a known structure with a certain structural similarity (at least 25 % sequence identity as a rule of thumb) is used to obtain the phases. Molecular replacement was performed with PHASER, implemented in CCP4. The Matthews coefficient was calculated to determine the number of molecules in the asymmetric unit by estimating the solvent content (Matthews, 1968).

2.5.4 - Refinement and structure analysis

As a better model gives back better phases, several rounds of refinement have to be performed starting with the initial model. For this, the structure was inspected and improved, using the program RESOLVE and the molecular graphics program COOT (Emsley and Cowtan, 2004). After a cycle of building, a refinement program was run to minimize the R-factor. The programs REFMAC (Murshudov et al., 1997) and ARP/WARP (Langer et al., 2008) were used for refinement. For the calculation of R_{free} , a test set of 5 % of the reflections was excluded from refinement to circumvent model bias (Brunger, 1993). The program ARP/WARP was used to build water molecules. A number of further programs and software were used for analysis of the structure. PYMOL (De Lano scientific) was used to create figures, CLUSTALX and CLUSTALW (Larkin et al., 2007) were used for sequence alignment. For geometry check of the structure the MOLPROBITY server was used (Chen et al., 2010).

3 – RESULTS

Summary of chapter “Results”

The data presented in this thesis are involving the structural and biochemical characterization of the Tudor domains (TDs) and the N-terminal MYND domain of TDRD1 protein. Given that multiple sDMA residues in various sequence contexts decorate the N-terminus of Mili, distinct specificities of the individual Tudor domains might drive complex formation.

We identified soluble and well-expressed protein fragments of multiple and single extended Tudor domains (eTud) and we examined their specificities for different sDMA-containing Mili peptides. We observed a binding preference of isolated TD2 and TD3 to consecutive peptides, R45me2 and R74me2 respectively. For that reason we examined whether TDRD1's tandem TD2-3 can bind more strongly to a longer peptide combining both these methylation marks. Isothermal Titration Calorimetry (ITC) measurements revealed no enhanced affinity suggesting for no-cooperativity. Moreover, we explored the binding capacity of single eTuds of TDRD1 to isolated symmetrically dimethylated (sDMAs), asymmetrically dimethylated (aDMAs) and monomethylated (MMAs) arginines. Our results indicate that the Tudor domains 2,3 and 4 of TDRD1 can bind strongly sDMAs and to a much less extent aDMAs and MMAs. Interestingly, we observed that a single amino acid substitution of the sDMA binding residues was responsible for the low binding affinity of TD1. This is also confirmed by pull-down assays from adult mouse testis, suggesting that TD1 is an “inactive” sDMA-binder.

Crystallization trials of single and tandem TDs with Mili's methylated peptides resulted in obtaining the structure of TD3-R45me2 complex, revealing a unique orientation of the peptide as a whole and also of the sDMA in the aromatic cage. Finally, Small Angle X-ray Scattering (SAXS) experiments of the four tandem Tudor domains of TDRD1 reveal a flexible, elongated entity lending support to a molecular scaffolding role for TDRD1 in the piRNA pathway.

At the last section, preliminary results are presented for the N-terminal MYND domain of TDRD1. MYND domains are protein-protein interaction modules, coordinated by zinc atoms. Using pull-down assays, we show that it interacts specifically with the inactive isomerase domain of the novel piRNA associated factor FK506-binding protein FKBP6.

Resume du chapitre “Resultats”

Les données présentées dans cette thèse concernent la caractérisation structurale et biochimique des domaines Tudor (TDs) et du domaine N-terminal MYND de la protéine TDRD1. Etant donné que des multiples résidus sDMA dans des différents contextes de séquences existent au N-terminal de Mili, des spécificités distinctes des domaines individuels Tudor pourraient conduire à la formation des complexes.

Nous avons identifié des fractions protéiques solubles et bien exprimées des domaines Tudor soit multiples soit individuellement étendus (eTud) et nous avons examiné leurs spécificités pour différents peptides Mili qui contiennent sDMA. Nous avons observé une préférence de liaison des TD2 et TD3 pour des peptides consécutifs R45me2 et R74me2 respectivement. Pour cela, nous avons examiné si le tandem de TDRD1 le TD2-3 peut se lier de façon plus forte à un peptide plus long qui combine tous ces deux marques de méthylation. Les mesures Isothermal Titration Calorimetry (ITC) n’ont pas montré une affinité plus forte, un résultat qui suggère non-coopérativité. En outre, nous avons exploré la capacité de liaison des eTuds de TDRD1 individuels pour sDMAs individuelles, des DMAs asymétriques (aDMAs) et des arginines monométhylées (MMAs). Nos résultats indiquent que les domaines Tudor 2,3 et 4 de TDRD1 peuvent se lier fortement aux sDMA et un peu moins aux aDMAs et MMAs. En fait, nous avons observé que la faible affinité de liaison de TD1 est associée à une substitution d’une seule acide amine des résidus responsables de liaison des sDMA. Ce résultat a été aussi confirmé par des essais pull-down des testicules de la souris adulte, indiquant que TD1 est inactive sDMA-binder.

Les essais de cristallisation des TD individuels et tandem avec des peptides méthylés Mili ont conduit à l’obtention de la structure du complexe TD3-R45me2 qui a révélé une orientation unique du peptide dans sa totalité et aussi du sDMA dans la cage aromatique. Finalement, des expériences de Small Angle X-ray Scattering (SAXS) des quatre domaines Tudor de TDRD1 tandem ont révélé une entité flexible et élongée qui soutient le rôle moléculaire d’échafaudage de TDRD1 dans la voie de l’ARNpi.

A la dernière section, nous présentons des résultats préliminaires pour le N-terminal du domaine MYND de TDRD1. Les domaines MYND sont des molécules d’interaction protéine-protéine coordonnées par des atomes de zinc. En réalisant des essais pull-down, nous avons montré que le MYND interagit spécifiquement avec le domaine d’isomérase inactive du nouveau facteur FK506-binding protein FKBP6 associé à l’ARNpi.

3.1 – Structural and functional studies on TDRD1- a molecular scaffold in the piRNA pathway

3.1.1 - Expression, purification and crystallization trials of Tandem eTudor domains of TDRD1

Tudor domain containing protein 1 (TDRD1) is a ~130 kDa multi-domain protein expressed in the germline, and has been characterized by various research groups as a piRNA pathway associated factor, indispensable for the normal progression of spermatogenesis (Chuma et al., 2006; Reuter et al., 2009; Wang et al., 2009). Mouse TDRD1 is comprised of an N-terminal MYND domain (myeloid translocation protein 8, Nery, and DEAF-1) and four tandem Tudor domains ([fig.3.1.1.A](#)). As a basis for structural and biochemical studies of the TDRD1 protein, the first trials were made on constructs comprising all four Tudor domains of TDRD1. The experimental protocol followed was including expression of protein constructs in *E.coli* with a N-terminal His₆-tag and purification in Ni²⁺ resin column, TEV cleavage, a second Ni²⁺ resin column to remove protein contaminants and a size-exclusion chromatography for further purity of the sample (see [materials and methods](#)).

All the constructs comprising the four Tudor domains (TDs) of TDRD1 were designed according to secondary structure prediction bioinformatic tools (NPS@, Combet et al., 2000), multiple sequence alignments, disorder prediction programs (RONN, Yang et al., 2005) and preliminary construct design performed in Ramesh Pillai's group (EMBL, Grenoble Outstation). The protein samples were soluble and expressed in high yields. Size exclusion chromatography on S200 column showed that the proteins are monomeric in solution ([fig 3.1.1.B](#)). Unfortunately, all tested protein samples were unstable and showed considerable degradation in all the purification steps ([fig. 3.1.1.C](#)). The most stable constructs among them (TD1-4, residues 232 to 1094; 1-4#4, residues 232 to 1113) were concentrated to 5, 10 and 15 mg/ml and sent for crystallization trials at HTX-laboratory at 20 °C and 4 °C, in order to prevent further degradation of the protein. Crystallization trials gave no crystal hits and the sample TD1-4 was subjected to

limited proteolysis with trypsin (see [materials and methods](#)) in order to identify more stable constructs of tandem or single Tudor domains ([fig.3.1.2.A](#)).

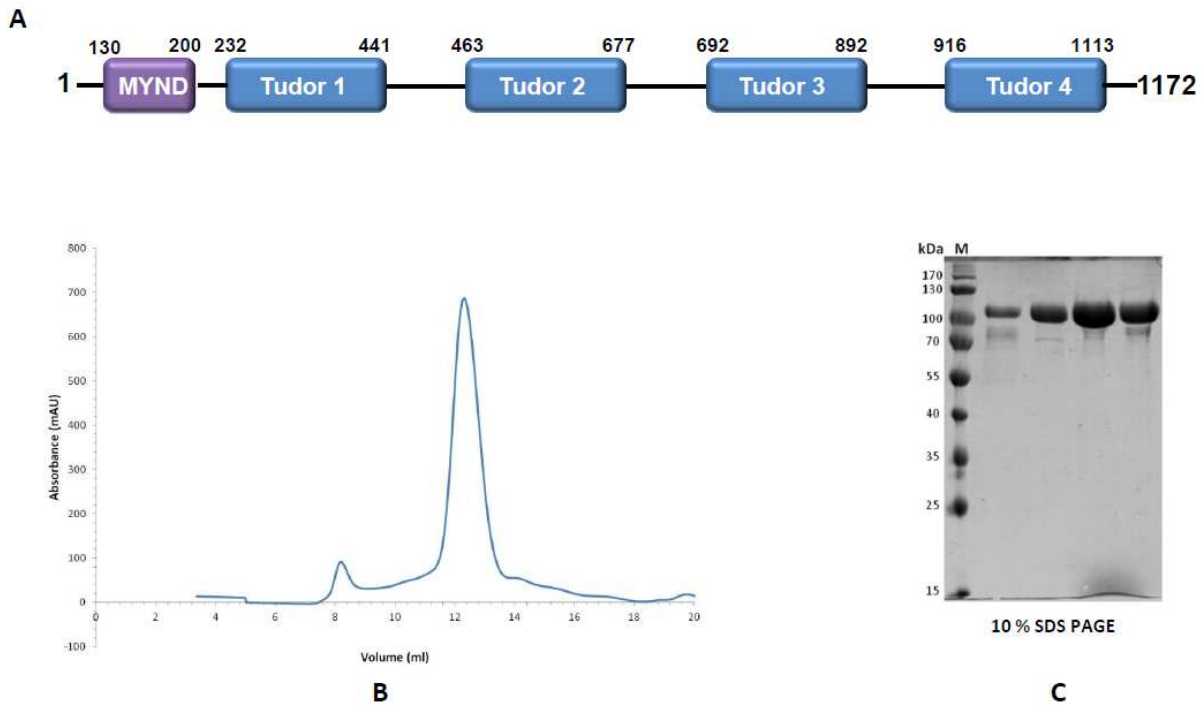


Figure 3.1.1. Purification of tandem Tudor domains TD1-4 (232-1094 aa). **A.** Schematic representation of mouse TDRD1. TDRD1 has an N-terminal MYND domain and four tandem extended Tudor domains. **B.** Size exclusion chromatography profile on a S200 gel filtration column (Blue line: Absorbance at 280 nm). **C.** SDS-PAGE of the peak fractions. Size marker in kilodaltons (kDa) is indicated.

Various samples were analyzed by Mass spectrometry (MS) and N-terminal sequencing and new constructs comprising the first two tandem Tudor domains were designed ([fig.3.1.3](#)). However, MS analysis could not identify polypeptides corresponding to other Tandem Tudor domain combination or single domains, since they were rapidly degraded over time. For that reason, using prediction programs, various protein fragments were obtained soluble harboring tandem Tudor domains; nevertheless, all of them were either degraded over time (see section [3.1.2](#)) or were not sufficiently homogeneous during purification steps ([fig.3.1.3](#)).

Protein constructs encompassing the Tandem Tudor domains 1 and 2 were expressed and purified to homogeneity in high yields. The best-behaved construct, TD1-2 (residues 232 to 677) ([fig.3.1.2.A-B](#)) was subjected to crystallization trials at HTX laboratory at 20 °C at three

concentrations: 18 mg/ml, 12 mg/ml and 7 mg/ml. After 7 days, micro crystals grew in various conditions. However, trials for manual reproduction of the crystals using the hanging drop diffusion method and sitting drop method failed.

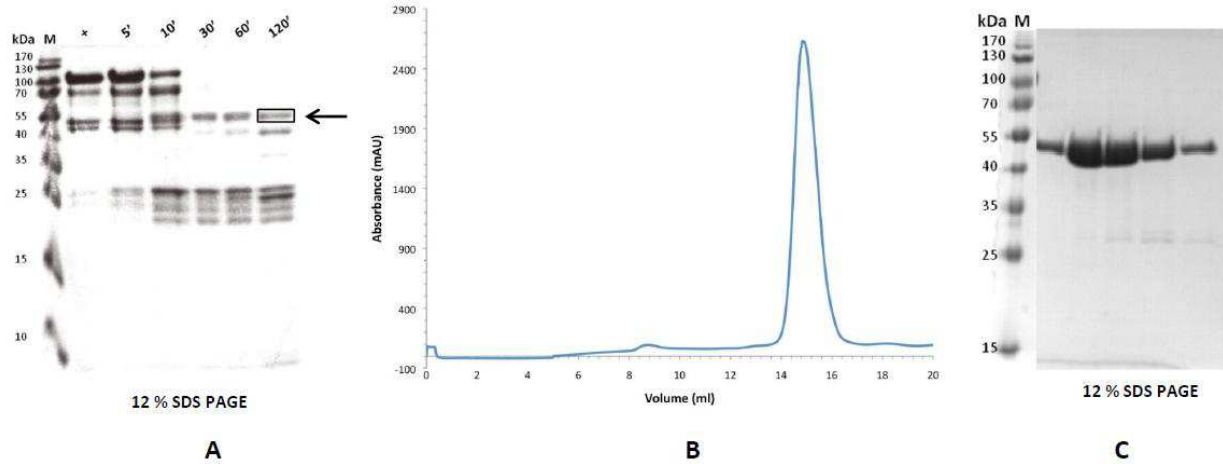


Figure 3.1.2. Purification of tandem Tudor domains TD1-2 (232-677 aa). **A.** Tryptic digest of TD1-4. Trypsin was added in a 1:1000 ratio and samples were taken at the indicated time points. One fragment (black box, arrow) was identified using N-terminal sequencing and Mass Spectrometry and corresponds to TD1-2. **B.** Size exclusion chromatography profile of TD1-2 on a S200 gel filtration column (Blue line: Absorbance at 280 nm). **C.** SDS-PAGE of the peak fractions. Size marker in kilodaltons (kDa) is indicated.






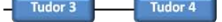
Construct	Protein domains	Amino acids	Solubility	Crystallization robot
1-4#1 (TD1-4)		232-1094	Soluble (degradation)	No hits
1-4#2		232-1155	Soluble (degradation)	No hits
1-4#3		232-1172	Soluble (degradation)	No hits
1-4#4 *		232-1113	Soluble (degradation)	No hits
1-2-3#1 *		232-917	Soluble	
2-3-4#1		476-1050	Soluble	
2-3-4#2 *		463-1113	Soluble	
1-2#1		232-750	Soluble	
1-2#2 (TD1-2) *		232-677	Soluble	microcrystals
1-2#3		245-750	Soluble	
1-2#4		245-677	Insoluble	
2-3#1		463-917	Soluble	No hits
2-3#2 (TD2-3) *		463-892	Soluble	No hits
3-4#1		575-1050	Insoluble	
3-4#2		692-1050	Soluble	
3-4#3		692-1155	Soluble	No hits
3-4#4		692-1172	Soluble	No hits
3-4#5 (TD3-4) *		692-1113	Soluble	No hits
3-4#6		717-1113	Insoluble	
3-4#7		717-1094	Insoluble	
3-4#8		717-1050	insoluble	

Figure 3.1.3. Constructs comprised of tandem Tudor domains of TDRD1. In this table are listed the various protein fragments expressed, purified and sent for crystallization trials. The constructs with an asterisk were expressed after the identification of single eTud domains, while in parentheses are denoted the protein fragments used for biochemical experiments described later.

3.1.2 - Identification of single eTudor domains

The unsuccessful trials to crystallize tandem Tudor domains of TDRD1 led us to try to identify the single Tudor domains of the protein. Trypsin was added to the purified protein TD1-2 at 1:1000 weight ratio, the reaction was left at room temperature, stopped at different time points and loaded on a 12% SDS-PAGE gel. After a 2 hour digestion, four bands show to be resistant ([fig.3.1.4.A](#)). Using Mass Spectrometry and N-terminal sequencing we identified single Tudor domains 1 and 2. Constructs expressed in high amounts were TD1 (232-476 aa) and TD2 (463-677 aa). Both constructs were monomeric in solution, as seen from their corresponding size exclusion chromatography profile ([fig.3.1.4.B-E](#)). TD1 and TD2 were concentrated to 21 mg/ml, 12 mg/ml, 5 mg/ml and 36.3 mg/ml, 25 mg/ml, 10 mg/ml respectively and sent for

crystallization trials at 20 °C at the HTX robot facility. After several weeks of inspection, no positive hits were observed. Finally, based on the results of the tryptic digest, sequence alignment, secondary structure and disorder predictions, several other constructs listed in [fig.3.1.5](#) were expressed with no crystallization success.

One of the identified protein fragments after limited proteolysis had a size of ~15 kDa ([fig.3.1.4.A](#)), similar to the already known prototypic Tudor domain structure of SMN protein (Selenko et al., 2001, PDB 1G5V) and the Tudor domain of TDRD2, the latter being involved in the piRNA pathway (Chen et al., 2009, PDB 3FDR). Remarkably, this Tudor protein fragment corresponding to Tudor domain 2 did not yield soluble protein, while protein constructs having similar size to TD1 and TD2 (22-25 kDa) were soluble and expressed in high amounts ([fig.3.1.5](#)). In addition, similar trials of expressing prototypic Tudor domains 3 and 4 of TDRD1 encompassing ~100 residues resulted in insolubility. Thus, Tudor domains of TDRD1, in terms of molecular weight, resemble more the extended Tudor-SN domain (Friberg et al., 2009, PDB 2WAC) of approximately the same size than the single Tudor domain of SMN protein, suggesting that apart from the prototypic Tudor domain they contain an additional folded part indispensable for solubility.

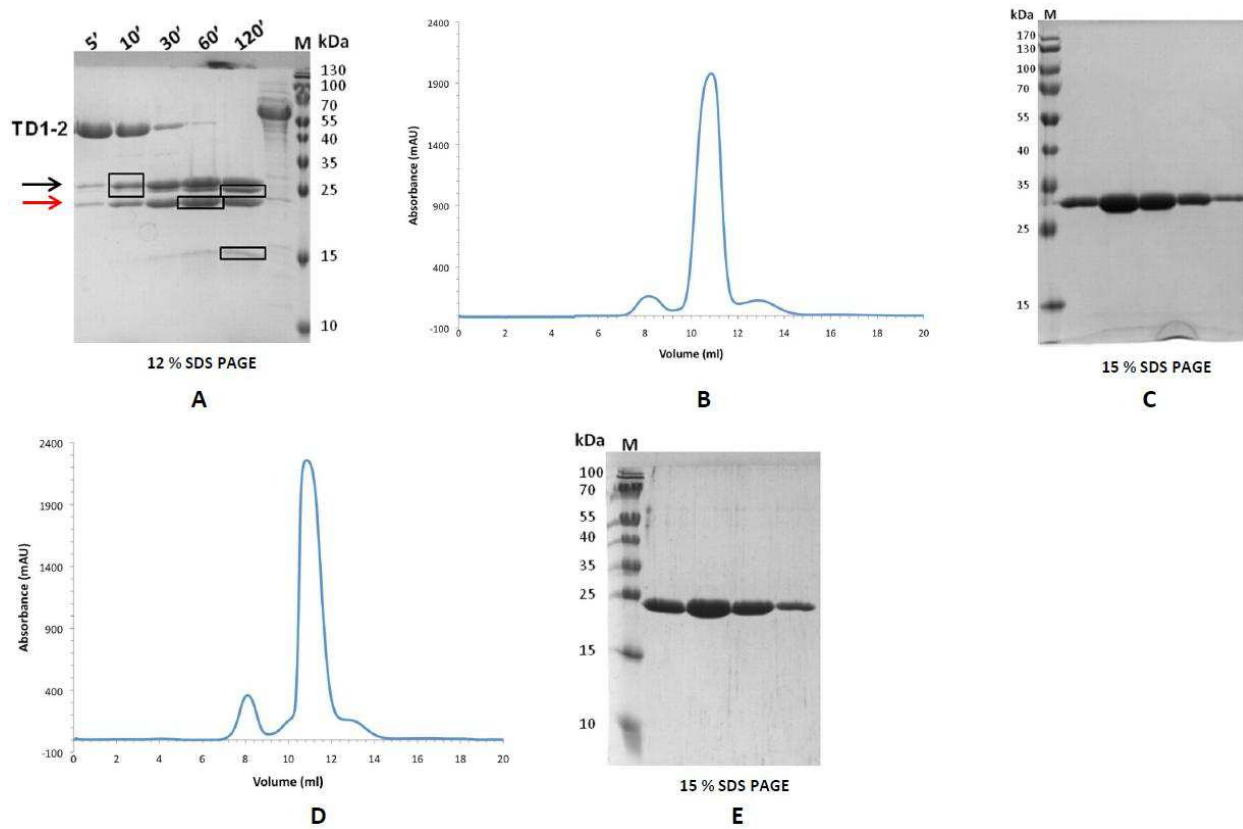


Figure 3.1.4. Purification of single eTudor domains TD1 (232-476 aa) and TD2 (463-677 aa). **A.** Tryptic digest of TD1-2. Trypsin was added in a 1:1000 ratio and samples were taken at the indicated time points. Four samples (black boxes) were identified using Mass Spectrometry and N-terminal sequencing. TD1 (black arrow) and TD2 (red arrow) were stable and homogeneous after purification. **B.** Size exclusion chromatography profile of TD1 on a S75 gel filtration column (Blue line: Absorbance at 280 nm). **C.** SDS-PAGE of the peak fractions. **D.** Size exclusion chromatography profile of TD2 on a S75 gel filtration column (Blue line: Absorbance at 280 nm). **E.** SDS-PAGE of the peak fractions. Size marker in kilodaltons (kDa) is indicated.

Construct	Amino acids	Solubility	Crystallization robot
1#1	217-476	Soluble	No hits
1#2 (TD1) *	232-476	Soluble	No hits
1#3 *	232-439	Soluble aggregates	
1#4 *	232-463	Soluble	No hits
1#5 *	232-468	Soluble	No hits
1#6	240-463	Soluble	No hits
1#7	240-458	Soluble	No hits
1#8 *	240-476	Soluble	No hits
1#9	240-468	Soluble	No hits
1#10	240-471	Soluble	No hits
1#11	255-439	Insoluble	
1#12	258-476	Soluble	
1#13	258-377	Soluble aggregates	
1#14	258-360	Soluble aggregates	
1#15	262-360	Soluble	
1#16	262-476	Soluble	
1#17	305-387	Insoluble	

Construct	Amino acids	Solubility	Crystallization robot
2#1 *	458-677	Soluble	No hits
2#2 *	463-686	Soluble	No hits
2#3 *	463-692	Soluble	No hits
2#4 *	463-697	Soluble	
2#5 *	463-666	Soluble	
2#6 *	463-610	Insoluble	
2#7	463-587	Insoluble	
2#8 (TD2) *	463-677	Soluble	No hits
2#9	477-610	Insoluble	
2#10 *	477-677	Soluble	No hits
2#11	488-666	Soluble	No hits
2#12	488-686	Soluble	
2#13 *	488-677	Insoluble	
2#14	488-587	Insoluble	
2#15	488-610	Soluble aggregates	
2#16	532-677	Soluble	No hits
2#17 *	531-677	soluble	No hits
2#18	540-677	insoluble	

Figure 3.1.5. Constructs comprised of single Tudor domains 1 and 2 of TDRD1. In this table are listed the various protein fragments expressed, purified and sent for crystallization trials. The constructs with an asterisk were expressed after the identification of single eTud domains, while in parentheses are denoted the protein fragments used for biochemical experiments described later.

The identification of the molecular weight that single extended Tudor domains (eTud) are soluble prompted us to try the expression of various constructs of approximately the same size. However, all of the trials resulted in either low solubility or in homogeneous samples that did not crystallize (fig.3.1.5). In parallel, we decided to clone constructs corresponding to the tandem Tudor domains 3 and 4, in order to further characterize their corresponding single Tudor domains. One of the tandem protein fragments, 3-4#2 (692-1050 aa, fig.3.1.3) was expressed in high yields and during purification was degraded into two stable protein parts. These fragments were degraded from their C-terminal region and correspond to the N-terminal regions of the protein since they are retained by Ni²⁺ resin during the purification (they carry an N-terminal His-tag). The presence of the N-terminal His-tag was confirmed by anti-His Western blot (data not shown). Limited proteolysis using 1:1000 weight ratio of trypsin and subsequent

mass spectrometry analysis identified two stable fragments corresponding to eTudor domain 3, while eTudor domain 4 was completely digested without forming a discrete band (fig.3.1.6.A).

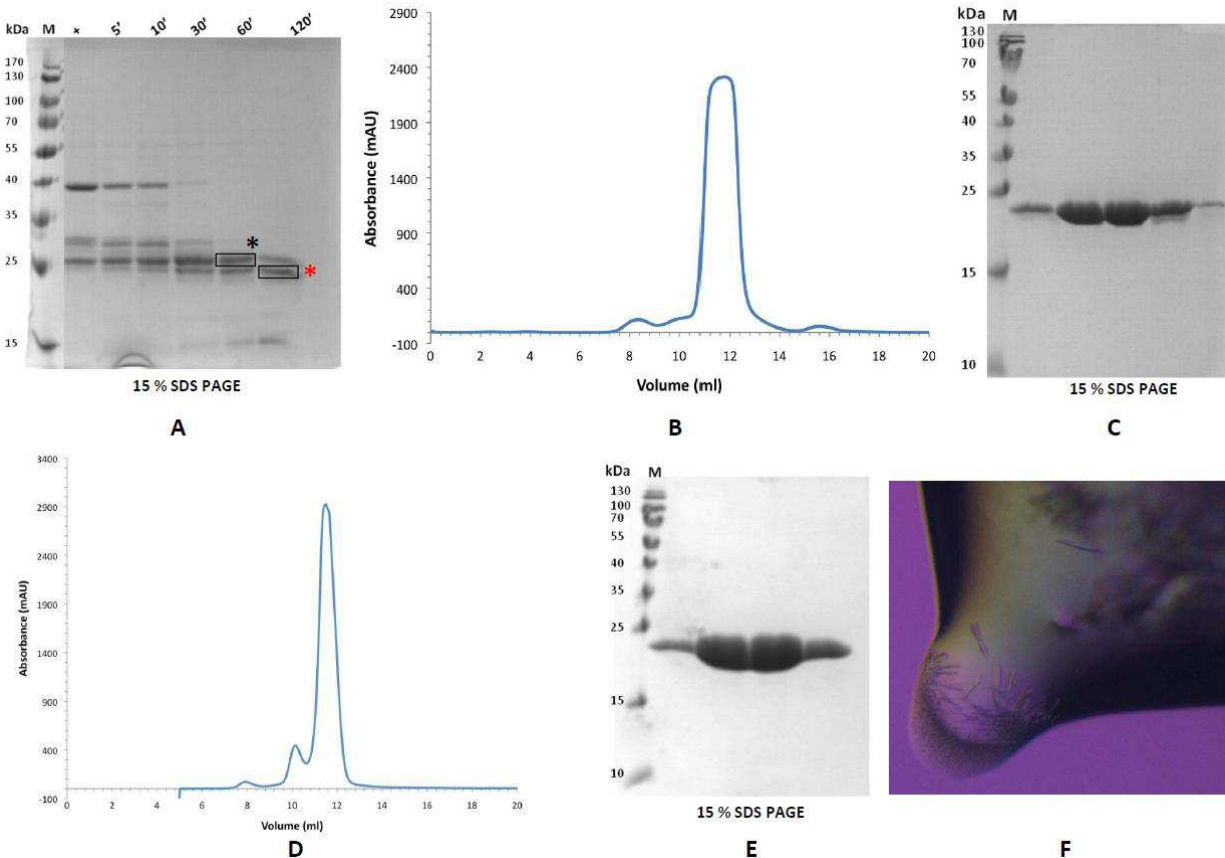


Figure 3.1.6. Purification of single eTudor domains TD3 (692-892 aa) and TD4 (916-1113 aa). **A.** Tryptic digest of the construct 3-4#2 (692-1050 aa). Trypsin was added in a 1:1000 ratio and samples were taken at the indicated time points. Two samples (black boxes) were identified using Mass Spectrometry. TD3 (black asterisk) is corresponding to apo-TD3 (692-917), which is further digested to TD3 (692-892 aa) (red asterisk). **B.** Size exclusion chromatography profile of TD3 on a S75 gel filtration column (Blue line: Absorbance at 280 nm). **C.** SDS-PAGE of the peak fractions. **D.** Size exclusion chromatography profile of TD4 on a S75 gel filtration column (Blue line: Absorbance at 280 nm). **E.** SDS-PAGE of the peak fractions. TD4 was identified after secondary structure prediction analysis. Size marker in kilodaltons (kDa) is indicated. **F.** Crystals of TD4.

The first one was TD3 (692-892 aa) that was soluble and stable throughout the purification (fig.3.1.6.B-C); it was sent for crystallization trials at the High-Throughput Robot facility at three

concentrations, 20 °C: 20 mg/ml, 15 mg/ml and 5 mg/ml. After 7 days, microcrystal plates growing out of sea urchin-like crystals appeared at condition 0.01 M magnesium chloride hexahydrate, 0.05 M HEPES-Na pH 7.0 and 1.6 M ammonium sulfate at protein concentration 15 mg/ml (fig.3.1.7.A). Using vapor diffusion in hanging drops, protein sample was mixed with reservoir solution in 1:1 ratio and crystals were reproduced in bigger size at various conditions. Crystals were briefly soaked in a solution containing the crystallization condition and different concentrations (from 20% to 30%) of cryoprotectant; different cryoprotectants were used, including glycerol, ethylene glycol and PEG200. During diffraction experiments the highest quality crystals proved to be the ones grown in 0.05 M HEPES-Na pH=7, 0.15 M magnesium chloride and 1.25 M ammonium sulfate. The best data set was collected at ID14-EH4 beamline (ESRF) at 3.4 Å resolution and processed as described in [Materials and Methods](#) (fig.3.1.7.B). For solving the structure, Molecular Replacement was carried out with PHASER (McCoy et al., 2005) using the closest protein homologue, protein Tudor-SN from *Drosophila Melanogaster* (PDB 2WAC) as a search model, but no solutions were obtained. To phase the structure, a crystal grown under native conditions was transferred in the same solution containing 2mM K₂PtCl₄ and soaked from two to eight hours before freezing. A heavy atom-derivitized data set was collected at the same beamline. The combination of incomplete incorporation, presence of native protein, small crystals and a low resolution of approximately 4.2 Å resolution did not allow obtaining phase information. For that reason, TD3 was expressed as Se-Methionine-substituted protein, purified identically to the native form and sent for crystallization trials at 20 °C at the HTX laboratory. After ~10 days, microcrystal plates appeared in a condition containing 0.1 M Tris-HCl pH=8.5, 12 % w/v glycerol and 1.5 M ammonium sulfate and were grown in bigger size after manual reproduction using the hanging-drop method (fig.3.1.7.C). Datasets with resolution of 3.8 Å resolution were collected at the inflection and peak wavelengths of the Se K-edge (fig.3.1.7.D) at ID29 beamline (ESRF) and the incorporation of SeMethionine was confirmed by MS (fig.3.1.7.E). Unfortunately, due to low resolution of the data, trials to solve to structure failed.

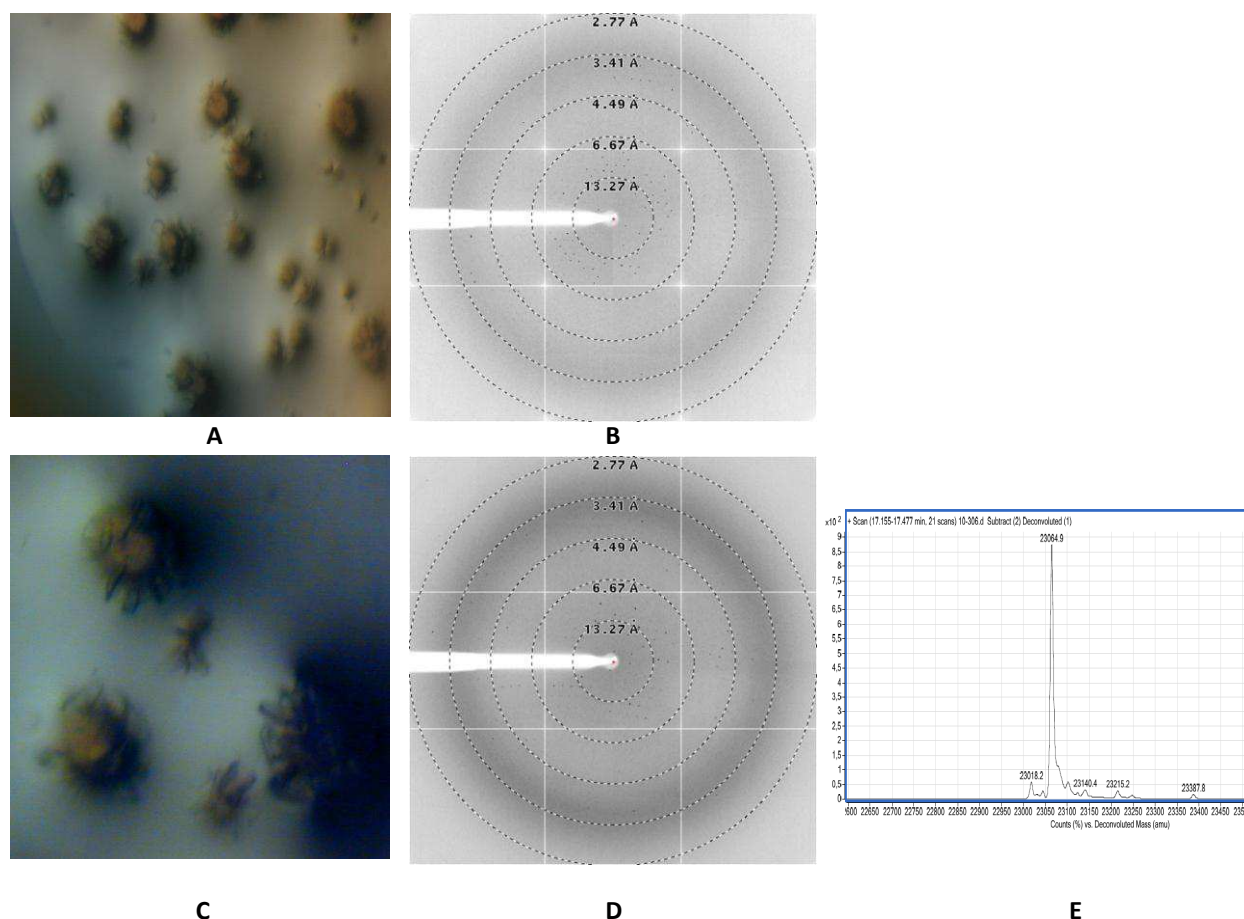


Figure 3.1.7. Crystallization of TD3 (692-892 aa). **A.** Native Crystals of TD3. **B.** Diffraction pattern of the native crystal at 3.4 Å resolution. **C.** Crystals of seleno-methionine TD3. **D.** Diffraction pattern of Se-Met TD3 at 3.8 Å resolution. **E.** Seleno-Methionine (SeMet) incorporation into TD3, confirmed by MS analysis.

The second protein construct identified after limited proteolysis was apo-TD3 (692-917 aa). Domain prediction analysis revealed that apo-TD3 had a predicted additional α -helix at its C-terminal to the TD3 construct. The protein was monomeric in solution and homogeneous (data not shown). Crystals grew in Cartesian robot drops within 3-4 days at 20 °C in a condition containing 10 % PEG 1000 and 10 % PEG 8000 (fig.3.1.8.A). Dataset was collected at ID14-EH4 beamline at ESRF at 2.8 Å resolution (fig.3.1.8.B). The data were processed but the structure could not be solved by Molecular Replacement using as a search model the Tudor-SN protein. The crystals could not be reproduced either by sitting or hanging drop techniques and the Se-Met substituted protein did not give any crystals. One possible reason is that the additional predicted α -helix is rapidly degraded, preventing the protein to crystallize, as seen in limited

proteolysis experiment (data not shown). Various other protein constructs were designed, either by truncating or adding residues to the C-terminal part of TD3 with no crystallization success (fig.3.1.9).

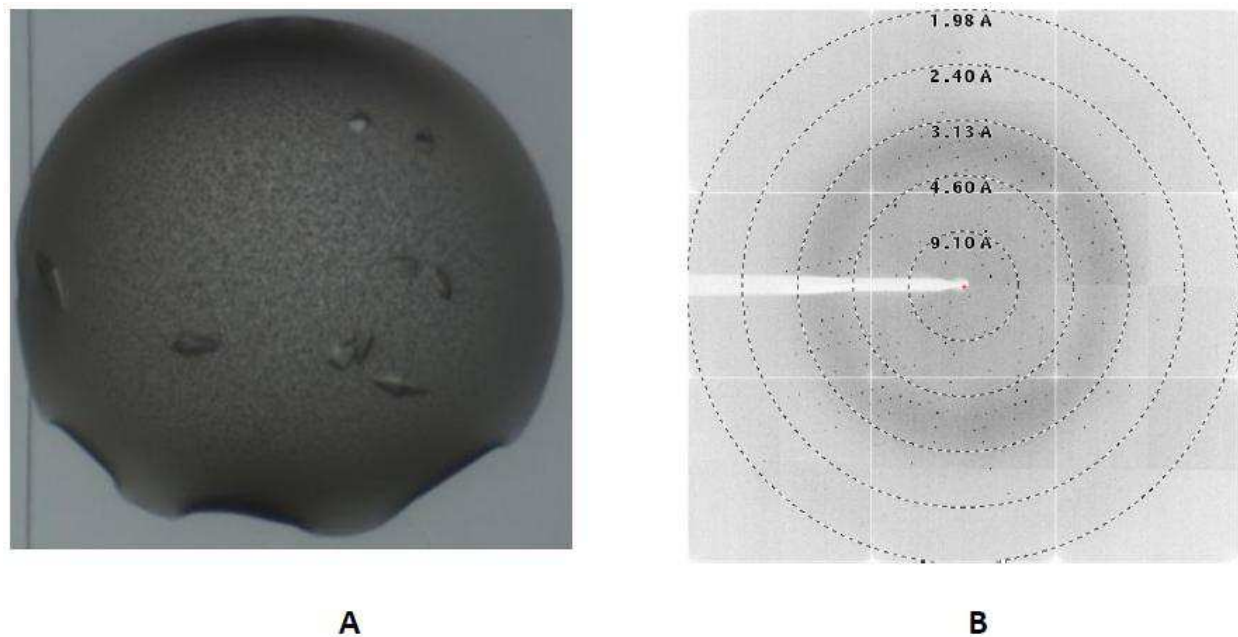


Figure 3.1.8. Crystallization of apo-TD3 (692-917 aa). A. Native Crystals of apo-TD3. B. Diffraction pattern of the native crystal at 2.8 Å resolution.

The information derived from the above experiments and prediction-based bioinformatic analysis was used to design constructs for the crystallization of Tudor 4 domain. TD4 construct (916-1113 aa) was purified using the standard protocol and subjected to crystallization trials at 32.35 mg/ml, 25 mg/ml and 15 mg/ml at 20 °C (fig.3.1.6.D-E). After seven days crystals appeared in all concentrations in a number of conditions (0.2 M magnesium chloride, 0.1 M Tris-HCl pH 8.5, 25 % w/v PEG 3350 and 0.2 M magnesium chloride, 0.1 M Tris-HCl pH 8.5, 15 % w/v PEG 4000) and were reproduced by the hanging drop method (fig.3.1.6.F). The crystals were flash-frozen in a cryoprotectant including 20 % glycerol and the crystallization condition, and diffracted at 8 Å resolution. Attempts to optimize the diffraction resolution included an additive screening (provided by the HTX laboratory), seeding, screening of various

cryoprotectants, sitting drop technique, with all not improving the quality of the crystals obtained. Designing new constructs based on TD4 gave no crystallization hits (fig.3.1.9).

The identification of soluble and prone to crystallization constructs allowed us to define additional constructs harboring tandem eTudor domains (TD2-3, TD3-4, fig.3.1.10). Despite the unsuccessful trials for crystallization, these protein fragments are implicated in biochemical and biophysical characterization of the TDRD1 protein.

Construct	Amino acids	Solubility	Crystallization robot
3#1	531-750	Soluble aggregates	
3#2	686-892	Soluble	No hits
3#3	686-885	Soluble	No hits
3#4	686-917	Soluble	No hits
3#5 *	692-824	Soluble	
3#6 (TD3) *	692-892	Soluble	Crystal structure
3#7 (apo-TD3) *	692-917	Soluble	Crystal structure
3#8	692-927	Soluble	No hits
3#9 *	692-872	Soluble	
3#10	692-903	Soluble	
3#11	703-917	Soluble	
3#12	707-892	Soluble aggregates	
3#13 *	707-917	Insoluble	
3#14	707-885	Soluble	
3#15	707-824	Soluble	
3#16	717-824	Soluble	No hits

Construct	Amino acids	Solubility	Crystallization robot
4#1	916-1050	Soluble	No hits
4#2	916-1028	Soluble aggregates	
4#3 (TD4) *	916-1113	Soluble	Crystals
4#4	916-1155	soluble	No hits
4#5	916-1172	Soluble aggregates	
4#6	916-1099	Soluble	No hits
4#7 *	916-1123	Soluble	No hits
4#8 *	916-1133	soluble	No hits
4#9	916-1094	soluble	No hits
4#10 *	916-1115	Soluble	No hits
4#11 *	926-1113	Insoluble	
4#12	926-1142	Insoluble	
4#13	930-1028	Insoluble	
4#14	930-1123	Insoluble	
4#15	930-1113	Insoluble	
4#16	930-1155	Soluble aggregates	
4#17	977-1123	Insoluble	
4#18	977-1172	Insoluble	
4#19	986-1113	insoluble	

Figure 3.1.9. Constructs comprised of single Tudor domains 3 and 4 of TDRD1. In these tables are listed the various protein fragments expressed, purified and sent for crystallization trials. The constructs with an asterisk were expressed after the identification of single eTud domains, while in parentheses are denoted the protein fragments used for biochemical experiments described later.

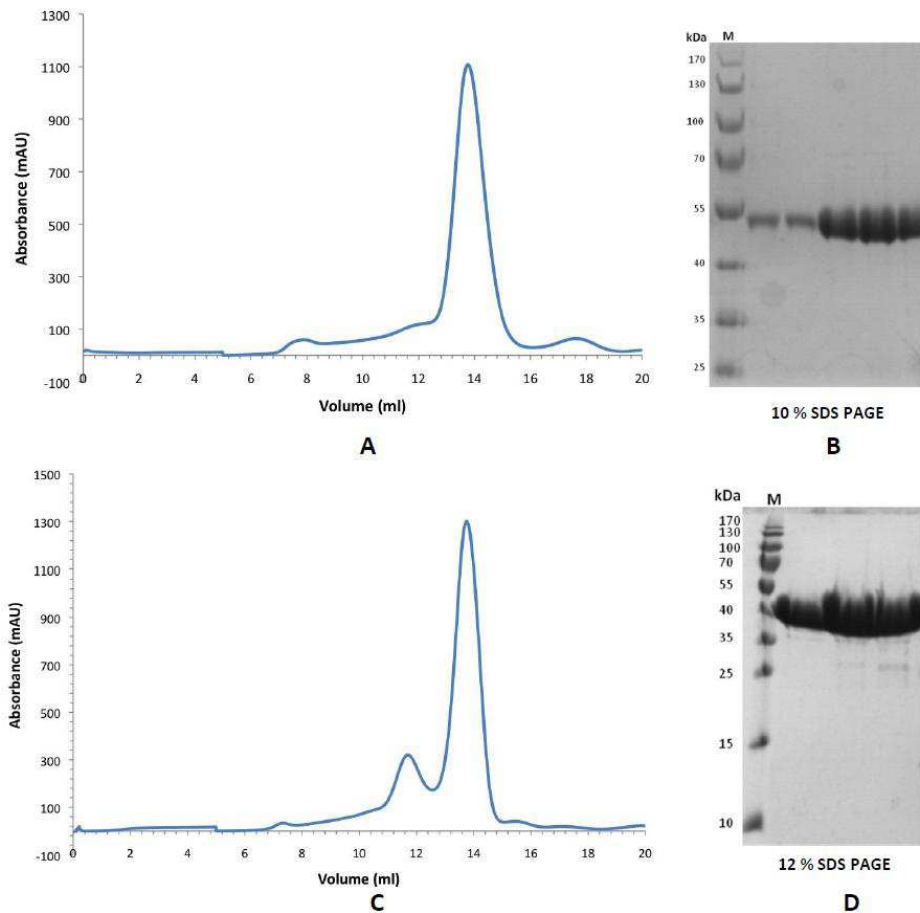


Figure 3.1.10. Purification of tandem Tudor domains TD2-3 (463-892 aa) and TD3-4 (692-1113 aa). **A.** Size exclusion chromatography profile of TD2-3 on a S200 gel filtration column (Blue line: Absorbance at 280 nm). **B.** SDS-PAGE of the peak fractions. **C.** Size exclusion chromatography profile of TD3-4 on a S200 gel filtration column (Blue line: Absorbance at 280 nm) **D.** SDS-PAGE of the peak fractions. Size marker in kilodaltons (kDa) is indicated.

3.1.3 - TDRD1 is involved in the piRNA pathway and interacts with methylated arginines at the N-terminus of Mili.

In 2009, it was revealed for the first time that Tudor domain containing protein 1 (TDRD1) interacts with the PIWI protein Mili (Kabsch, 1993; Kojima et al., 2009; Reuter et al., 2009; Vagin et al., 2009; Wang et al., 2009). This interaction is critical for the maintenance of genome integrity, since loss of TDRD1 elevates the levels of transposons that cause male sterility.

Furthermore, Mili's N-terminus is decorated by multiple Arginine-Glycine (RG) repeats and the symmetrical dimethylation of arginines is responsible for the recruitment of the Tudor domains of TDRD1 (Reuter et al., 2009; Vagin et al., 2009). More specifically, Reuter et al demonstrated that a specific 13-aa long peptide (R74me2) including a symmetrically dimethylated arginine (sDMA) derived from Mili can specifically interact with the four tandem Tudor domains of TDRD1, excluding the N-terminal MYND domain from this interaction ([fig.3.1.11.A-B](#))

As Mili and TDRD1 interact via symmetrically dimethylated arginines, in the first place we wanted to gain more information into the specificity of this interaction. For that reason, we decided to repeat the pull down assay described in [fig. 3.1.11.B](#). *In vitro* translated TD1-4 (232-1094 aa) with TnT Coupled Reticulocyte Lysate System and subsequent protein labeling with ³⁵S-Methionine was incubated with R74me2 peptide that was conjugated to biotin at its C-terminus (see [materials and methods](#)). The exposure of the gel revealed that TD1-4 specifically interacts with R74me2 peptide, while the unmethylated version of the peptide (R74) did not interact ([fig.3.1.11.C](#)), showing that the methylation is responsible for this specific association. Then, in order to explore if the methylated peptide has a binding preference for a specific Tudor domain, we decided to repeat the pull-down assay using tandem Tudor domains 1-2 (TD1-2 construct), 3-4 (TD3-4 construct) and single Tudor domains (constructs TD1, TD2, TD3, TD4). Even if the *in vitro* translation of each construct was successful, we always observed unspecific binding to the unmethylated peptide (data not shown). As explained later, there is no interaction of the non-methylated peptide with Tudor domains of TDRD1.

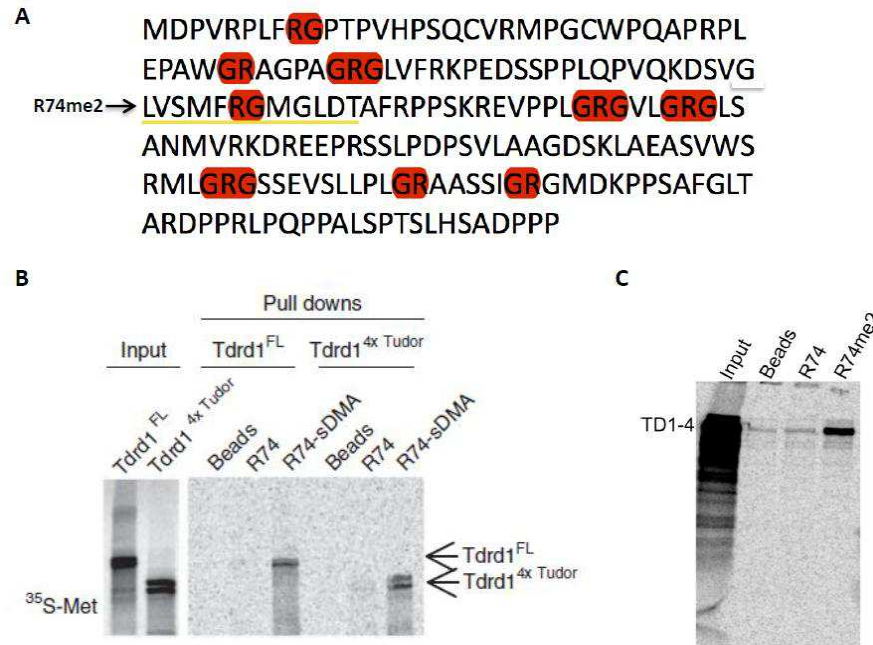


Figure 3.1.11. The N-terminus of Mili contains multiple arginines and interacts with TDRD1 protein. **A.** Arginine-glycine (RG, highlighted in red) repeats, a common target for arginine methyltransferases. Underlined in yellow is the sequence of the R74me2 peptide that when symmetrically dimethylated, interacts with the eTudor domains of TDRD1 (4X Tudor or TD1-4) (Reuter et al., 2009). **B.** Pull-down assay proving methylation-dependent interaction of R74me2 with 4X Tudor (Reuter et al., 2009). **C.** Pull-down experiment (B) reproduced.

3.1.4 - Recognition of sDMA-peptides by single eTudor domains of TDRD1.

Proteomic studies on endogenous PIWI proteins from *Drosophila melanogaster* and mouse combined with Mass Spectrometry analysis identified specifically the modified arginine residues at their N-terminus (Chen et al., 2009; Kirino et al., 2009; Nishida et al., 2009; Reuter et al., 2009; Vagin et al., 2009). More specifically, Mili has eight arginines found to be symmetrically dimethylated, while four of them are also found to be monomethylated (MMA), (fig.3.1.12). MIWI, on the other hand, has two arginines symmetrically dimethylated and three monomethylated, while an additional RG cluster is not modified. MIWI2 is not identified so far to have post-translational modifications despite the presence of characteristic arginine-glycine repeats (fig.3.1.12). Fly Aubergine and AGO3 have three sDMAs each and no monomethylated marks.

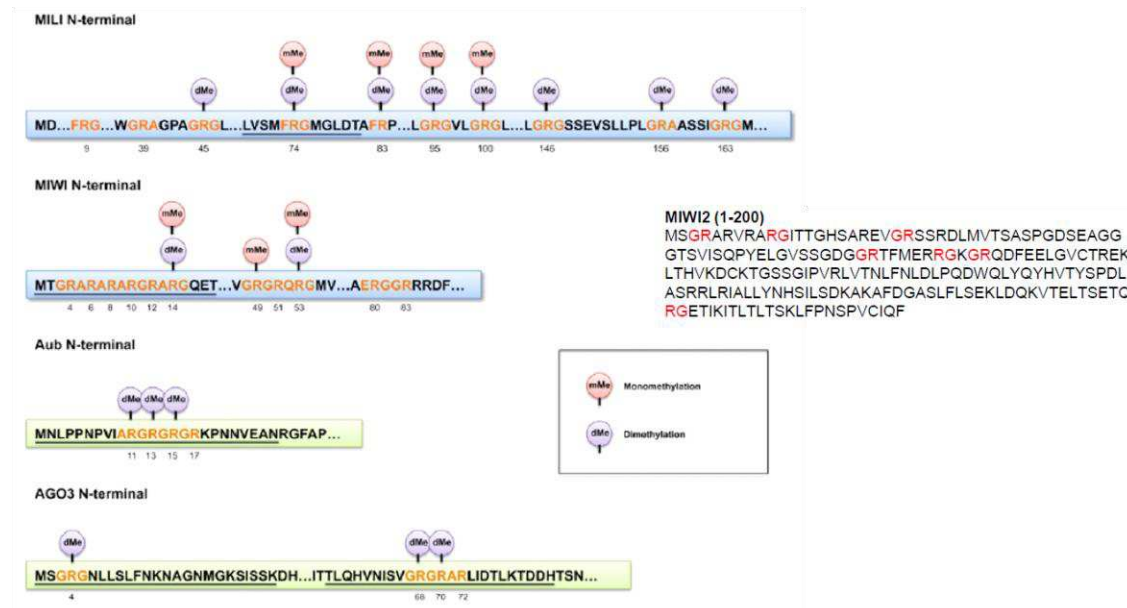


Figure 3.1.12. The N-terminus of PIWI proteins contains methylated arginines. In mouse, Mili and Miwi contain multiple arginines that are symmetrically dimethylated (purple circle) or monomethylated (red circle) (Siomi et al., 2010). Miwi2 contains multiple RG repeats, even though the methylation status of the arginines is not identified yet. In *Drosophila*, Aubergine and AGO3 contain symmetrically dimethylated arginines. For fly protein Piwi, there is no information for post-translational modifications.

The presence of multiple methylation marks at the N-terminus of Mili prompted us to question how they interact with each eTud domain of TDRD1. To address this issue we took advantage of the well behaved constructs of tandem Tudor domains (TD1-4, TD2-3, TD3-4) and the single Tudor domain constructs TD1, TD2, TD3 and TD4. We then measured by Isothermal Titration Calorimetry (ITC) the affinity of each single domain to three different Mili peptides previously shown to contain *in vivo* a symmetrical-dimethylated arginine (fig.3.1.13). The peptides tested were at successive positions and denoted R45me2 (38-GRAGPAGRme2GLVFR), R74me2 (69-LVSMFRme2GMGLDT) and R83me2 (76-MGLDTAFRme2PPSKR) (fig.3.1.12, fig.3.1.13). It is noteworthy that, whereas previous structural studies have used peptides in which the methylated arginines have been flanked both sides by glycine or alanine residues (Liu et al., 2010a; Liu et al., 2010b), this is only true of the R45me2 Mili peptide.

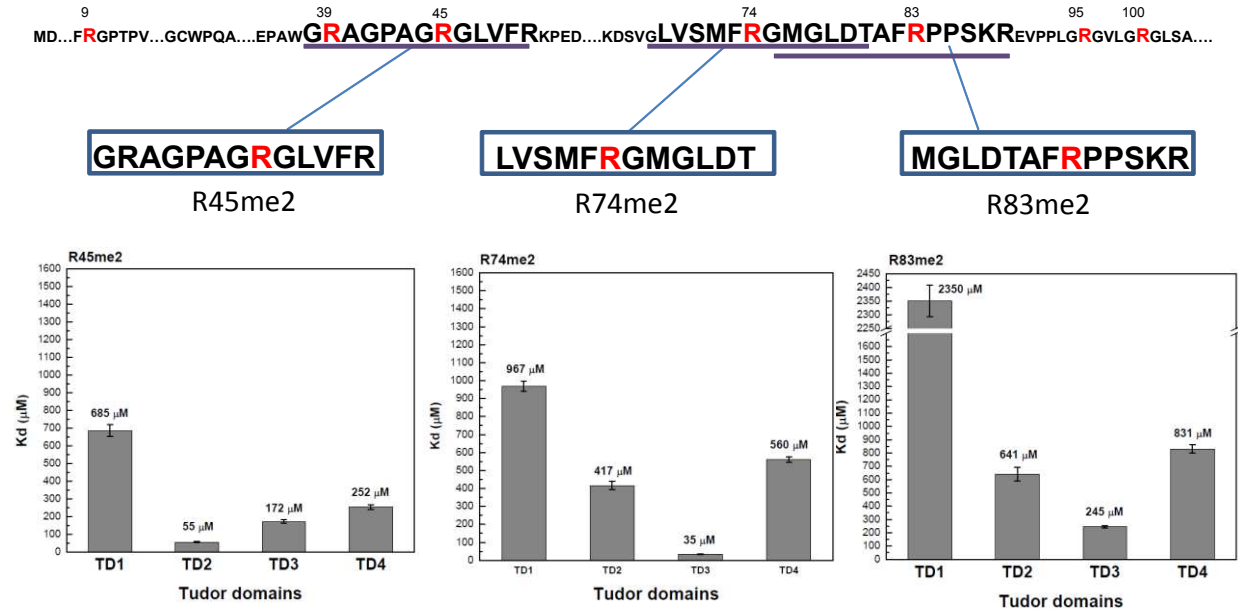
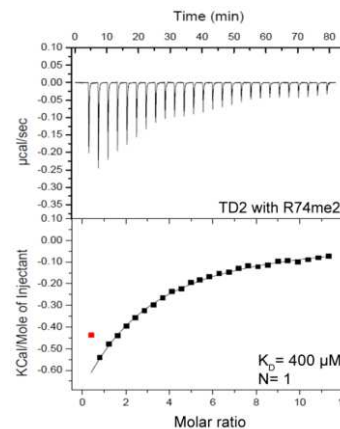
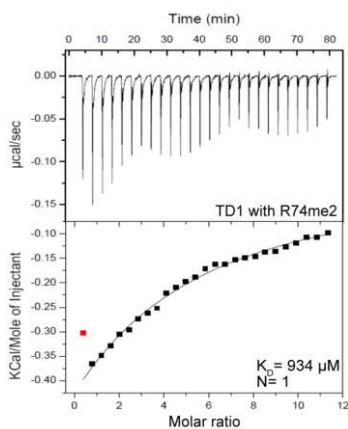
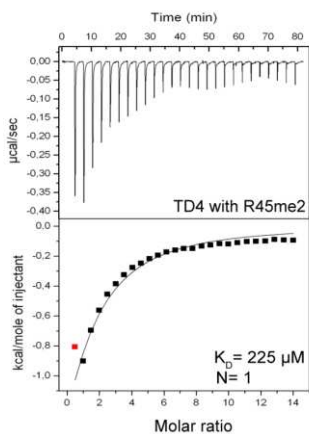
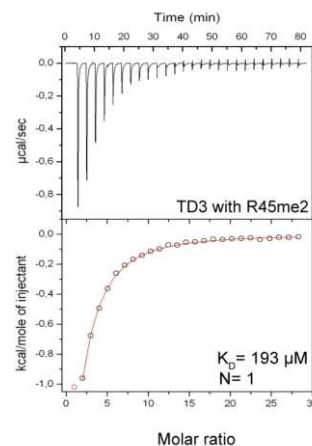
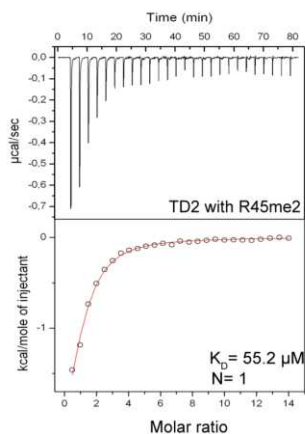
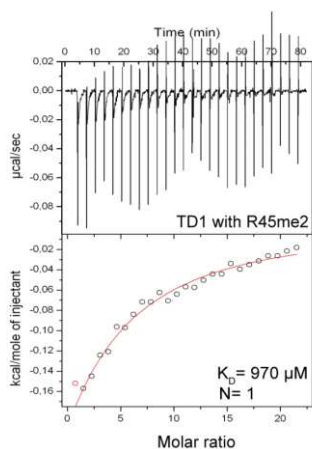


Figure 3.1.13. Location and binding affinity of sDMA containing peptides of Mili to TDRD1's eTud domains. K_Ds derived from ITC measurements for three sDMA containing peptides of Mili binding to the four individual eTud domains of mouse TDRD1. The three methylated peptides used in this study are indicated at the top. Each bar corresponds to mean values obtained from duplicate and triplicate ITC titrations.

The ITC results are summarized in [fig.3.1.13](#) with representative experimental data and fits shown in [fig.3.1.14](#). The R45me2 peptide had the highest affinity for TD2 (K_D= 55 μM) and TD3 (172 μM) domains, with TD4 having a slightly lower affinity (252 μM) and the TD1 domain showing poor binding (685 μM). The R74me2 peptide, in which the symmetrical dimethylarginine (Rme2) is only flanked in one side by a glycine, showed high affinity only for the TD3 domain (35 μM) and for apo-TD3 (141 μM), with the other domains showing substantially weaker binding ([fig.3.1.14](#)). Indeed, when we titrated the R74me2 against the tandem domain constructs TD2-3 and TD3-4 the measured K_D of the major binding site in a two site model is similar to that of the single TD3 domain (respectively 45 and 51 μM) with a binding to a second site being 6-10-fold weaker (respectively 312 and 495 μM) ([fig.3.1.15](#)). The importance of the methylation mark on the arginine in the context of the peptide is denoted by the non-binding titration of TD3 with R74 ([fig.3.1.14](#)). The R83me2 peptide, which has an unusual Phe-Rme2-Pro tripeptide, showed low affinity with all single eTud domains: TD3 domain had the lowest K_D (245 μM), with TD2 (641 μM) and TD4 (831 μM) following. TD1 domain, as for the rest of the peptides, showed a much higher K_D value (2.3 mM) ([fig.3.1.14](#)).

These results clearly indicate that the central TD2 and TD3 domains of TDRD1 have the highest affinities for the tested sDMA-containing peptides of Mili, with TD4 and especially TD1 having significantly lower affinity.



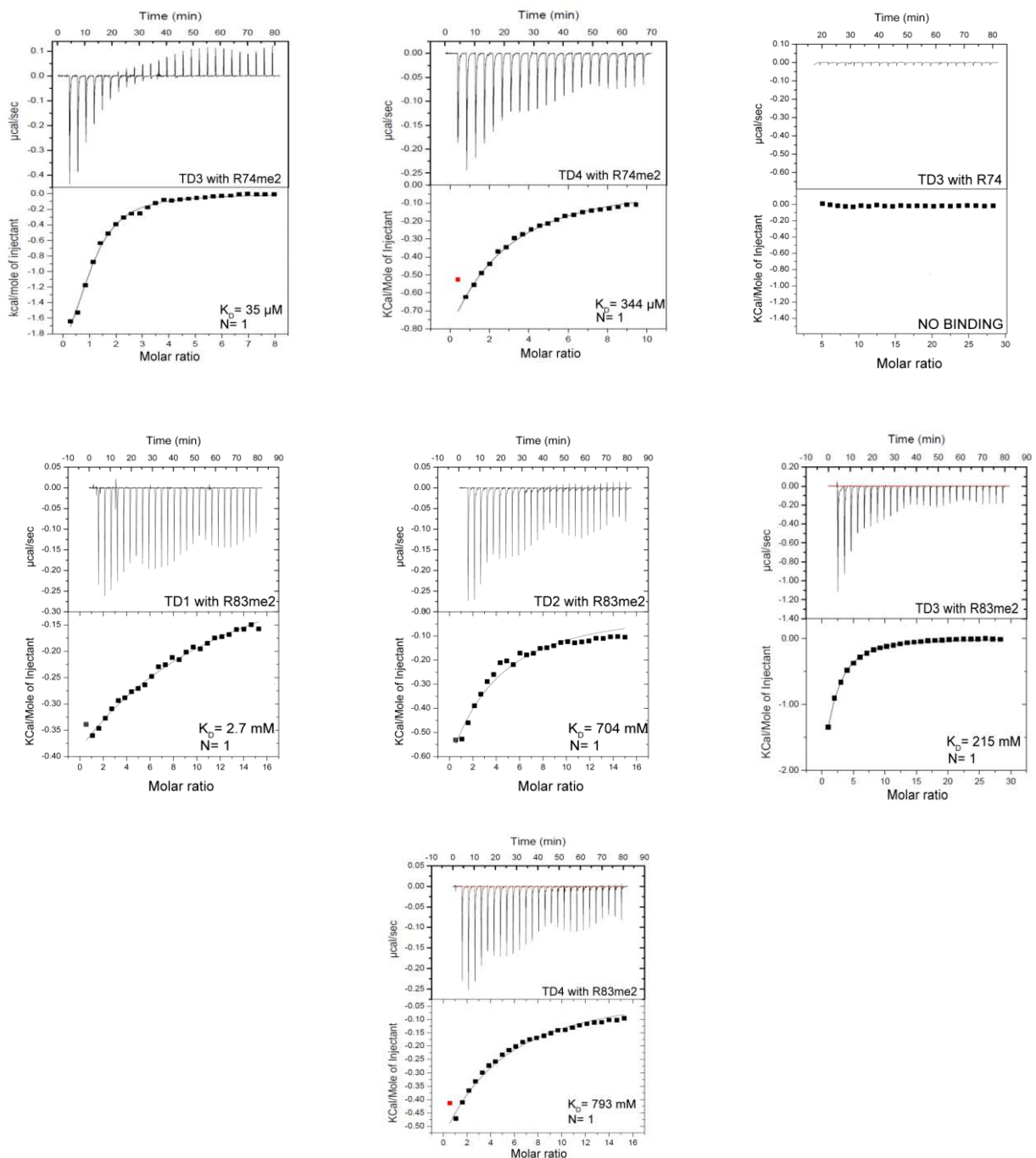


Figure 3.1.14. Representative ITC experiments for the four single eTud domains (TD1, 2, 3 and 4) of TDRD1 with the R45me2, R74me2 and R83me2 peptides of Mili. In each panel, the ITC raw data (top) and the fitting of the data (bottom) are depicted. The data were fitted to a single-site binding model. The dissociation constants are shown, as well as the number of binding sites. The R74 peptide with no symmetrical dimethylarginine shows no binding to TD3. ITC parameters are summarized in [Table 1](#).

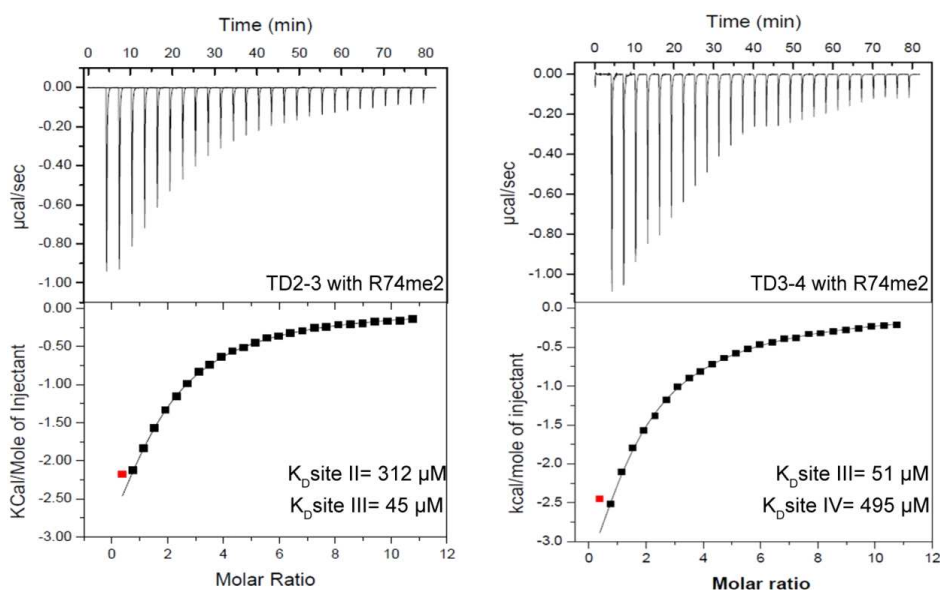


Figure 3.1.15. ITC titrations of the double eTud domains TD2-3 and TD3-4 with the R74me2 peptide of Mili. Representative ITC experiments for tandem Tudor domains TD2-3 and TD3-4 with R74me2 peptide of Mili. The data were fitted to a sequential binding site model (The model corresponds to two different binding sites as the peptide binds differently to each site of the Tudor domain. The stoichiometry is 2:1 -peptide:protein-, or 1 peptide per Tudor domain site). ITC parameters are summarized in [Table 1](#).

TD2 and TD3 showed the highest affinity for consecutive singly arginine methylated Mili peptides, respectively R45me2 and R74me2. This prompted us to test the binding of tandem TD2-3 and TD3-4 domains to the doubly methylated peptide 43-AGR45me2...FR74me2GMG-77, denoted R4574me2s, which encompasses both the original peptides, for indications of cooperativity. Satisfactory fits were achieved with one binding site ([fig.3.1.16](#)) and gave a K_D of 48.5 μ M for TD2-3. This is comparable to the single TD2-R45me2 and TD3-R74me2 affinities (respectively 55 and 35 μ M), showing that there is no enhancement of affinity. A possible explanation for no cooperativity could be that the two recognized arginines and the two Tudor domains are far enough, and thus they can act in an “independent” like manner. Both the long linker, about 30 residues, between TD2 and TD3, and the long sequence connecting the two interacting arginine (28 residues), support this hypothesis. For TD3-4, titration of R4574me2s showed a K_D of 29 μ M, which is comparable with the TD3-R74me2 interaction (35 μ M) and consistent with little contribution from TD4 binding (252 μ M for R45me2), suggesting that weak cooperativity may take place. Furthermore, titration with R4574me2s and single TD2 and TD3

showed similar K_D s for the tandem TD2-3 and TD3-4 (39 μ M and 66 μ M for TD2 and TD3 respectively) (fig.3.1.16). Taken together, these results suggest that the central TD2 and TD3 domains are binding methylarginine peptides independently.

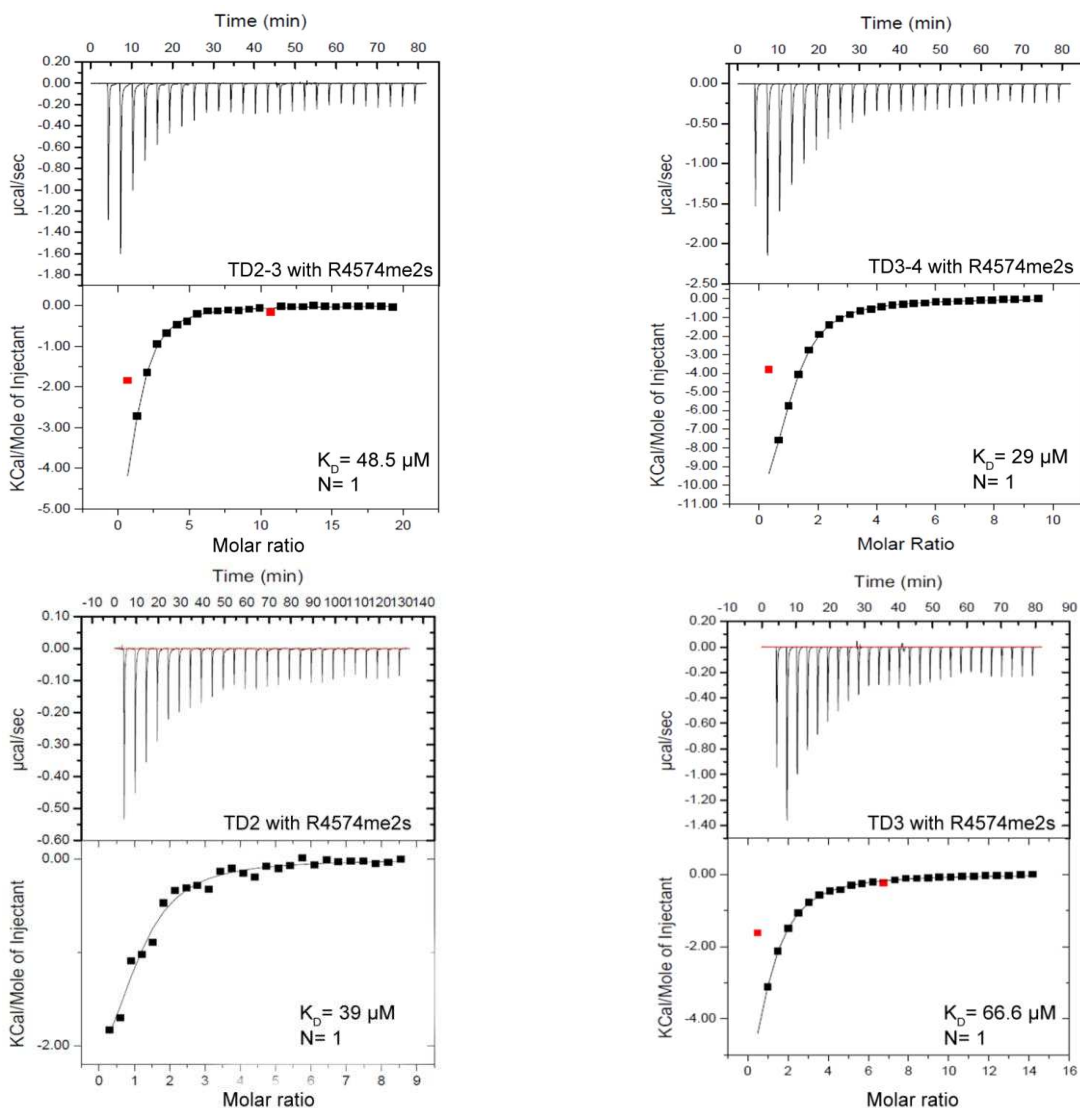


Figure 3.1.16. The double eTud domains TD2-3, TD3-4 and single TD2 and TD3 with the R4574me2s doubly methylated peptide. Representative ITC experiments of tandem constructs TD2-3 and TD3-4 and single eTuds TD2 and TD3 binding to a long peptide R45R74me2s of Mili, harboring two methylated arginines (sequence as shown). The data were fitted automatically to a single-site binding model. ITC parameters are summarized in Table 1.

3.1.5 - Crystallization trials of eTudor domains-peptide complexes

In order to better understand the recognition of Mili derived sDMA-peptides by the Tudor domains of TDRD1, we tried to co-crystallize each of the domains with different peptides. In the case of R74me2 peptide, TD3, apo-TD3 and TD4 proteins were mixed with the peptide by directly adding a three- or fivefold molar excess of the latter to the protein solution. Crystals of apo-TD3 with R74me2 in 1:5 molar ratio appeared in a condition containing 0.05 M Potassium phosphate and 20 % w/v PEG8000 at 20 °C by the robot sitting drop vapor diffusion method ([fig.3.1.17.A](#)). Complete native data set was collected at ID14-EH4 beamline (ESRF) at 3.4 Å resolution ([fig.3.1.17.B](#)). Data were processed as described in [Materials and Methods](#) and the structure was solved by Molecular Replacement, using the apo-TD3 structure (see [section 3.1.7](#)) as a search model. However, no electron density was found for the peptide, suggesting that is not included in the structure. TD4-R74me2 crystallized in the same condition as TD4 and diffracted also poorly at 8 Å resolution, implying that also the peptide is not present in the crystal ([fig.3.1.17.C](#)). Crystallization trials for TD3 in complex with R74me2 with three-, five- and eightfold molar excess were unsuccessful.

Crystallization trials were also set up for R45me2 peptide. The TD2-R45me2 complex crystals were obtained by mixing the protein and the peptide in 1:5 molar ratio and grew at 20 °C in a solution containing 0.1 M Bis-Tris pH 6.5 and 45 % w/v polypropylene glycol P400 and in Tris-HCl pH 8, 20 % PEG MME 500. The crystals had a pin-like shape and were reproducible in all the conditions by handset drops, but did not diffract ([fig.3.1.17.D](#)). Attempts to crystallize this peptide with TD4 were also unsuccessful, while crystallization trials of the long peptide R4574me2s with tandem and single eTud domains gave no positive hits.

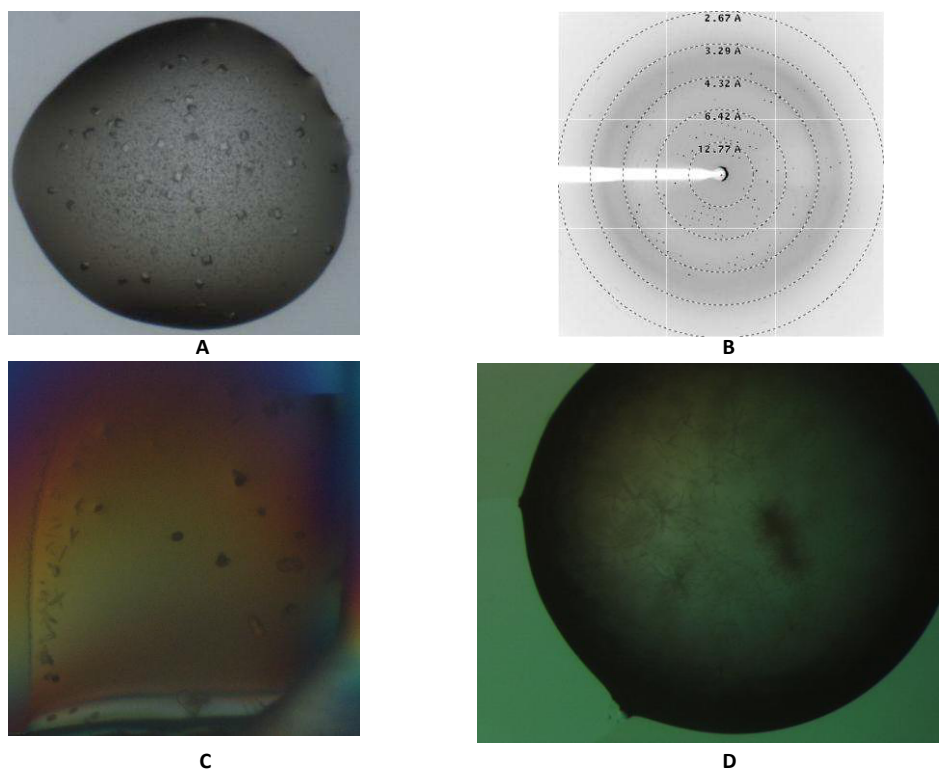


Figure 3.1.17. Crystallization of eTud domains of TDRD1 with Mili peptides. **A.** Protein crystals of apo-TD3 with R74me2 peptide. **B.** Diffraction pattern of the crystals in **A** at 3.4 Å resolution. **C.** Crystals of TD4 with R74me2 peptide. **D.** Crystals of TD2 with the R45me2 peptide.

3.1.6 - Crystal structure of the TD3-R45me2 peptide complex

3.1.6.1 - Crystallization and data analysis

The TD3 (692-892) complex with R45me2 peptide was prepared by mixing protein and peptide in a 1:5 molar ratio, and co-crystals grew at 20°C in 0.2 M ammonium acetate, 0.05 M sodium cacodylate pH 6.5, 30 % PEG8000, 0.01 M magnesium acetate tetrahydrate (fig.3.1.18.A). A complete native data set was collected to a resolution of 2.1 Å (fig.3.1.18.B) on beamline ID14-EH4 at the ESRF (Grenoble, France) and processed with XDS (Kabsch, 1993). Since molecular replacement with the known eTud structures did not work, the structure of the TD3-R45me2 complex was solved *de novo* using seleno-methionine labeled protein and the MAD method. Crystals of the Se-Met-labelled protein-peptide complex were grown at 20 °C in slightly different conditions, 0.2 M NaCl, 0.1 M Bis-Tris pH 7.0 and 29 % PEG 3350 (fig.3.1.18.C).

Datasets with resolution of 2.7-2.8 Å were collected at the inflection point and peak wavelengths of the Se K-edge (fig.3.1.18.D-E). Six SeMet sites were identified using SHARP (de La Fortelle et al., 1997) and phases improved using RESOLVE. The initial model was built manually using COOT (Emsley and Cowtan, 2004) and finally automatically built by ARP-wARP (Perrakis et al., 1997). Crystals belong to the space group of *P1* with two complexes per asymmetric unit. The structure was refined at 2.1 Å resolution using Refmac5 (with TLS refinement) (Murshudov et al., 1997) to a final R-factor of 21.9 % and R_{free} of 28.4 % with all residues in allowed (97.7 % in favored) regions of the Ramachandran plot, as analyzed by MOLPROBITY (Chen et al., 2010) (fig.3.1.19).

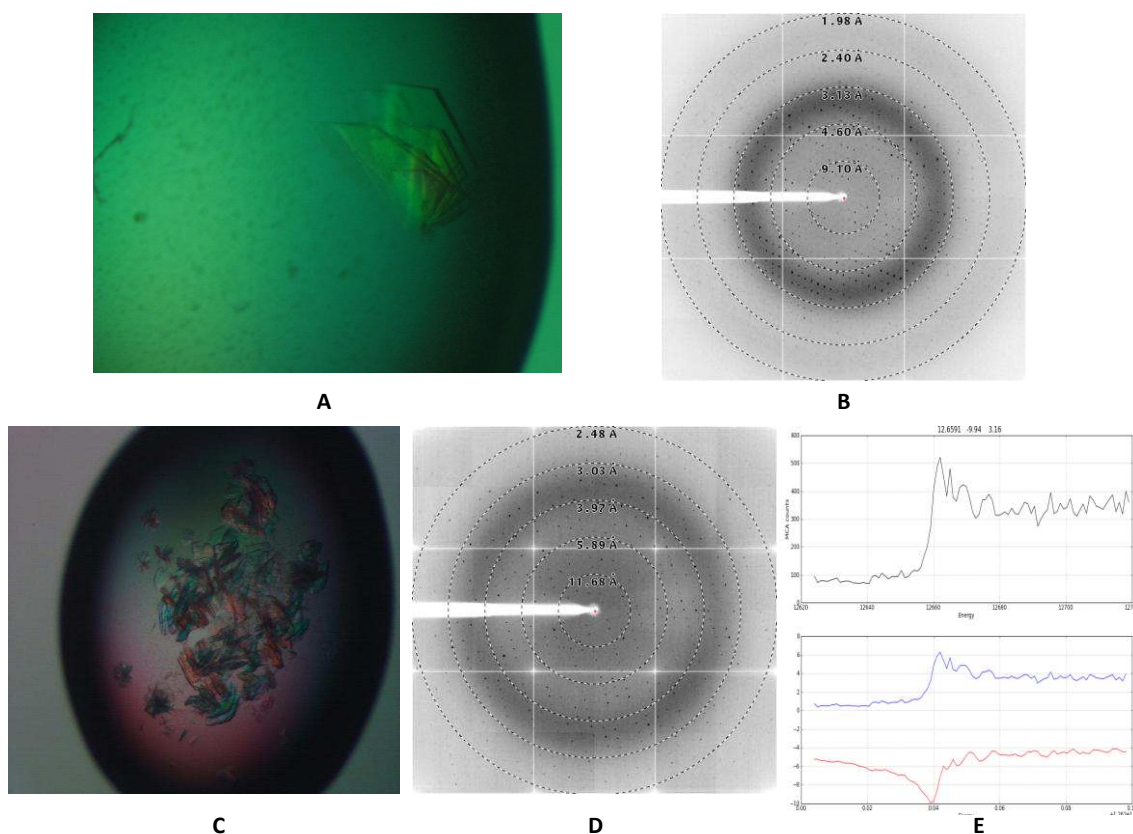


Figure 3.1.18. Crystallization of TD3-R45me2 complex. **A.** Native Crystals of TD3-R45me2. **B.** Diffraction pattern of the native crystal at 2.1 Å resolution. **C.** Crystals of seleno-methionine TD3-R45me2. **D.** Diffraction pattern of Se-Met TD3-R45me2 at 2.8 Å resolution. **E.** X-ray absorption spectrum of the peak wavelength of the Se K-edge.

	TD3/R45me2		TD3	
Data collection	Native	SelenoMet Peak	SelenoMet Inflection	Native
Space group	<i>P1</i>	<i>P1</i>	<i>P1</i>	<i>P4₁2₁2</i>
Cell dimensions				
<i>a, b, c</i> (Å)	39.94, 53.96, 60.79	40.05, 53.63, 61.04	40.51, 54.29, 61.56	62.96, 62.96, 132.48
α β γ (°)	93.69, 98.56, 111.67	93.04, 98.42, 111.97	93.27, 98.22, 111.76	90.00, 90.00, 90.00
Resolution (Å)	50-2.1 (2.2-2.1) ^a	50-2.6 (2.7-2.6)	50-2.9 (2.8-2.7)	50-2.8 (2.9-2.8)
<i>R</i> _{merge}	11.3 (39.3)	11.7 (46.9)	12.3 (42.8)	7.8 (60.3)
<i>I</i> / σ <i>I</i>	7.28 (2.55)	10.05 (2.50)	10.42 (2.2)	14.80 (2.75)
Completeness (%)	95.5 (95.8)	98.5 (97.8)	88.2 (95)	98.8 (93.0)
Redundancy	2.4 (2.4)	3.32 (3.3)	3.51 (2.08)	5.87 (6.06)
Refinement				
Resolution (Å)	37-2.1			132.28-2.8
No. refls used (free)	24480 (1275)			6638 (330)
<i>R</i> _{work} / <i>R</i> _{free}	21.9/28.4			22.7/25.5
No. atoms total	3415			
Protein	3089			1545
Peptide	128			-
Water	186 + 2 glycerol			-
Average <i>B</i> -factors				
Protein	18.3			72
Peptide	19.4			-
Water	21.3			-
R.M.S. deviations				
Bond lengths (Å)	0.01			0.007
Bond angles (°)	1.38			1.011
Ramachandran				
Favoured (%)	97.5			
Allowed (%)	100			

Figure 3.1.19. Crystallographic data and refinement statistics.

3.1.6.2 - Description of the TD3-R45me2 structure

TDRD1 TD3 domain folds into a bilobal α/β extended Tudor (eTud) structure with 196/201 residues visible (697-892). It comprises a Tudor core (prototypic) domain (residues 697-702, 731-818) linked by three connections to a SN (staphylococcal nuclease)-like domain (703-730, 819-892) (fig. 3.1.20.A). The peptide is bound in a cleft between these two lobes and makes contacts with both. The SN-like domain of TD3 is comprised of two β strands (β 1 and β 2) exposed to the peptide binding cleft, and three β strands (β 8-10) and two α helices (α 2 and α 3) at the C-terminal part of the structure. The first two β -strands are connected to the Tudor core domain via a long α helix (α 1). The Tudor core domain is a barrel comprised of four β -strands

(β 3-6), each of which contributes key residues to build the hydrophobic/aromatic cage that binds the methylated arginine side chain (fig.3.1.20.B).

As expected, the overall structure of TD3 resembles the known structures of the eleventh extended Tudor (eTud11) domain of *Drosophila* TUDOR (PDB 3NTI) and human SND1 (PDB 2OMC) with RMSDs of respectively 1.72 Å and 1.61 Å for respectively 153 and 173 matched residues (fig.3.1.21.A). The TD3 SN-like domain superposes well with staphylococcal nuclease (RMSD 1.70 Å for 84 residues aligned, PDB 3T16) but, like eTud11, lacks the long C-terminal helix, which is however found in the SND1 structure. Despite the overall similarity there are some notable differences between TD3 and the other known eTud domains. Firstly, TD3 has an N-terminal extension (residues 697-707) beyond the N-terminal end of the eTud11 and SND1 structures, which crosses back from the SN-like domain to become an integral part of the Tudor core domain. Notably aromatic residues W699 and W701 form hydrophobic interactions with F817 (Tudor core) and F704 with F727 and M820 (SN-like domain) (fig.3.1.20.C). K729 interacting with both E708 and the carbonyl-oxygen of E703 additionally stabilizes the conformation of this unique extension. The extension, which is found identically positioned in the apo-structure (apo-TD3), probably does not directly affect peptide binding, since it is underneath the binding cleft. However, it presumably rigidifies the two lobe structure since it provides a third connection between them. It also explains why constructs modeled on the eTud11 and SND1 structures were not well-behaved since they lacked this extension, which is clearly important for stable folding of TD3. Furthermore it defines a different direction for the linker to the TD2 domain. A second significant difference is the orientation of the helix α 2. Compared with the eTud11 and SND1 structures, the C-terminal end of this helix is tilted away from the peptide binding site (fig.3.1.21.A). This is important in disfavoring contacts to peptides in the previously observed positions and favoring the quite different peptide path observed in the current structure (fig.3.1.20.D, fig.3.1.21.B).

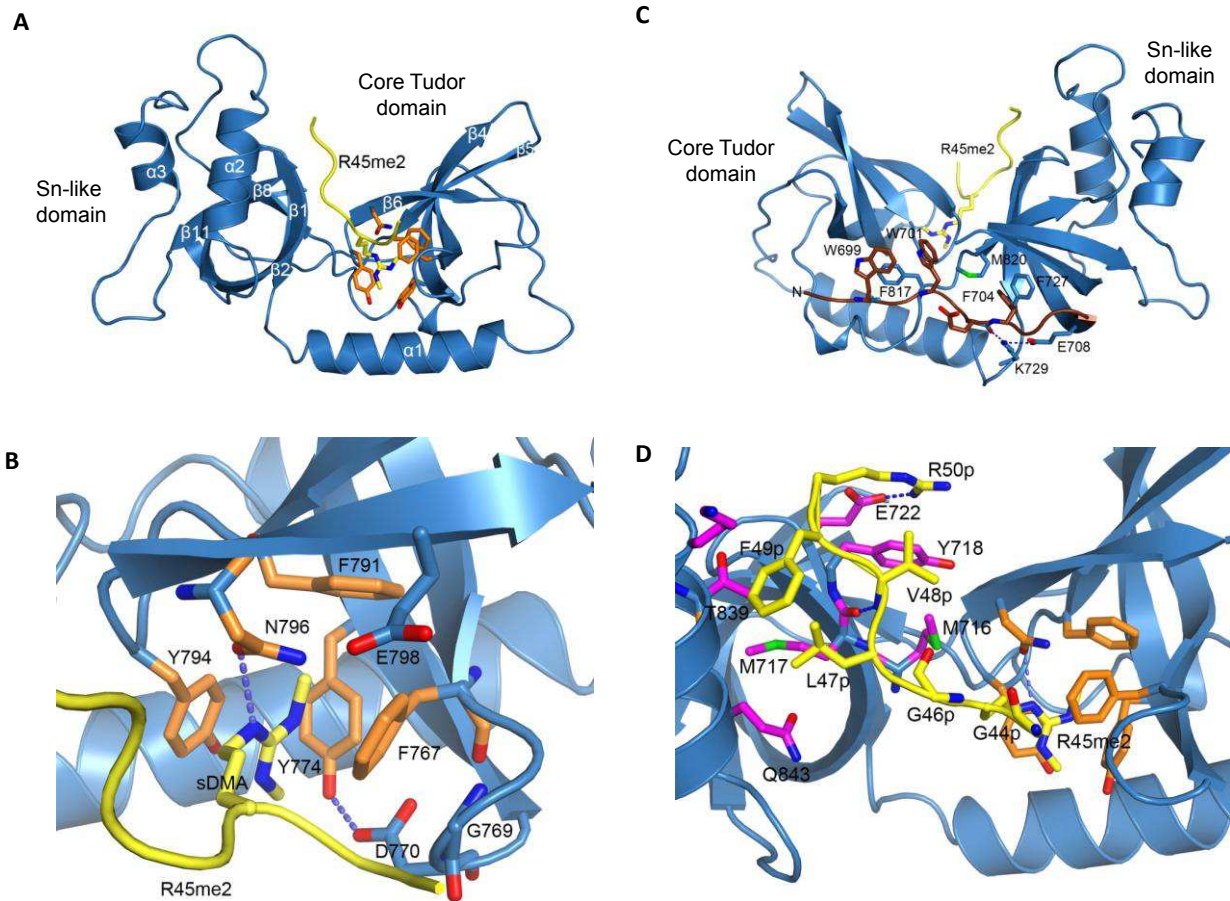


Figure 3.1.20. TDRD1 TD3 structure in complex with R45me2 peptide. **A.** Cartoon representation of TD3 (blue) in complex with the Mili R45me2 peptide (yellow). Secondary structure elements are labeled in white. The aromatic cage residues (orange) and the sDMA (yellow) are represented as sticks. **B.** The sDMA in the aromatic cage of TD3 colored as in **A.** The dipeptide GD stabilizing the cage and E798 interacting with N796 are also shown. **C.** The N-terminal extension (chocolate color) of TD3 makes a third connection between the Tudor core (left) and the SN-like (right) sub-domains against which it packs via large hydrophobic residues such as W699, W701 and F704 as shown. **D.** Details of the interactions between the R45me2 peptide and TD3. Colors are as in **A** and **B.** Residues of the SN-like domain interacting with the peptide are in magenta. Blue dashed lines indicate putative hydrogen bonds.

3.1.6.3 - Recognition of the sDMA by the aromatic cage

The Tudor core domain binds the sDMA of R45me2 via its aromatic binding cage. The residues forming the cage are highly conserved in all Tudor domains that are known to bind sDMAs. The methylated side chain enters the cage through an opening from the top of the structure and is involved in extensive cation- π interactions with the aromatic residues F767, Y774, Y794 and hydrophobic interactions with F791, which constitute the cage (fig.3.1.20.B). Y774 forms the back of the cage, with Y794 and F767 sandwiching the methylated side chain from each side.

F791 and the highly conserved asparagine N796 form the ceiling of the binding pocket. Finally, the dipeptide, 769-GD, reinforces the cage with G769 peptide stacking on F767 and D770 hydrogen bonding to Y774 (fig.3.1.20.B).

By superimposing the TD3-R45me2 with either the eTud11-R4me2 or SND1-peptide structures, a remarkable observation is that the methyl-arginine groups do not overlay but the methylated guanidine group is rotated 120° about the CZ position (fig.3.1.21.C). This is because the side chain enters the cage from a completely different direction. Thus one methyl group of TD3-R45me2 (CQ2, that pointing towards F791) coincides with CQ1 in the other structures, whereas CQ1 of TD3-R45me2 coincides with the CD in the other structures (and reciprocally the CD of TD3-R45me2 coincides with CQ2 in the other structures) (fig.3.1.21.C). A consequence of this is that the conserved asparagine N796 (which is held in position by an interaction with E798) makes a hydrogen bond to the Nε atom of the arginine rather than to NH2 as in other Tudor domain structures. Nevertheless, N796 remains critical for methyl-arginine binding, as reported for related structures, since the single mutant N796A is enough to significantly weaken the interaction with the R45me2 peptide ($K_D \sim 1.4$ mM) (fig.3.1.26.A, fig.3.1.27). Overall, these observations suggest that the aromatic cage of extended Tudor domains displays plasticity and is capable of accommodating methylated arginines with a different direction of insertion.

3.1.6.4 - Orientation of the R45me2 peptide in the TD3 domain structure.

The Mili R45me2 peptide in the crystal structure reveals seven ordered residues, 44-GRme2GLVFR-50, with A43 poorly ordered. Unexpectedly, the structure reveals a unique orientation of the bound peptide compared to the reported eTud-sDMA-peptide structures of fly TUDOR-Aub and human SND1-Miwi complexes (Liu et al., 2010a; Liu et al., 2010b). The C-terminus of the peptide is at the 'top' of the TD3 domain whereas in the other structures, the peptide is bound at the 'bottom' of the eTud domain (fig.3.1.21.B). However, as in the other known structures, apart from the sDMA, the peptide makes additional contacts with residues in the SN domain as well as the Tudor core domain (fig.3.1.20.D).

The peptide plane of G44p (p denotes a peptide residue), which precedes the methylated arginine, is stabilized through cation- π interactions with F767, part of the aromatic cage. This interaction is a consequence of the rotated position of the methylated arginine in the cage and requires the presence of G769 (mentioned above) to avoid steric clash. In SND1 and eTud11 this glycine is substituted by a valine or glutamate, respectively. Residues C-terminal to the R45me2 form an extended chain and interact principally with residues 716-718 from the distal part of strand β 1 and residues 836-843 from the proximal part of helix α 2 of the SN-like domain (fig.3.1.22.A). G46p and L47p make hydrophobic interactions with M716 (a hydrophilic E or Q in other structures) and M717 respectively, both from strand β 1, as well as Gln843 on helix α 2. The amino group of V48p hydrogen bonds with the carbonyl oxygen of M717 and its side chain makes hydrophobic interactions with M716 and Y718 (fig.3.1.20.D). Next, the benzene ring of F49p, which is stacked on L47p, makes van der Waals contacts with K836, which is on helix α 2. The last ordered amino acid of the peptide is R50p, which points out of the structure into the cleft between the SN-like domain and the Tudor core domain. Its amino group makes a hydrogen bond with the carbonyl oxygen of Y718 and its carbonyl group makes a water-mediated hydrogen bond with T839, which is also on helix α 2 (fig.3.1.20.D). Finally, the side chain of R50p stacks on Y718 (strand β 1) and makes a salt bridge with the carboxyl group of E722 (strand β 2). In an attempt to confirm the novel orientation of the peptide, we mutated E722 to alanine and measured the affinity by ITC. However, the K_D had the same value as the wild-type protein (data not shown). It should be pointed out that there are additional interactions of the peptide with a crystallographic related TD3 molecule (denoted *); these involve two main chain-interactions with the main chain of I781* and a salt bridge of Glu722* with R50p. For reasons given below we do not think that these interactions influence strongly the mode of peptide binding observed.

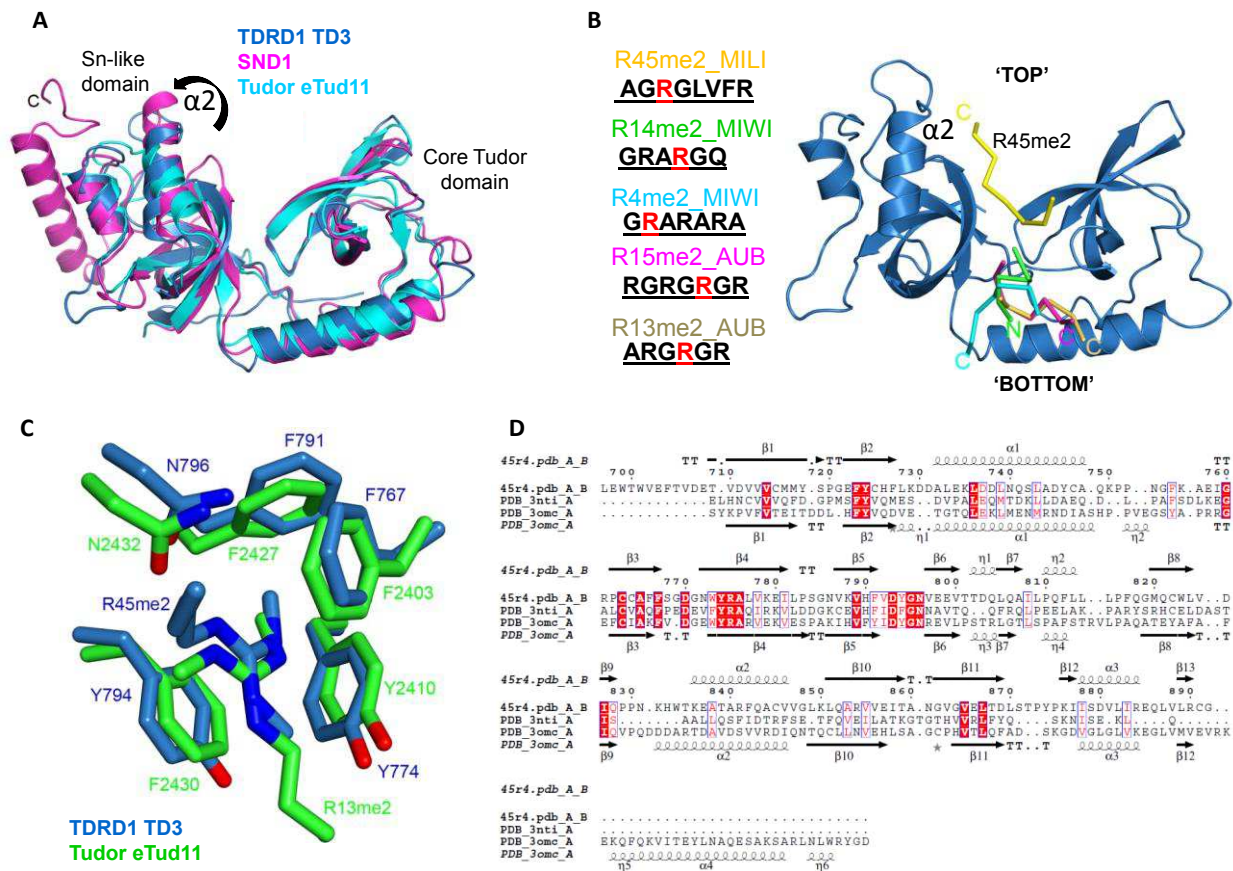


Figure 3.1.21. TD3 binds the R45me2 peptide in a completely different orientation than previously observed in other eTud domains-peptide complexes. **A.** Superposition of the structures of TD3 (blue), SND1 (magenta) and eTud11 (cyan). The C-terminal helix of SND1 is absent from the other two structures. The different orientations of the distal end of the helix $\alpha 2$, which could play a determining role in defining the orientation of the peptide, are indicated by an arrow. **B.** Diagram showing the orientation of various sDMA containing peptides with reference to the TDRD1 TD3 structure (blue) obtained by superposition of eTud-peptide complexes of known structure. R14me2 (PDB 3OMG, green) and R4me2 (PDB 3OMC, cyan) were co-crystallized with the SND1 protein, while R15me2 (PDB 3NTI, magenta) and R13me2 (PDB 3NTH, wheat) were crystallized with eTud11 from *Drosophila* TUDOR. All the above mentioned peptides enter the aromatic pocket from the bottom while the R45me2 peptide (yellow) enters from the top of the structure. The observed peptide residues in each structure are shown at the left with the sDMA highlighted in red. N- and C- termini of the peptides are also depicted. **C.** Structural comparison of the aromatic cages of TD3 (blue) and eTud11 (green) in complex with their corresponding peptides highlighting the different direction of entry of the two methyl-arginine groups. All other such structures exhibit the same conformation as eTud11/R13me2. **D.** Structure-guided sequence alignment between TD3 (45r4), eTud11 (PDB code 3NTI) and SND1 (PDB code 3OMC). Residues with similarity above 70% are displayed in red. The alignment was generated with ClustalX (Larkin et al., 2007) and displayed with ESPript (Gouet et al., 1999, <http://esprict.ibcp.fr/ESPript/ESPript/>).

A significant difference between the TD3 structure and those of eTud11 and SND1 is the orientation of helix $\alpha 2$ of the SN-like domain (fig.3.1.21.A). For SND1 and eTud11 that accommodate their peptides from the bottom (fig.3.1.21.B), the peptide is contacted by residues from the C-terminal (distal) end of this helix (e.g. Asn823 in SND1 equivalent to G848 in TD3, fig.3.1.21.D) as well as proximal parts of strand $\beta 1$ (e.g. F686 in SND1, equivalent to V713 in TD3). In TD3 the helix is tilted such that its N-terminal (proximal) end contacts the peptide, together with distal part of strand $\beta 1$ (as described above), but its distal end is shifted away from the binding cleft and does not contact the peptide (fig.3.1.20.C, fig.3.1.21.A-B). This appears to be an important structural adaptation together with, for instance, the presence of M716 and G769 (mentioned above), which permits the unusual orientation of the peptide in the TD3 structure. Also, while the peptides of eTud11 and SND1 are contacting the long α -helix (helix $\alpha 1$), R45me2 does not show any interaction with it (fig.3.1.22). Overall, we observe that peptide residues are making important contacts with the SN-like domain of eTuds but with different residues comprising it, thus confirming the different orientation of binding (fig.3.1.22). These observations underscore another remarkable plasticity outside the aromatic cage of the eTud fold to accommodate different peptides in quite different orientations.

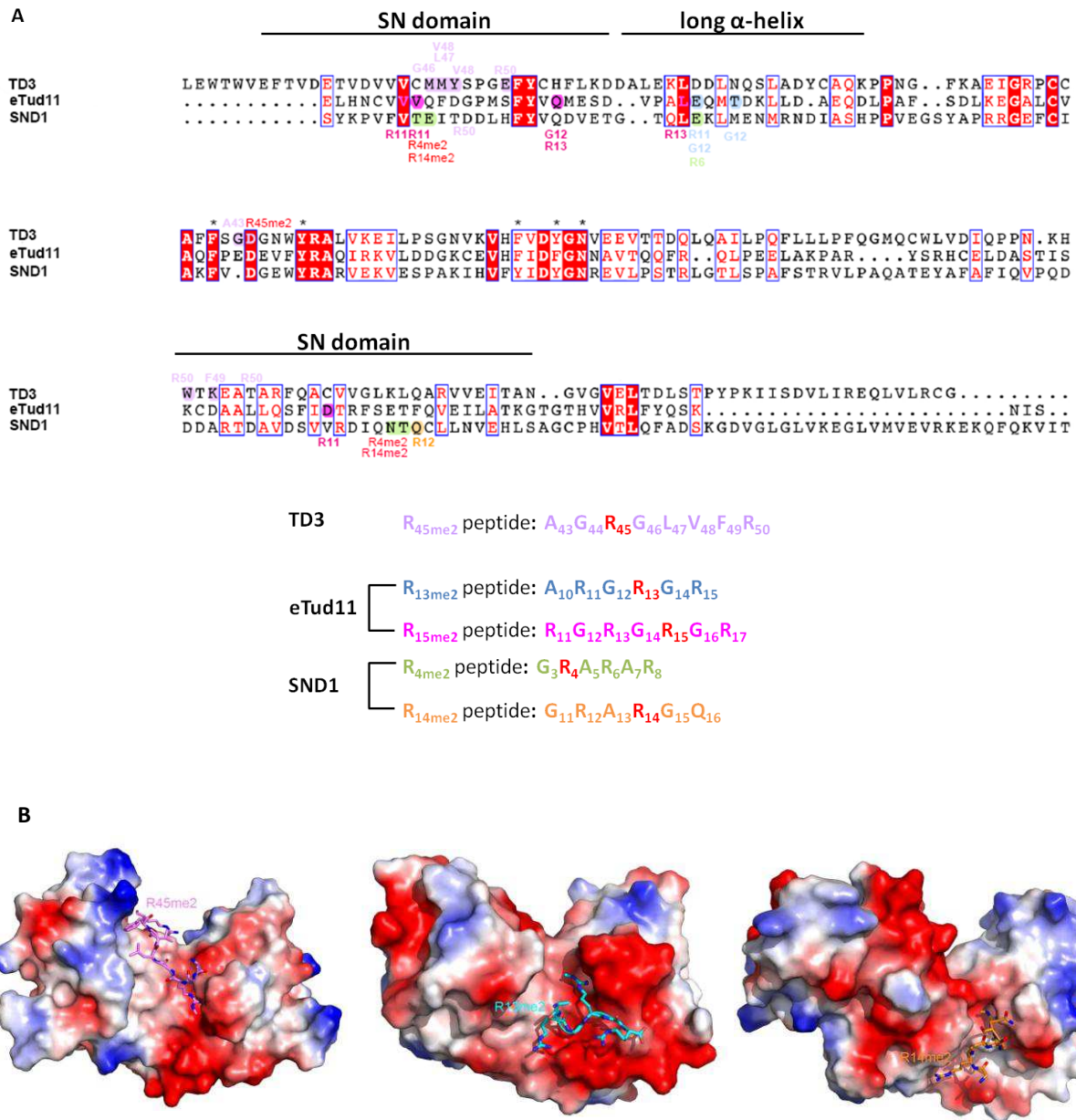


Figure 3.1.22. eTud domains bind PIWI methylated peptides in different orientations. A. Top: Multiple sequence alignment of TD3, eTud11 and SND1. **Bottom:** Sequences of the peptides that are ordered in their corresponding crystal structure. At the top, residues flanking the methylated arginines are shown above and below the residues that interact with. PIWI methylated peptides make extensive contacts with the SN-like domain, albeit with different amino acids, confirming their different orientation. The long α -helix that connects the core Tudor domain with the SN-like domain makes contacts with eTud11 and SND1 methylated peptides, but not with TD3 peptide. **B.** Electrostatic surface potential representation of eTud domains: TD3 (left), eTud11 (middle) and SND1 (right) in complex with their methylated peptides. Blue and red represent positive and negative electrostatic potentials, respectively.

3.1.7 - Structure of unliganded TDRD1 TD3.

3.1.7.1 - crystallisation and structure determination.

Unliganded TD3 (692-917, apo-TD3) crystals were obtained and diffracted as described before ([fig.3.1.8.A-B](#)). They belong to space group $P4_12_12$ and a complete data set was collected on beamline ID14-EH4, processed with XDS and the structure solved by molecular replacement with PHASER (McCoy et al., 2005) using the TD3/R45me2 complex structure as a search model. The structure was refined with Refmac5 (with TLS refinement) to a final R-factor of 22.7 % and R_{free} of 25.5 %. 99.76% of residues are in allowed (95.8% in favored) regions of the Ramachandran plot ([fig.3.19](#)).

3.1.7.2 - Structure description

The structure was solved at 2.8 Å resolution in the absence of bound peptide. Compared to TD3-R45me2, the structures are extremely similar (RMSD 0.65 Å for all 191 common visible residues), with small perturbations likely due to crystal contacts. The extra C-terminal residues are not observed, confirming our limited proteolysis experiments (see section [3.1.2](#)) and further suggesting that TD3 does not have the last C-terminal helix characteristic of SN domains and observed in SND1 but not eTud11. The peptide binding site is unchanged except for a slight expansion of the aromatic cage largely due to movement of the 767-FGD loop ([fig.3.1.23](#)). Thus there is only a small element of induced fit in the peptide binding.

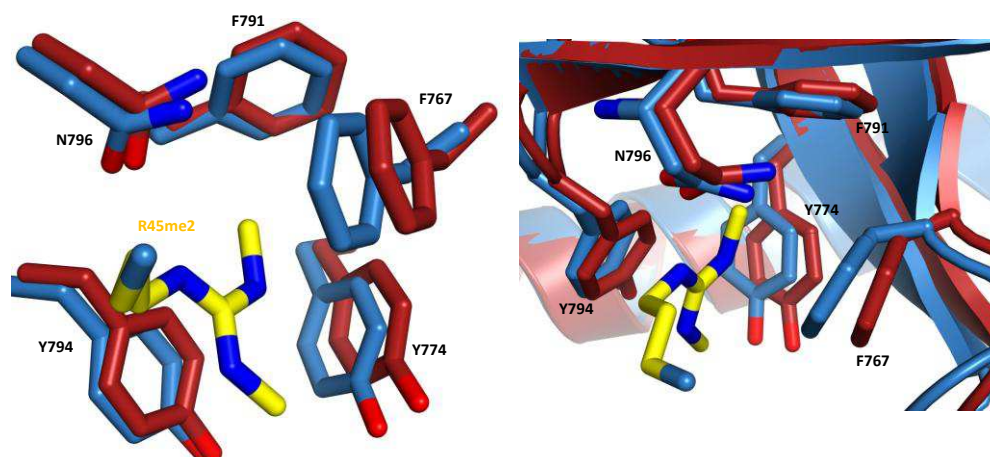


Figure 3.1.23. Comparison of the aromatic cage of TD3-R45me2 complex and apo-TD3. Zoom-in representation of the aromatic cages of Tudor3-R45me2 (blue) and apo-Tudor3 (red) crystal structures. Probably the slight expansion of the aromatic residues is not induced by the methylarginine binding.

3.1.8 - Conserved residues in eTud domains.

The interactions of the R45me2 peptide, apart from that of the sDMA itself, clearly depend on the extended nature of TD3 and indeed it has previously been noted that Tudor domains of TDRD protein family are all found in the extended version (Liu et al., 2010a; Liu et al., 2010b; Vourekas et al., 2010). We therefore used a multiple sequence alignment of eTud domains from mouse TDRD family proteins to identify possible sequence signatures for such extended domains apart from the aromatic residues forming the cage (fig.3.1.24.A). Firstly we identified highly conserved residues that seem important for establishing the eTud double domain fold and these include, in TD3, V714, 723-FY, L736, L743, R775 and 815-LP, whose location is indicated in fig.3.1.24.B. V714 and F723 are conserved residues from respectively strands β 1 and β 2 of the SN-like domain and form part of the hydrophobic core of the domain. Adjacent residue Y724 (W in several domains) makes van der Waals contacts across the inter-domain interface with G795, a highly conserved feature of the aromatic cage. L736 and L743 are conserved hydrophobic residues found in the amphipathic helix α 1 that connects the two domains and pack against the underside of the Tudor core domain contributing to its hydrophobic core. Perhaps the most striking signature of the mouse eTud domains is the

correlated conservation of R775, D793 and 815-LP ([fig.3.1.24.A](#)). The absolutely conserved R775 (adjacent to aromatic cage residue Y774) found in all eTud core domains, but not Tudor core domains (prototypic), plays a key role in pinning together different parts of the structure and stabilizing formation of the aromatic cage. The side chain is largely buried, the aliphatic side-chain making hydrophobic contacts with L739, I808 and F817 and the guanidinium group making a strong salt bridge with conserved D793 (adjacent to aromatic cage residue Y794), as well as three hydrogen bonds to the carbonyl oxygens of F812, L815 and P816 ([fig.3.1.24.B](#)). The LP dipeptide is conserved in SND1 but not in eTud11. These considerations suggest that most of the Tudor domains of the mouse TDRD subfamily are indeed extended Tudor domain.

Secondly, we examined whether there are any conserved residues, outside the aromatic cage, that are involved in extended peptide binding. Equivalent residues E2374 in eTud11 and E708 in SND1 ([fig.3.1.21.D](#)), found in the connecting helix $\alpha 1$ interact via hydrogen bonds with the non-methylated R11 in the R13me2 peptide and R6 in the R4me2 peptide respectively (Liu et al., 2010a; Liu et al., 2010b). In TD3 the equivalent residue D737 does not interact with the peptide. Furthermore Q2365 in eTud11 and Q699 in SND1, at equivalent positions on strand $\beta 2$, interact with R13 and A7 from respectively the R15me2 and R4me2 peptides. This residue is conserved as a glutamine in many mouse eTud domains but is H726 in TD3 ([fig.3.1.24.A](#)). It does not bind the peptide but preserves an important hydrogen bond stabilizing residue D765, one of the aromatic cage loop residues. The fact that these residues do not have conserved functions in TD3 with regard to peptide binding is not surprising given the very different mode of binding of the R45me2 peptide. Indeed we have highlighted above adaptations in the TD3 that appear to favor the mode of peptide interaction observed. Indeed, this may be related to the particular nature of the bound peptide, namely that it is of the form **Rme2G Φ x Φ x** where Φ is a largish hydrophobic residue (M,L,F). Interestingly, the R74me2 peptide 74-**Rme2GMGLD**, which binds the best to TD3 ([fig.3.1.13](#)) has this pattern and is likely that this interaction is driven by hydrophobic interactions, since the entropic forces ($T\Delta S = 3.4$ kcal/mol) indicate that this binding is hydrophobically-driven ([Table 1](#)). Finally, some other, non-tested, Mili modifications, conserve at least the first hydrophobic position e.g. 95-**Rme2GVLG**, 100- **Rme2GLSA**, 163-**Rme2GMDK**. These peptides may all bind in a similar way to that observed for the R45me2

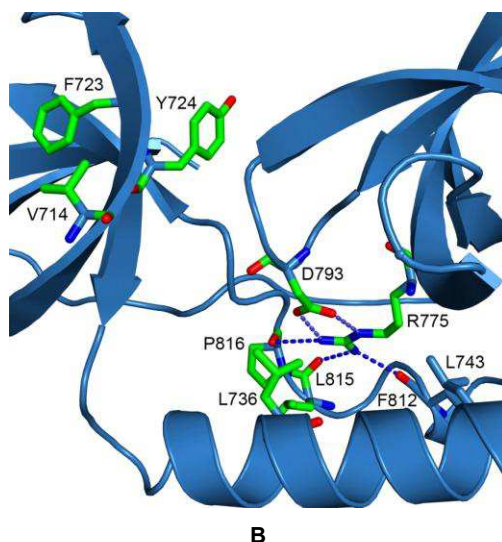


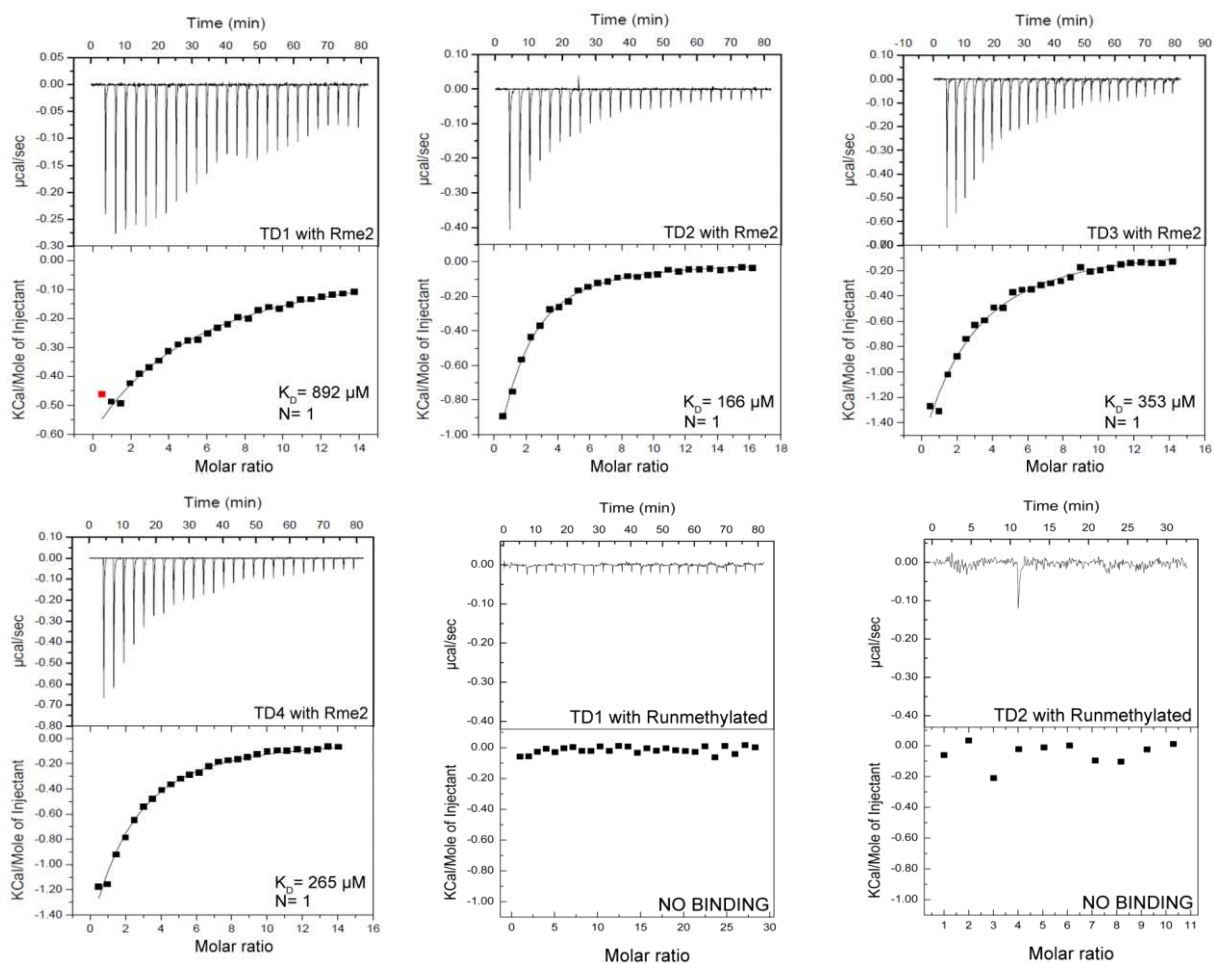
Figure 3.1.24. Signature for the extended Tudor fold in mouse TDRD proteins. **A.** Multiple sequence alignment of extended Tudor domains from murine TDRD proteins. Blue arrows show the conserved aromatic cage residues and red arrows indicate other significant conserved residues present in the extended Tudor domain (eTud) fold. Domains potentially inactive for methyl-arginine binding are indicated with an asterisk. **B.** Conserved residues in murine TDRD eTud domains. In a stick-model representation, the residues conserved in most of the eTud domains of TDRD proteins are shown (green). The absolutely conserved R775 is involved in extensive interactions with residues of the Tudor domain, thus stabilizing the aromatic cage. Blue dashed lines indicate hydrogen bonds.

3.1.9 - Active and inactive eTudor domains

3.1.9.1 – Characterization of ‘active’ – ‘inactive’ eTudor domains of TDRD1

Above we have argued that all the mouse TDRD domains are likely to have the eTud fold, but how many are actually active in binding sDMA containing peptides? Our binding data clearly suggests that TD2, TD3 and TD4 can bind methylated arginine peptides of Mili albeit with differing affinities. However, TD1 showed only weak binding to all the peptides tested. To investigate this further we used ITC to measure the affinity of each separate TDRD1 eTud domain with free dimethylated (sDMA or Rme2) or unmethylated arginines (Runmethyl) (fig.3.1.25). The single arginines used were N-terminally acetylated and C-terminally amidated to mimic a peptide context. The isolated sDMA bound to TD2, 3 and 4 with respective affinities of 172, 353 and 275.5 μM (mean values after duplicates), consistent with the results for the peptides and confirming that the non sDMA parts of the various peptides tested contribute to the affinity observed (fig.3.1.26.A). TD1 domain showed the weakest binding with approximately 830 μM affinity (fig.3.1.26.A, 3.1.25) again suggesting that it may not be involved

methylation-dependent interactions. As expected, the unmethylated arginine showed very low or undetectable binding to all the single eTudor domains (fig.3.1.25).



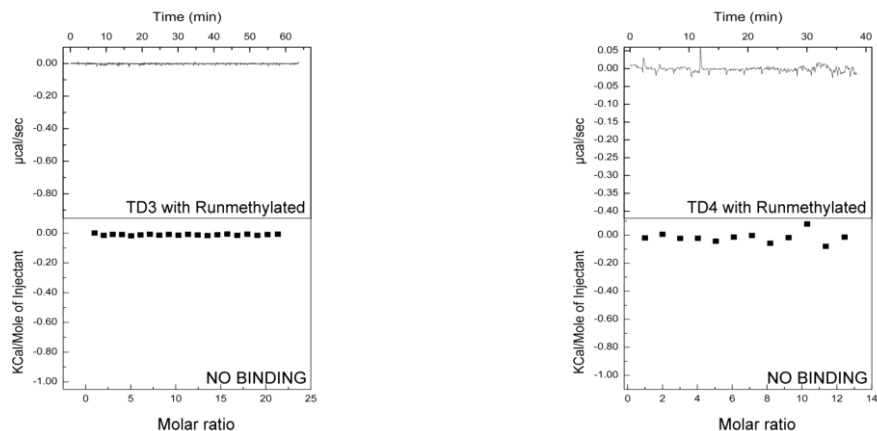


Figure 3.1.25. Representative ITC experiments for the four single eTud domains (TD1, 2, 3 and 4) of TDRD1 with isolated symmetrically dimethylated arginine (Rme2) and with single arginine (Runmethylated). ITC parameters are summarized in [Table 1](#).

Inspection of the TD1 sequence in comparison to the other three TDRD1 domains shows that N325 is found in place of a usually conserved tyrosine (Y774 in TD3) in the aromatic cage ([fig.3.1.26.B](#)). An asparagine at this position is found in mammalian (human, mouse and bovine) TDRD1 but zebrafish and medaka TDRD1 maintain the tyrosine. To test whether this substitution was responsible for the observed low binding to TD1, we made the mutation N325Y in TD1. ITC measurements with an isolated sDMA or with the R45me2 peptide show that this single substitution restores binding comparable to the best of the other eTud domains; the affinity for sDMA being 117 μM and for the R45me2 peptide 47 μM ([fig.3.1.26.A](#), [fig.3.1.27](#)). On the other hand, when the reverse substitution, Y774N, was made in the TD3 domain, binding to the single sDMA was very weak (2.2 mM), although the affinity to the R45me2 peptide was reduced from wild-type but not dramatically (270 μM), showing that the additional affinity for the non-sDMA residues could maintain the binding ([fig.3.1.26.A](#), [fig.3.1.27](#)). This was not the case however for the TD3 mutation Y774A, which failed to bind to either Rme2 or R45me2 ([fig.3.1.27](#)).

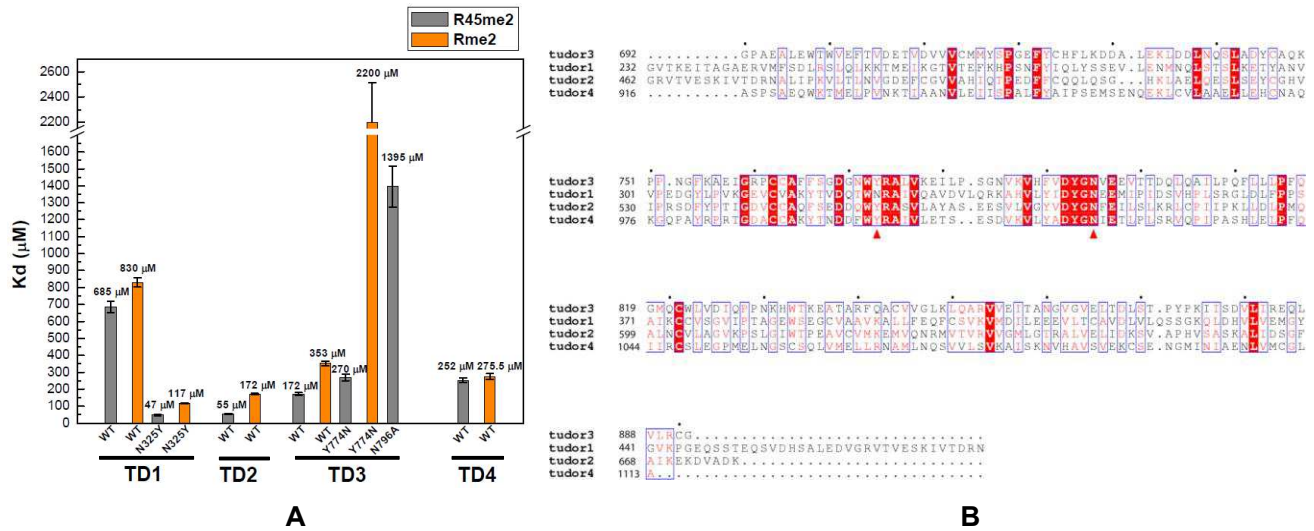


Figure 3.1.26. Active and inactive extended Tudor domains of TDRD1. A. K_D s derived from ITC measurements for the binding of isolated symmetrically dimethylated arginine (Rme2) and R45me2 peptide to individual wild-type (wt) TDRD1 TDs and N325Y, Y774N and N796A mutants. Bars represent mean values of duplicate and triplicate ITC titrations. **B.** Multiple sequence alignment of the four individual Tudor domains of TDRD1. The first red arrow highlights the position of N325 of TD1 and the second arrow the Y774 of the aromatic cage of TD3. Mutations on both residues are critical for binding sDMAs (see A).

The results above show that substitutions of the conserved asparagine (e.g. N796A in TD3) or an aromatic cage tyrosine (e.g. N325 in TD1, Y774A in TD3) drastically reduce sDMA binding rendering the domain effectively inactive. Indeed, the signature motif of ‘active’ Tudor domains appears to be FX_nW(F)YRX_nF(Y)XDY(F)GN, consistent with the fact that all reported crystal structures have these residues. On this basis we predict that at least half of the mouse TDRD protein eTud domains have very low or lack binding affinity for sDMA due to the presence of one or more significant deviations from the ‘active’ motif (fig.3.1.24.A).

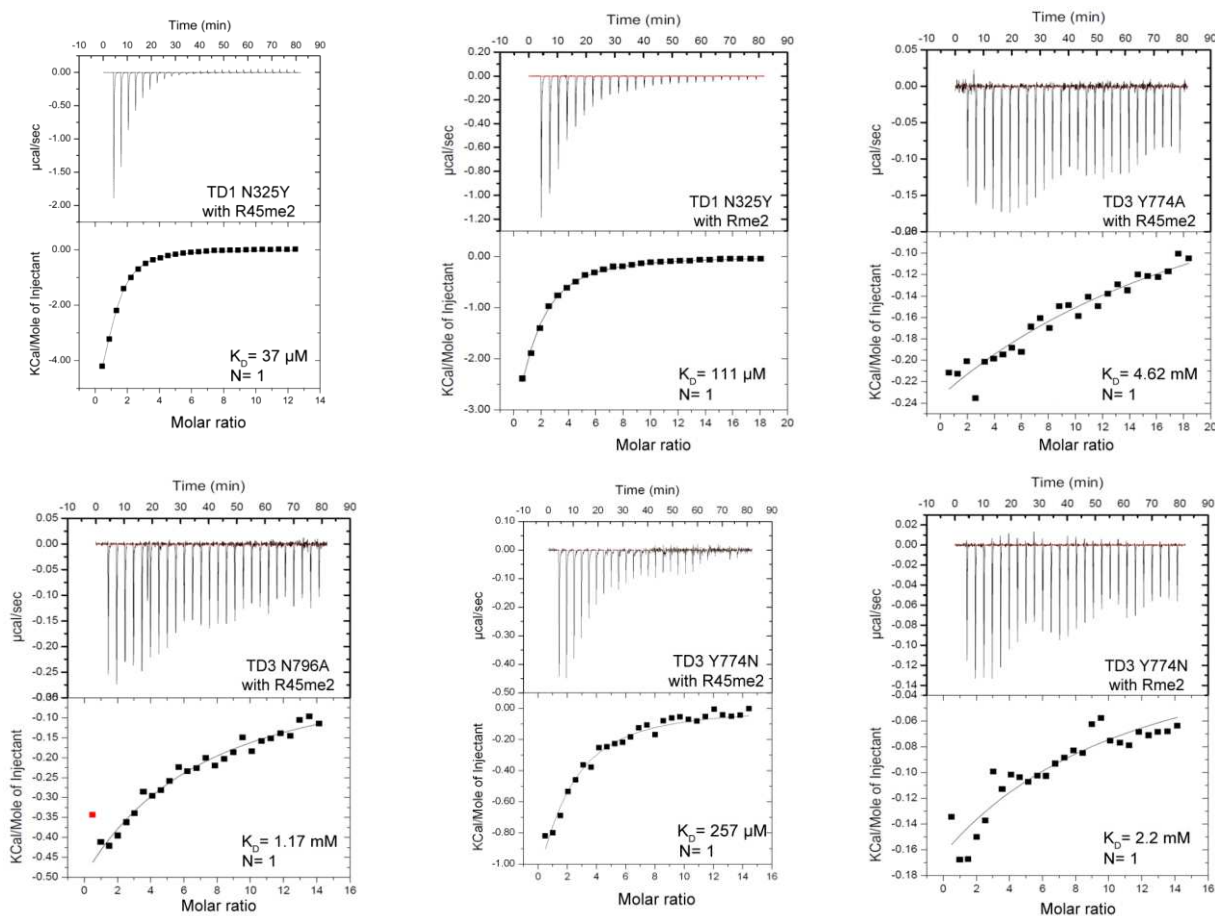


Figure 3.1.27. Single point mutants can render eTud domains of TDRD1 ‘active’ or ‘inactive’ methylarginine binders. Representative ITC experiments for single point mutants of TD1 and TD3 with R45me2 peptide of Mili and Rme2. These mutants either abolish or restore the methylarginine binding activity of the aromatic cage. ITC parameters are summarized in [Table 1](#).

3.1.9.2 - Interaction of endogenous murine proteins by TDRD1 eTudor domains

In order to further investigate whether TD1 is ‘active’ or not in the binding of methylated marks on piRNA pathway proteins, we performed pull-down assays from the lysate of adult mouse testes. We tested the binding of the four tandem eTud domains of TDRD1 (TD1-4) as well as each single eTud domain to endogenous murine proteins Mili and Miwi both of which are reported to contain sDMAs ([fig.3.1.12](#)). Our results show that both proteins can bind TD1-4

and single domains TD2, TD3 and TD4, but only very weakly to TD1 ([fig.3.1.28.A](#)). The germline-specific RNA helicase mouse Vasa Homologue (MVH) is a secondary piRNA biogenesis factor (Kuramochi-Miyagawa et al., 2010) that is reported to carry both symmetrical and asymmetrical dimethyl-arginine modifications (aDMAs) (Kirino et al., 2010). MVH associates with TDRD1 and TDRD6 and is a component of Mili and Miwi complexes (Kirino et al., 2010; Leroy et al., 1989; Tanaka et al., 2000). As shown in [fig.3.1.28.A](#), while MVH was retained on beads coated with TD2, TD3 or TD4, it failed to interact with TD1. Western blotting confirmed the presence of the respective arginine modifications on all the three proteins ([fig.3.1.28.B](#)). We then wanted to explore the binding of the 'active' TD1, carrying the mutation N325Y, with the endogenous PIWI protein Mili. Pull-down experiments indicated that TD1 N325Y could bind Mili, albeit much more weakly than TD2 and TD3 domains. Hence, we could not make safe conclusions about the regain of activity for TD1 carrying this point mutation ([fig.3.1.28.C](#)). It should be noted here that in ITC experiments TD1 N325Y could bind with high affinity single sDMA or Mili peptide, while the endogenous Mili from adult testis carries multiple methylated arginines at its N-terminus ([fig.3.1.12](#)). One possibility is that in the endogenous environment of mouse testis, additional factors could contribute to the binding of Mili with TDRD1, while another potentiality is that the overall fold of TD1 domain does not allow binding of methylated peptides or this specific domain is binding to another partner.

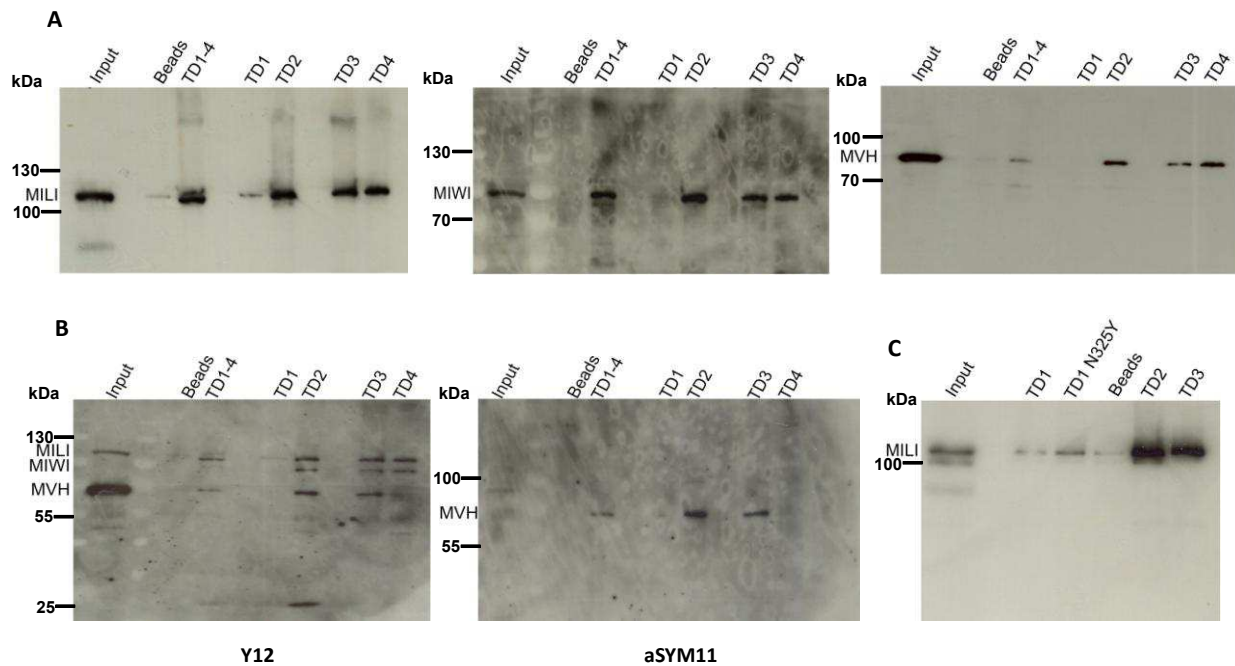


Figure 3.1.28. Active and inactive extended Tudor domains of TDRD1. **A.** Pull-down assays of endogenous murine proteins Mili, Miwi and mouse Vasa homologue (MVH) by individual and multiple TDs of TDRD1. His-tagged constructs with four Tudor domains (TD1-4) and single Tudor domains (TD1, TD2, TD3, TD4) were used. **B.** Detection of methylated arginines in endogenous murine proteins, using antibodies against symmetrical (Y12) and asymmetrical (aSYM11) dimethylated arginines. **C.** Pull-down assay with the TD1 N325Y. Size markers in kilodaltons are indicated.

The observation that apart from sDMAs PIWI proteins and MVH contain monomethylated arginines (MMA) and asymmetrically dimethylated arginines (aDMA) respectively, prompted us to explore their binding affinity to single eTud domains of TDRD1. Single aDMA bound very poorly to all TDs, except for TD2 that showed no binding, while for single MMA we observed low affinity in mM range (fig.3.1.29, fig.3.1.30). Among these titrations, TD3 showed the highest affinity for aDMA (690 μ M) and crystallization trials were set up by mixing the protein with single aDMA in five and ten excess molar ratios. Despite the appearance of crystals after 3 days, they proved to be salt crystals after diffraction experiments. Overall, these results suggest that TDRD1 is recognizing only symmetrical dimethylarginines found on Mili or other PIWI proteins and the ‘inactive’ TD1 is also unable to bind other types of modified arginine, thus having an unknown so far role. It cannot be ruled out that TD1 binds a non-methylated peptide from Mili or some other protein.

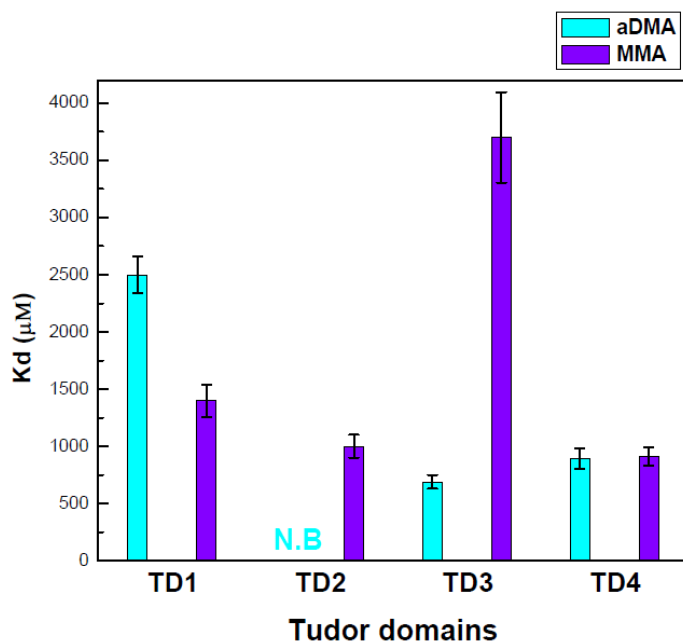
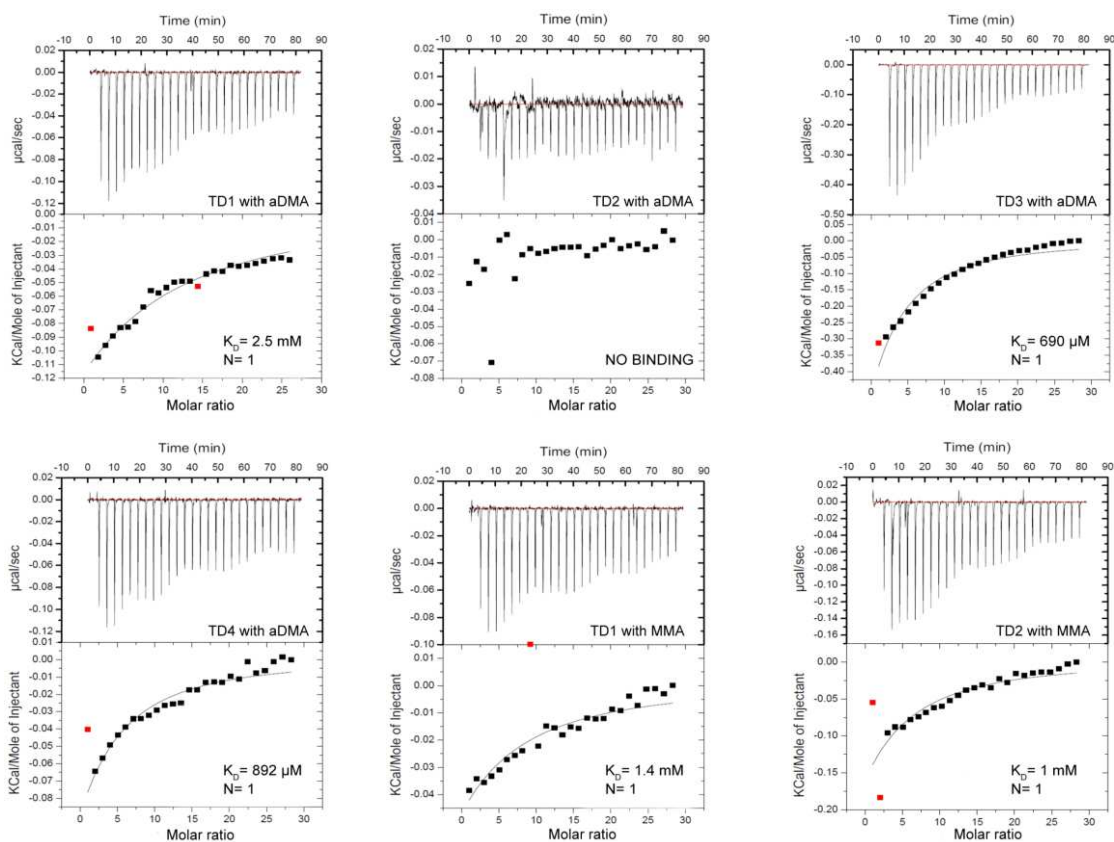


Figure 3.1.29. Single eTud domains of TDRD1 with aDMAs and MMAs. K_D s derived from ITC measurements for the binding of isolated asymmetrically dimethylated arginines (aDMAs) and monomethylated arginines (MMA) to individual TDRD1 TDs.



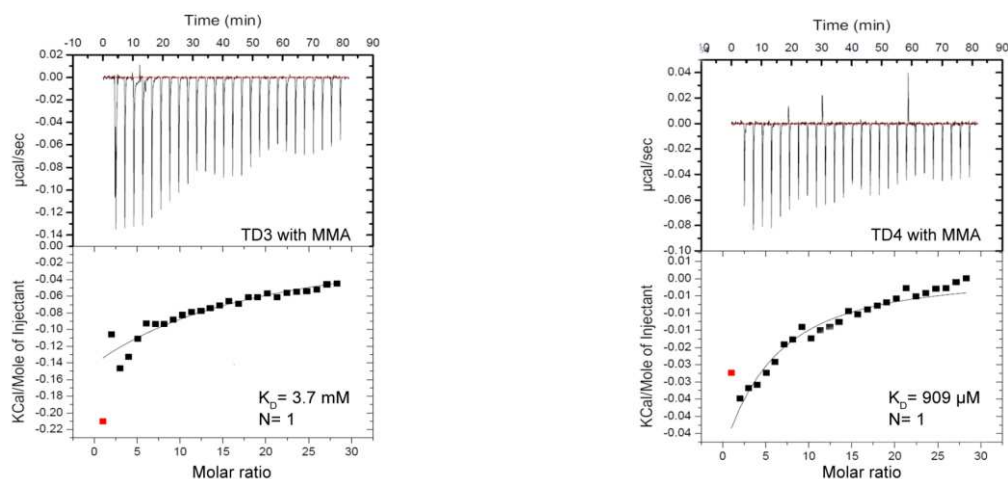


Figure 3.1.30. Representative ITC experiments of the four single eTud domains (TD1, 2, 3 and 4) of TDRD1 with isolated asymmetrically dimethylated arginine (aDMA) and isolated monomethylated arginine (MMA). ITC parameters are summarized in [Table 1](#).

3.1.10 - SAXS analysis of the tandem Tudor domains of TDRD1

Since our considerable efforts to obtain the crystal structure of tandem or single eTudor domains failed, we decided to use small-angle X-ray scattering (SAXS) to determine the overall conformation of the tandem Tudor domains of TDRD1 ([fig.3.1.31](#)). SAXS data was measured on both the four Tudor domain construct (TD1-4) and the tandem TD1 and TD2 domains (construct TD1-2). *Ab initio* modeling was done with DAMMIF and 40 individual models (for both TD1-4 and TD1-2) were averaged using the program DAMAVER. The output of DAMAVER was refined using it as a starting model for the program DAMMIN to yield the final calculated envelope (see [materials and methods](#)). The envelope together with model independent parameters such as radius of gyration and Dmax ([fig.3.1.31.D](#)) show that TD1-4 has an elongated shape with four distinct lobes which are comparable in size to individual eTud domains and arranged end-to-end ([fig.3.1.31.A](#)). The envelope derived from the TD1-2 data is consistent with a double-domain structure similar to either the first two or last two domains in the TD1-4 envelope ([fig.3.1.31.A](#)).

In order to better characterize the SAXS envelope of TD1-4, we performed SAXS-based rigid body modeling, using the program CORAL (Petoukhov and Svergun, 2005). This program translates and rotates the atomic models of individual domains against the SAXS data, with the missing flexible inter-domain linkers being modeled as random chains. Atomic models for the TD1, TD2 and TD4 domains were derived by homology modeling from the crystal structure of the extended Tudor-SN protein from *D. melanogaster* (PDB code 2WAC), while for TD3 the TDRD1 crystal structure reported here was used. Rigid body modeling simultaneously to both the TD1-4 data and TD1-2 consistently resulted in poor fits to the experimental data suggesting that the assumption that TD1-2 and TD3-4 have the same conformation may not be true, for instance due to inter-domain flexibility. Using just the TD1-4 data, the final four domain pseudo-atomic model has a χ^2 of 2.05 (fig.3.1.31.C). The CORAL model fits within the independently derived *ab initio* TD1-4 envelope, although there are some deviations consistent with flexibility and/or incorrect modeling of the linkers (fig.3.1.31.B). To investigate the discrepancy with the TD1-2 data, the sub-complexes TD1-2 and TD3-4, derived from the TD1-4 pseudo-atomic model, were compared to the TD1-2 data with Crysol and Oligomer. This showed that individually they do not match the TD1-2 data but that a combination of both conformations (61 % and 39% TD1-2 and TD3-4 respectively) matches the experimental data ($\chi^2 = 2.5$) (fig.3.1.31.C). This shows that within the limits of the analysis the elongated flexible model presented for TD1-4 is consistent with data for the TD1-2.

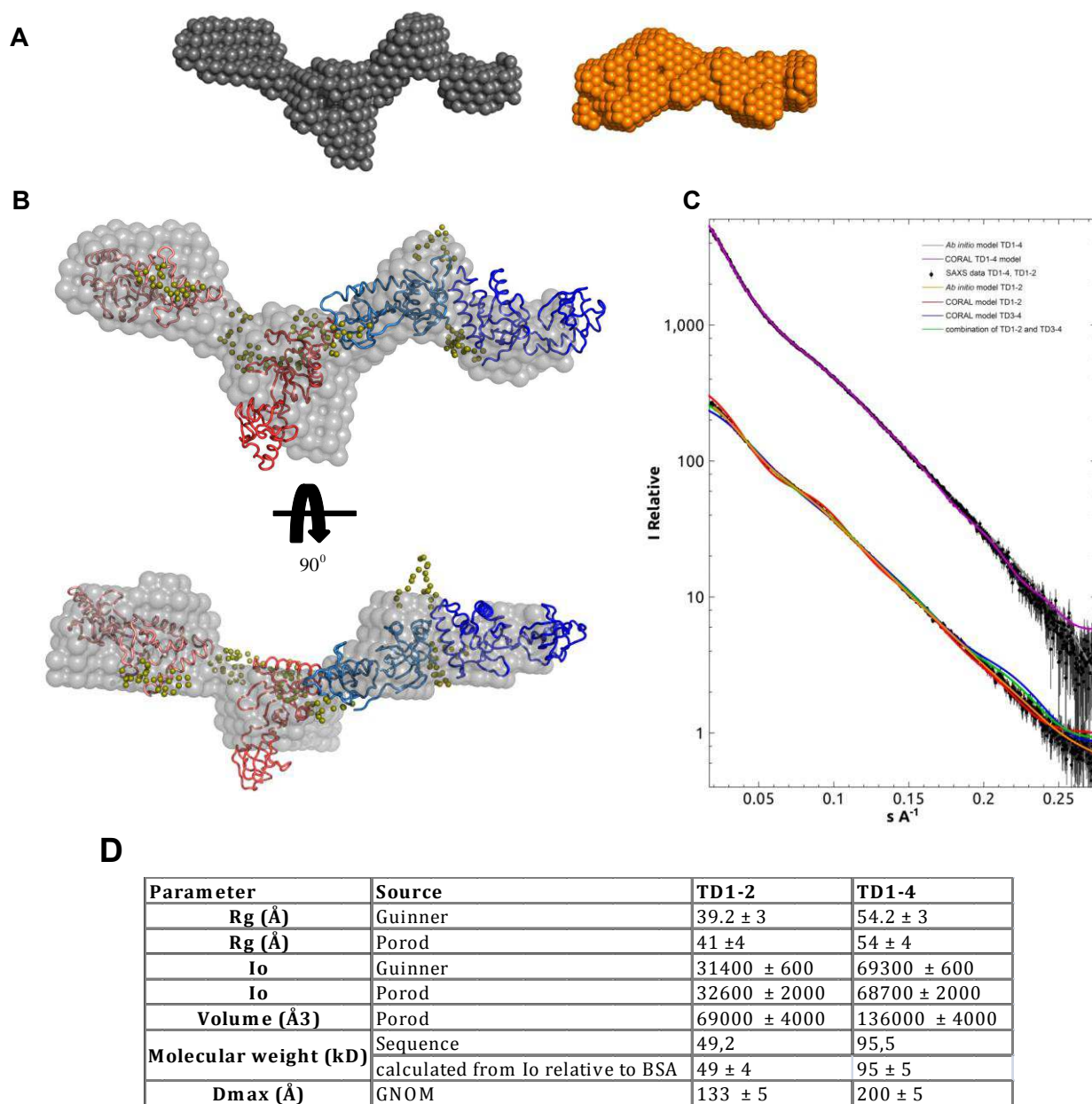


Figure 3.1.31. Architecture of the TD1-4 domains of TDRD1 as determined from small-angle scattering data. A. *Ab initio* models of TD1-4 (232-1094 aa) (grey) and TD1-2 (232-677 aa) (orange) derived from SAXS data. The TD1-2 SAXS data and model are consistent with either half of the TD1-4 model. **B.** Individual eTud domains were fitted to the TD1-4 *ab initio* model (grey envelope) using CORAL. TD1 (salmon), TD2 (light red) and TD4 (deep blue) models were based on Tudor-SN (PDB code 2WAC). TD3 (light blue) was fitted using the TDRD1 crystal structure determined here. Flexible linkers connecting the eTud domains, positioned by CORAL, are depicted as spheres (olive green). **C.** SAXS data (black dots with error bars) and fits for TD1-4 and TD1-2. **Top:** TD1-4 data. Fit of *ab initio* model shown in (A) (grey line) and the multi-domain model created by CORAL (purple line). **Bottom:** TD1-2 data. Fit of *ab initio* model shown in (B) (orange line). Fits of the tandem TD1-2 (red line) and TD3-4 (blue line) domain models taken from the CORAL model (shown in A) show systematic deviations. A linear combination of these two models (green line) better represents the data ($\chi^2 = 2.5$), suggesting flexibility. **D.** Model independent parameters derived from SAXS data.

Tudor domain	Methylated ligand	Ka (1/M)	Kd (µM)	ΔG (kcal/mol)	ΔH (kcal/mol)	minus TAS	ΔS (cal/mol/K)
TD1	R45me2	1.03E3 ± 53.2	685.5	-4,3685	-1,984	-2,405	8,06
TD1	R45me2	2.65E4 ± 1.54E3	47.5	-5,896	-6,361	0,465	-1,53
TD2	R45me2	1.81E4 ± 1.05E3	55	-5,8055	-3,1935	-2,6145	8,78
TD3	R45me2	5.18E3 ± 103	172	-5,134	-6,759	1,6125	-5,44
TD3	R45me2	3.89E3 ± 307	270	-4,865	-4,239	-0,629	2,11
TD3	R45me2	216 ± 27.7	4620	-3,182	-18,05	14,877	-49,9
TD3	R45me2	851 ± 55.1	1395	-3,898	-8,9515	5,0535	-16,95
TD4	R45me2		252,5	-49,065	-45085	0,401	1,345
TD1	R74me2	1.07E3 ± 31.9	967	-4,0085	-5,436	1,426	-4,79
TD2	R74me2	2.49E3 ± 46.2	417	-4,6065	-5,7805	1,17	-3,925
TD3	R74me2	2.82E4 ± 1.87E3	35	-6,067	-2,639	-3,428	11,5
TD4	R74me2	2.90E3 ± 81	560	-4,48	-5,0965	0,612	-2,05
TD2-3	R74me2	1) 2.00E4 ± 2.0E3 2) 3.02E3 ± 1.3E2	1) 50 2) 331	1) -5.863 2) -4.744	1) -4.365 2) -4.99	1) -1.502 2) 0.251	1) 5.04 2) -0.844
TD3-4	R74me2	1) 1.94E4 ± 2.6E3 2) 2.17E3 ± 1.4E2	1) 51 2) 495	1) -5.845 2) -4.54	1) -5.2 2) -7.43	1) -0.649 2) 2.8	1) 2.18 2) -9.65
TD1	R83me2	369 ± 12.8	2350	-3,575	-13,69	10,107	-33,9
TD2	R83me2	1.42E3 ± 119	641	-4,355	-4,785	0,426	-1,43
TD3	R83me2	4.64E3 ± 111	245	-4,925	-6,487	1,557	-5,225
TD4	R83me2	1.26E3 ± 53.5	831	-4,199	-4,518	0,316	-1,06
TD1	sDMA	1.12E3 ± 36.8	830.5	-4,2	-7,132	2,924	-9,8
TD1	sDMA	8.97E3 ± 278	117	-5,3	-8,268	2,9	-9,7
TD2	sDMA	5.99E3 ± 187	172	-5,12	-4,11	-1,01	3,41
TD3	sDMA	2.83E3 ± 155	353	-4,7	-8,13	3,42	-11,5
TD3	sDMA	454 ± 75.2	2200	-3,6	-4,85	12,25	-4,11
TD4	sDMA	3.76E3 ± 174	275.5	-4,8	-6,43	1,57	-5,29
TD1	adMA	400 ± 26.6	2500	-3,5	-3,5	-10,88	-0,03
TD2	adMA	NO BINDING	NO BINDING				
TD3	adMA	1.45E3 ± 136	690	-4,3	-4,2	-0,03	0,11
TD4	adMA	1.12E3 ± 122	892	-4,1	-1,06	-3,1	10,4
TD1	MMA	707 ± 88.8	1400	-3,8	-0,88	-3,01	10,1
TD2	MMA	988 ± 122	1000	-4,08	-2,15	-1,9	6,48
TD3	MMA	266 ± 36.4	3700	-3,3	-7,04	-3,7	-12,5
TD4	MMA	1.10E3 ± 103	909	-4,14	-0,613	-3,54	11,9
TD2-3	R4574me2s	2.06E4 ± 1.64E3	48.5	-5,8	-7,7	1,84	-6,19
TD3-4	R4574me2s	3.42E4 ± 1.07E3	29	-6,18	-13,98	7,78	-26,1
TD2	R4574me2s	2.57E4 ± 5.71E3	39	-6,012	-2,92	-3,1	10,4
TD3	R4574me2s	1.50E4 ± 628	66,6	-5,69	-10,3	4,62	-15,5

Table 1. Parameters from ITC data fitting. All values derive from duplicate ITC experiments. Ka values derive from the single representative experiments shown in fig. 3.1.14-16, 3.1.25, 3.1.27, 3.1.30.

3.2 - Characterization of the MYND domain of TDRD1 as a protein-protein interaction module.

3.2.1 - Introduction

The basis for the research work presented in this chapter lies on the characterization of FK506-binding protein FKBP6's interaction with TDRD1 protein. FKBP6 was identified recently in Ramesh Pillai's group (EMBL, Grenoble Outstation) as a novel component in the piRNA pathway. In addition, it was shown that the MYND domain of TDRD1 is involved in the binding of FKBP6 protein and the impact of this interaction remains unknown (Xiol et al., submitted). The preliminary results presented below focus on the characterization of the MYND domain-FKBP6 interaction in the piRNA pathway, using biochemical and biophysical approaches.

3.2.1.1 - FK506-binding proteins (FKBPs)

Proline is the only amino acid that can exist in both *cis* and *trans* conformations. The change between these two states is a rather slow process that is accelerated by enzymes known as prolyl *cis-trans* isomerases (PPlases) or rotamases (reviewed in Lu et al., 2007). FKBP6 belongs to the proteins that catalyze this reaction along with the unstructurally related family of cyclophilins, which are targets of immunosuppressive drugs that are commonly used to prevent rejection in organ transplantation (Fischer et al., 1989; Harding et al., 1989; Siekierka et al., 1989; Takahashi et al., 1989). Prolyl *cis-trans* isomerization events are commonly found in a broad range of cellular pathways, providing a backbone switch that can be used for protein folding and regulation. The isomerization can mediate conformational changes, either by causing a change in the accessibility of a given residue for post-translational modification or by modifying the spatial orientation of an entire domain (Nelson et al., 2006; Wang et al., 2010; Zhou et al., 2000).

FKBPs have been reported to have some functions that are independent of their catalytic activity. FKBP12, the first identified member in the family, regulates palmitoylation of h-Ras through prolyl *cis-trans* isomerization (Ahearn et al., 2011), but does not require this activity for

binding ryanodine receptor (Ryr) calcium release channel and helping to maintain it in a closed conformation (Timerman et al., 1996; Timerman et al., 1995). Some members of the FKBP family (including FKBP6) contain a tetratricopeptide repeat (TPR) domain at their C-termini. TPR domains are found in numerous proteins and they serve as interaction modules and multiprotein complex mediators, involved in diverse biological processes (reviewed by (Zeytuni and Zarivach, 2012). Some TPR motifs can recognize the C-terminal petapeptide MEEVD of the molecular chaperone HSP90 (Young et al., 1998) (fig.3.2.2.C). As a result, these FKBP6s could act as co-chaperones and few demonstrations have shown that the isomerase activity may not be required in this context (Ma and Whitlock, 1997; Meyer et al., 1998; Riggs et al., 2003).

Hsp90 has already been shown to play a role in small RNA pathways, both in PIWI and Ago complexes. Various studies indicated that Hsp90 together with Hsc70 mediates loading of small RNA duplexes into Ago protein in an ATP-dependent fashion in plants, humans and *Drosophila* (Iki et al., 2010; Iwasaki et al., 2010; Miyoshi et al., 2010), while target cleavage is Hsp90 and ATP-independent. Hsp90 is thought to have a role in protecting Argonaute proteins from degradation by the proteasome, as human Ago2 is destabilized and mislocalized upon Hsp90 inhibition (Johnston et al., 2010; Pare et al., 2009; Suzuki et al., 2009). In addition, there are links between Hsp90 and the piRNA pathway. More specifically, the silencing of Stellate in *Drosophila's* male germ line is defective in Hsp83 (*Drosophila's* Hsp90) and transposons are derepressed in both testes and ovaries (Specchia et al., 2010). Moreover, Hsp90 and the co-chaperone Hsp70/Hsp90 Organizing Protein (HOP) interact with Piwi *in vivo* and influence Piwi's phosphorylation status (Gangaraju et al., 2011). Taken altogether, it is suggested that Hsp90 may have a role in the piRNA pathway, although not as defined as for the miRNA and siRNA pathways.

3.2.1.2 - FKBP6 is a novel piRNA pathway-associated factor

A yeast two-hybrid screen for the N-terminal part of TDRD1 (residues 1-232), containing the MYND domain, identified a fragment of FKBP8 including its TPR domain (Pillai lab, unpublished data). In parallel, proteomic studies on endogenous Mili and Miwi2 complexes identified the presence of FKBP3, FKBP5 and FKBP6 (Vagin et al., 2009). FKBP8 participates in neural tube

development (Wong et al., 2008), while FKBP6, like TDRD1, is required for progression of meiosis in mouse males (Chuma et al., 2006; Crackower et al., 2003) and interacts with *in vitro* translated Tdrd1 (fig.3.2.1.A) and Mili (Xiol et al., submitted). Furthermore, Line1 retrotransposon expression is derepressed in *Fkbp6* mutant mice and DNA methylation in Line1 promoters is lost (Xiol et al., submitted). Finally, there are striking similarities in the expression pattern of FKBP6 orthologue in *Drosophila*, Shut-down (Munn and Steward, 2000) and the PIWI protein Aub, indicating participation in a shared pathway. These results suggest that FKBP6 participates in the piRNA pathway.

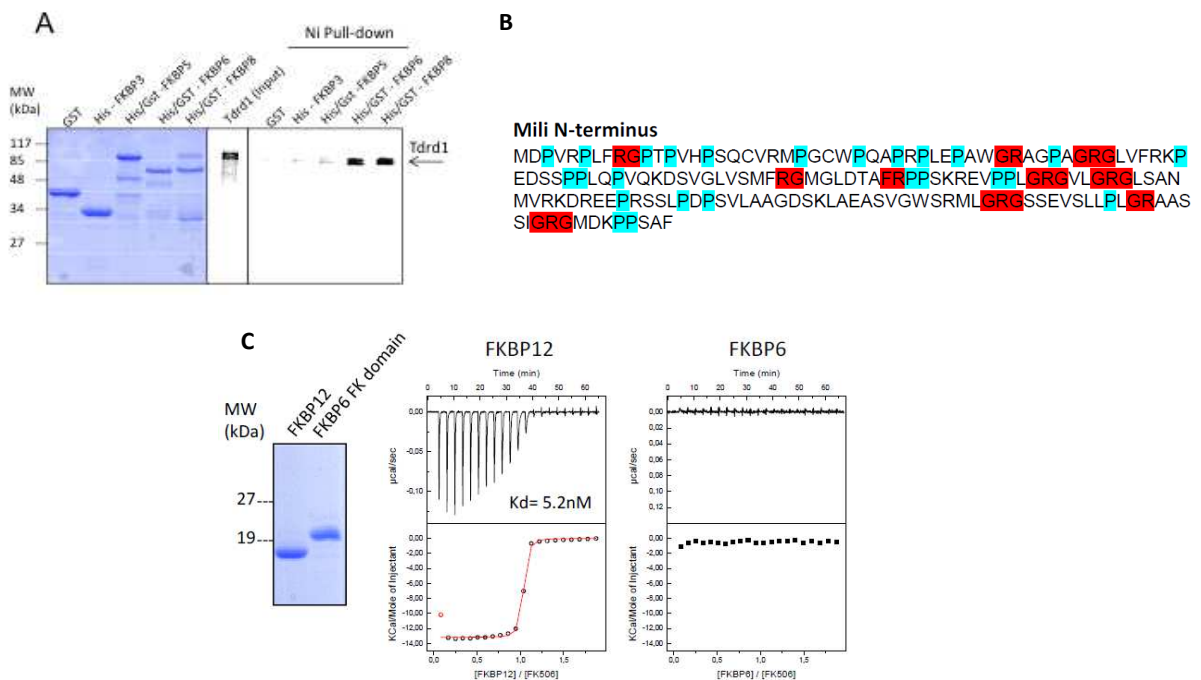


Figure 3.2.1. FKBP6 interacts with TDRD1 and is inactive as an isomerase (Xiol et al., submitted). **A.** Pull down assays between recombinant FKBP6s and *in vitro* translated TDRD1. GST was used as control. **B.** Amino acid sequence of Mili N-terminus. RG repeats, potential methylation sites, are highlighted in red, while all the prolines are highlighted in cyan. **C.** ITC measurements of FKBP12 and FKBP6 binding to FK506. 1 mM FKBP was added to 10 μ M of FK506. In the lower panels, the circles represent experimental data and the continuous line corresponds to the best fit to a model with one binding site.

Analysis of the piRNA profile levels in an *Fkbp6* mutant revealed that loss of FKBP6 has no effect on primary biogenesis, since piRNAs are still present in Mili complexes, even though the overall

profile seems to be altered as a consequence of Line1 activation (Xiol et al., submitted). Recovered Mili and Miwi2 RNPs from new born (P0) control and *Fkbp6* mutant animals were obtained and analysis of their associated piRNAs followed, showing that although Mili remained associated with piRNAs in the *Fkbp6* mutant, Miwi2-bound species were drastically reduced. Also, in the *Fkbp6* mutant gonocytes, accumulation of Miwi2 in the nucleus was reduced, while the enrichment of Tdrd1 and Mili in pi-bodies remains unaffected in the mutant (Xiol et al., submitted). Overall, these observations suggest that Mili remains competent to initiate biogenesis of Miwi2 piRNAs, but the subsequent poorly-understood steps, including loading of Miwi2 fails in the absence of *Fkbp6*.

3.2.1.3 - FKBP6 as a co-chaperone

FKBP6 is a protein of 37 kDa and is comprised of an N-terminal FK isomerase domain (19-145 aa) and a C-terminal TPR domain (145-301 aa). Considering the TPR-Hsp90 association in some FKBP proteins, it was tested whether Hsp90 could be recruited by the TPR domain of FKBP6. Multiple sequence alignment with TPR domains revealed that residues involved in Hsp90 interaction are conserved in FKBP6. More specifically, an absolutely conserved lysine was previously shown to be indispensable for FKBP52-Hsp90 interaction (Riggs et al., 2003). (fig.3.2.2.A). After design of deletion constructs of FKBP6 and K254A Hsp90 mutants, Western blot analysis with antibodies against endogenous Hsp90 revealed specific association of FKBP6's TPR domain to Hsp90, which is abolished upon mutation of K254 (fig.3.2.2.B). This confirms that FKBP6 could have a co-chaperone role by bridging PIWI complexes to Hsp90 (Xiol et al., submitted).

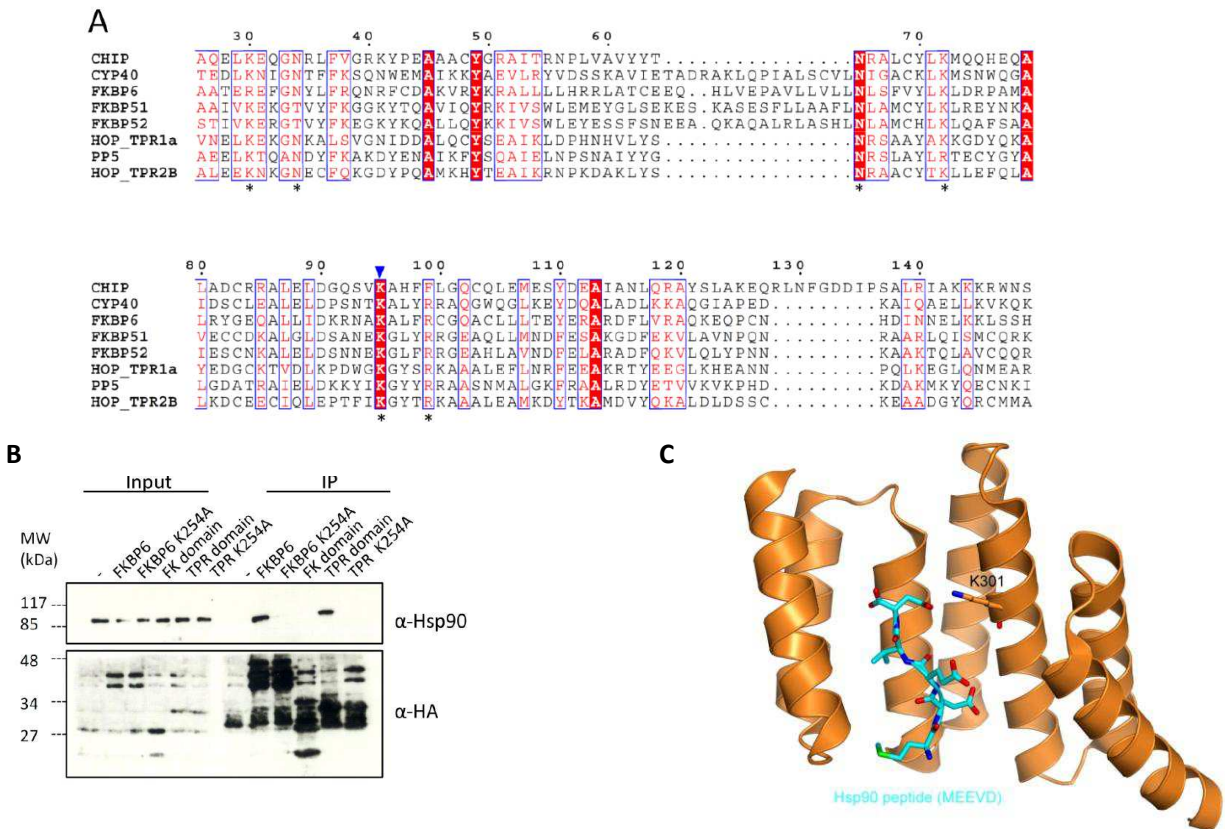


Figure 3.2.2. FKBP6 interacts with Hsp90. **A.** Sequence alignment of TPR domains from the indicated proteins. All the sequences are from human origin. Asterisks indicate residues that participate in interaction with Hsp90. Mutation of the lysine pointed with a blue triangle abrogates binding to Hsp90. **B.** HA-FKBP6 or the shown deleted versions (including K254A point mutants) were expressed in HEK293 cells and immunoprecipitated using an anti-HA affinity matrix. Interaction with Hsp90 was detected by Western blot, using antibodies against endogenous Hsp90 (Xiol et al., submitted). **C.** Cartoon representation of the TPR2a domain of HOP protein structure (Scheufler et al., 2000, PDB 1ELR) (orange) with the MEEVD C-terminal peptide of Hsp90 (cyan, sticks). The lysine K301 that is essential for this interaction is also shown.

3.2.1.4 - The FK domain of FKBP6 is inactive as an isomerase

As described before, the N-terminus of Mili contains many RG/RA repeats that are modified by symmetrical dimethylation on arginines (Vagin et al., 2009). Interestingly, apart from multiple arginines, it was noted that abundant prolines are also present between these arginines (fig.3.2.1.B). In order to test if FKBP6's catalytic domain interacts with these residues, a prolyl *cis-trans* isomerization assay was set up using full-length FKBP6 and the FK domain alone. Surprisingly, none of the proteins displayed activity, while positive controls (FKBP3, 5, and 12) were highly active (Xiol et al., submitted). Moreover, some active FKBP6s are inhibited by FK506, an immunosuppressive drug that binds the catalytic pocket (Kang et al., 2008; Van Duyne et al.,

FKBP6 protein (Walker et al., Structural Genomics Consortium, 2007, PDB 3BFX). Residues forming the binding pocket are highlighted in green stick model and the PPLIP motif is depicted in pink.

3.2.1.5 - Description of MYND domains

The MYND domain (myeloid translocation protein 8, Nery, DEAF-1) is a ~60 amino acid cysteine-rich structure present in proteins generally implicated in transcriptional repression and association with cancer (Gross and McGinnis, 1996; Huggenvik et al., 1998). It is defined by seven conserved cysteine residues and a single histidine residue that are coordinating the binding of two zinc atoms and are structurally homologous to the PHD and RING finger families of proteins as revealed by structures solved by NMR ([fig.3.2.4.A-B](#)) (Liu et al., 2007).

MYND domains have been shown to participate in protein-protein interactions. Among the MYND domain-containing proteins, BS69 has been shown to be a co-repressor that associates with various transcription factors (Ansieau and Leutz, 2002). In addition, the MYND domain of BS69 associates with the viral proteins E1A and EBNA2 by specifically recognizing their PXLXP peptide motif (X=any amino acid) (Ansieau and Leutz, 2002; Hughes-Davies et al., 2003). However, it has been reported a diversification in MYND domain binding properties, since there are examples of MYND domain-binding partners that do not have the PXLXP motif (Sugawara et al., 2001). AML1/ETO's MYND domain interacts specifically with the co-repressors SMRT and N-COR, contributing to repression of proliferation and differentiation of primary bone marrow cells (Liu et al., 2007). This interaction is mediated by the PPPLI motif in SMRT and N-CoR, which resembles the PXLXP motif, suggesting that a proline-rich region can contribute to the recruitment of a MYND domain ([fig.3.2.4.B](#)).

In order to gain more insights into the MYND domain-FKBP6 interaction, in a multiple sequence alignment of FKBP proteins we observed the presence of PXLXP motif at the FK domain of FKBP6 mouse and human proteins ([fig.3.2.3.A](#)). The crystal structure of human FKBP6 (Walker et al., Structural Genomics Consortium, 2007) reveals that the PPLIP motif is exposed at the surface of the protein, suggesting that it can be involved in the interaction with the MYND domain of TDRD1 ([fig.3.2.3.B](#)). Finally, an uncharacterized *Drosophila* Tudor protein containing

a MYND domain (CG9925) was reported to interact with a putative PPIase from the cyclophilin family (CG3511) in a genome wide two-hybrid screen (Giot et al., 2003). The results presented below, focus on the identification of soluble fragments of MYND domain of TDRD1 and interaction experiments with the FKBP6 protein.

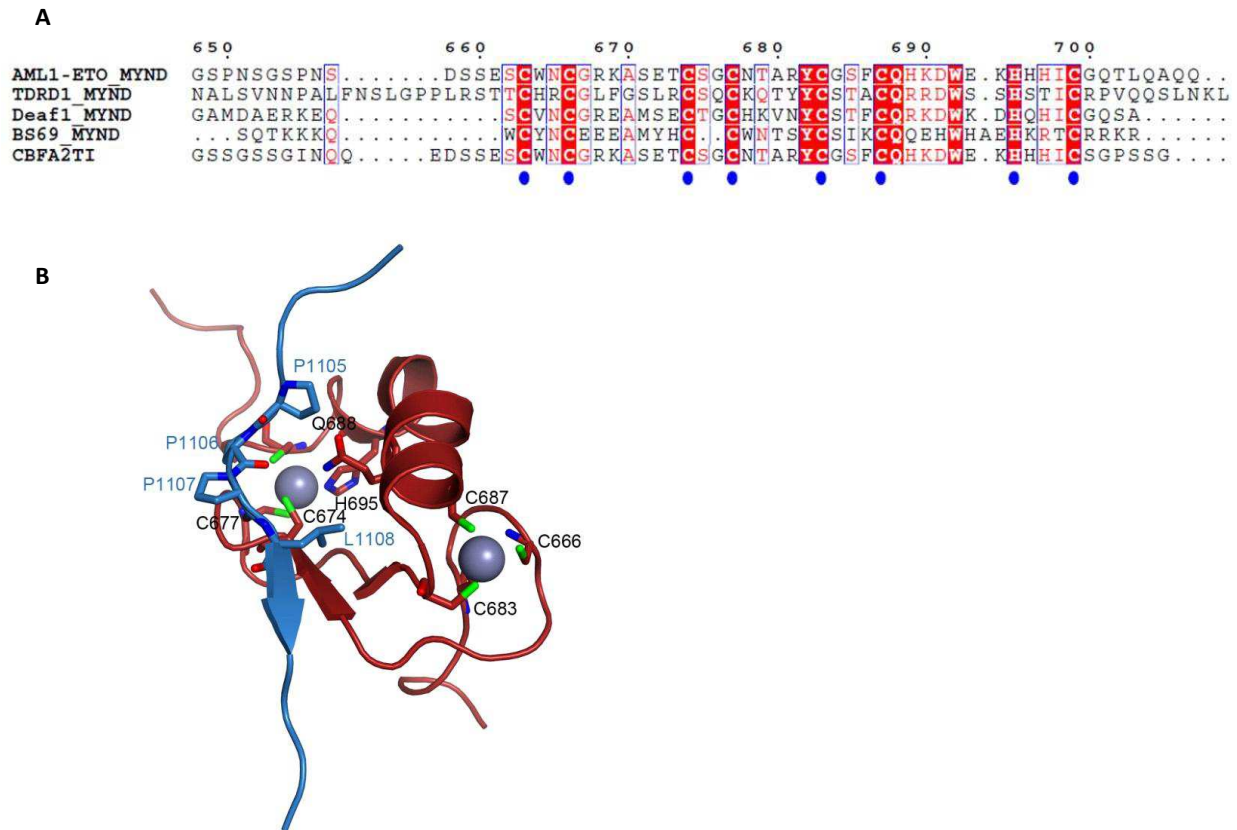


Figure 3.2.4. The MYND domains are zinc finger binding domains. A. Multiple sequence alignment of the MYND domains from different proteins. Blue dots indicate the absolutely conserved residues involved in the binding of zinc atoms. (AML1-ETO: PDB 20D1, Liu et al., 2007, Deaf1: PDB 2JW6, Spadaccini et al., 2006. BS69: MYND domain of the BS69 protein. CBFA2TI: PDB 2DJ8: Solution structure of the zinc finger MYND domain of protein CBFA2TI (Niraula et al., RIKEN Structural Genomics/Proteomics Initiative, 2006). **B.** Solution structure of the AML1-ETO MYND domain with the peptide of SMRT bound (Liu et al., 2007, PDB 2ODD). The two zinc atoms are represented as spheres and are coordinated by cysteines and a histidine residue. In blue stick model, the PPPLI motif of SMRT is shown. The absolutely conserved Q688 is forming hydrogen bond to the SMRT P1106.

3.2.2 – Results

3.2.2.1 – Expression and purification of FKBP6 and TDRD1's N-terminal fragments

As a first step for the characterization of MYND-FKBP6 interaction, we tried to identify soluble fragments of the N-terminus of TDRD1 including the bait region used in the yeast two-hybrid screen (residues 1-232). Using prediction tools, we expressed various constructs and as it can be seen from the [fig. 3.2.5](#), most of them had a low solubility yield or were forming soluble aggregates. In order to increase the solubility, the purification protocol (see [Materials and Methods](#)) was slightly altered by varying the concentration of NaCl up to 500 mM. In parallel, addition of ZnCl₂ in various concentrations and reducing agent in high molarity (to disrupt disulfide bonds between cysteines) did not alter the solubility profile. The observation that the behavior of the N-terminal fragments did not improve suggests that apart from the ~60 amino acid region corresponding to the MYND domain (130-200 aa), the rest may be unstructured. As a result, we were able to identify a soluble fragment corresponding to the MYND domain (residues 152-232) that was monomeric in solution as it can be seen from its size-exclusion chromatograph ([fig.3.2.6.D-E](#)), even though it was prone to precipitation. The best behaved constructs of the N-terminal region of TDRD1 (constructs MYND#2, 5 and 12) was used for interaction studies as it is described below.

Construct	Amino acids	Solubility	Crystallization robot
MYND#1	1-232	low solubility (most as soluble aggregates)	
MYND#2	56-232	low solubility (most as soluble aggregates)	
MYND#3	30-210	low solubility (most as soluble aggregates)	
MYND#4	56-210	Soluble aggregates	
MYND#5	127-232	Soluble aggregates	
MYND#6	1-196	Soluble aggregates	
MYND#7	56-196	insoluble	
MYND#8	30-170	Soluble aggregates	
MYND#9	30-184	Soluble aggregates	
MYND#10	127-196	Insoluble	
MYND#11	152-196	Insoluble	
MYND#12	152-232	soluble	

Figure 3.2.5. Constructs comprised of N-terminal fragments of TDRD1. In this table are listed the various protein fragments expressed and purified.

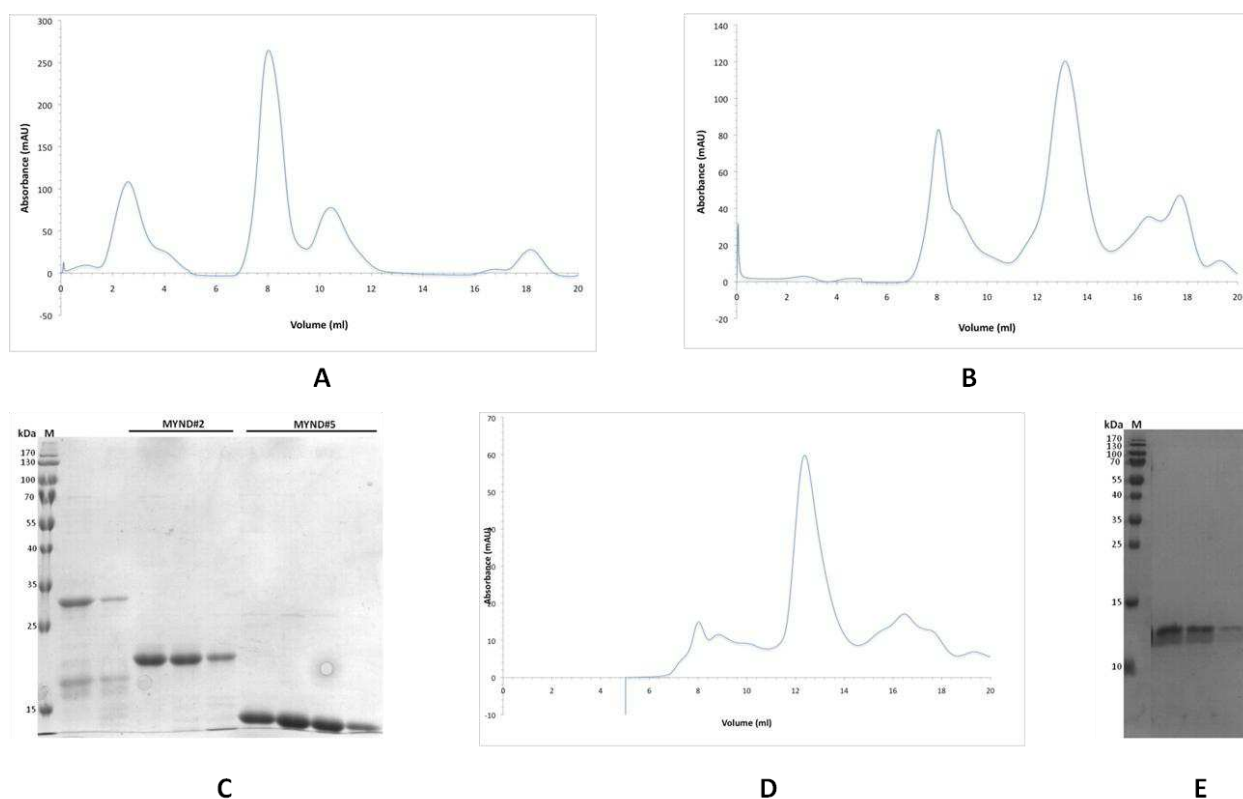


Figure 3.2.6. Purification of N-terminal fragments of TDRD1. **A.** Size exclusion chromatography profile of MYND#2 (56-232 aa) on a S75 gel filtration column (Blue line: Absorbance at 280 nm) **B.** Size exclusion chromatography profile of MYND#5 (127-232 aa) on a S75 gel filtration column (Blue line: Absorbance at 280 nm). **C.** SDS-PAGE of the peak fractions for MYND#2 and MYND#5. **D.** Size exclusion chromatography profile of MYND#12 (152-232 aa) on a S75 gel filtration column (Blue line: Absorbance at 280 nm). **E.** SDS-PAGE of the peak fractions for MYND#12. Size marker in kilodaltons (kDa) is indicated.

Various constructs of the FKBP6 protein were expressed in high amounts at Pillai's group (EMBL, Grenoble outstation). Protein fragment with residues 19-301 (FKBP6) including both the FK and TPR domains was homogeneous only in high salt concentration during purification, while in standard low salt trials, it was rapidly degraded to its two domains (fig.3.2.7.A). The FK domain (19-145 residues) was monomeric in solution in all salt concentrations while the TPR domain (145-301 residues) had a better solubility profile in 500 mM NaCl (fig.3.2.7.B-D). Crystallization trials for TPR domain in high salt at 20 °C and 4 °C gave no positive hits.

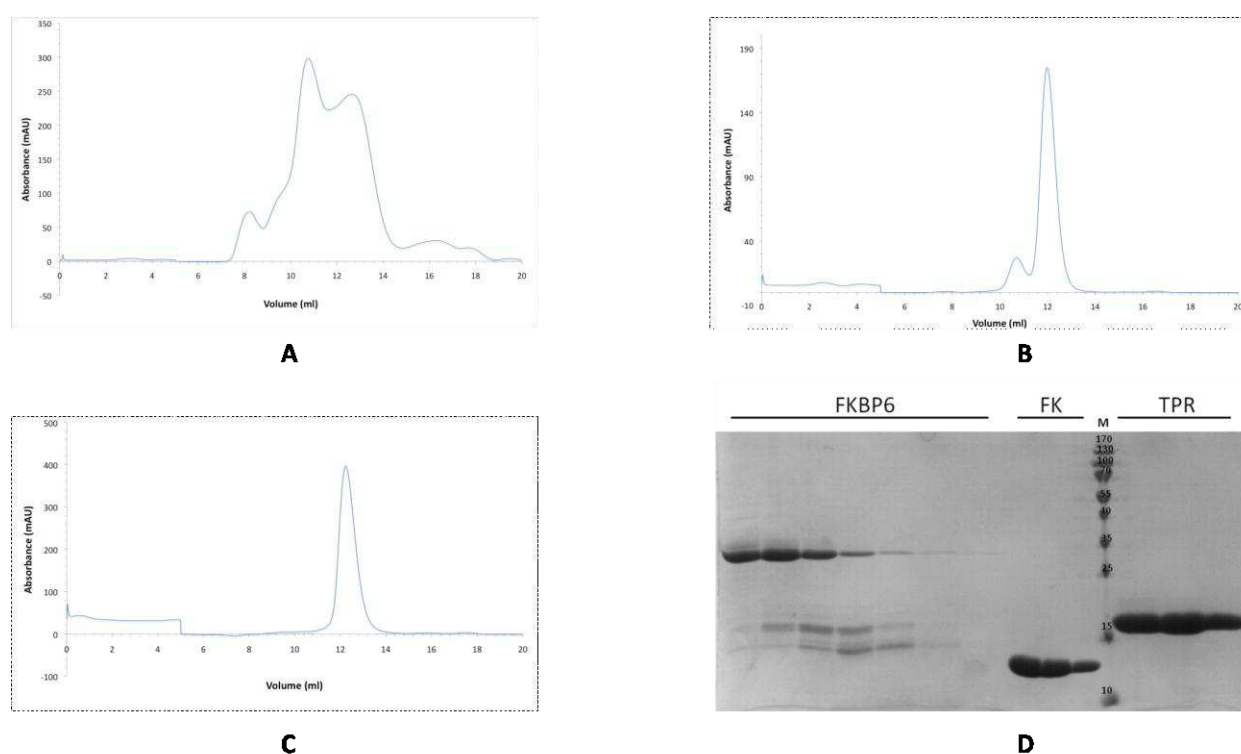


Figure 3.2.7. Purification of FKBP6 protein. A. Size exclusion chromatography profile of FKBP6 (19-301 aa) on a S200 gel filtration column (Blue line: Absorbance at 280 nm) B. Size exclusion chromatography profile of FK domain (19-145 aa) on a S75 gel filtration column (Blue line: Absorbance at 280 nm). C. Size exclusion chromatography profile of TPR domain (145-301 aa) on a S75 gel filtration column in 500 mM NaCl (Blue line: Absorbance at 280 nm). D. SDS-PAGE of the peak fractions for FKBP6, FK and TPR domains. Size marker in kilodaltons (kDa) is indicated.

3.2.2.2 - The MYND domain of TDRD1 interacts with the FK domain of FKBP6

In order to test the hypothesis that the N-terminal region of TDRD1 interacts with the FKBP6, we performed pull-down assays using recombinantly expressed proteins (see [Materials and Methods](#)). MYND#2 fragment (56-232 residues) interacted strongly with the FK domain and possibly with the TPR domain, since they have similar molecular weight ([fig.3.2.8.A](#)). To overcome this problem, the MYND#5 construct was used for pull down assays, having a smaller size. As it can be seen in [fig.3.2.8.B](#), it also interacts with the FK domain of FKBP6 and less with the TPR domain, suggesting that probably the N-terminal part of FKBP6 is binding the N-terminal region of TDRD1. However, Isothermal Titration Calorimetry (ITC) experiments between the two domains showed no interaction (data not shown). One possible reason is the low amounts of MYND#5 that can be obtained after purification, since the maximum soluble fraction corresponds to 1 mg/ml.

The protein fragment including only the MYND domain (MYND#12, 152-232 aa) also bound to the FK domain but showed no interaction with the TPR domain ([fig.3.2.8.C](#)). This interaction proved to be specific, since the eTudor 1 domain (TD1, 232-476 aa) showed no association with neither the FK nor the TPR domains. A further characterization of this interaction is not progressed, since these two proteins are not co-eluting in a size-exclusion chromatography experiment. The rapid precipitation of the MYND domain-containing fragment and the subsequent low solubility levels have not allowed us so far to gain more insights into its binding to the FK domain. Finally, we tried to co-express the two proteins, with the various constructs of FKBP6 having an N-terminal his tag (in pETM-11 vector) and MYND domain protein fragments having no affinity tag (in pCDF-Duet vector). Despite considerable efforts, there was no expression of the proteins.

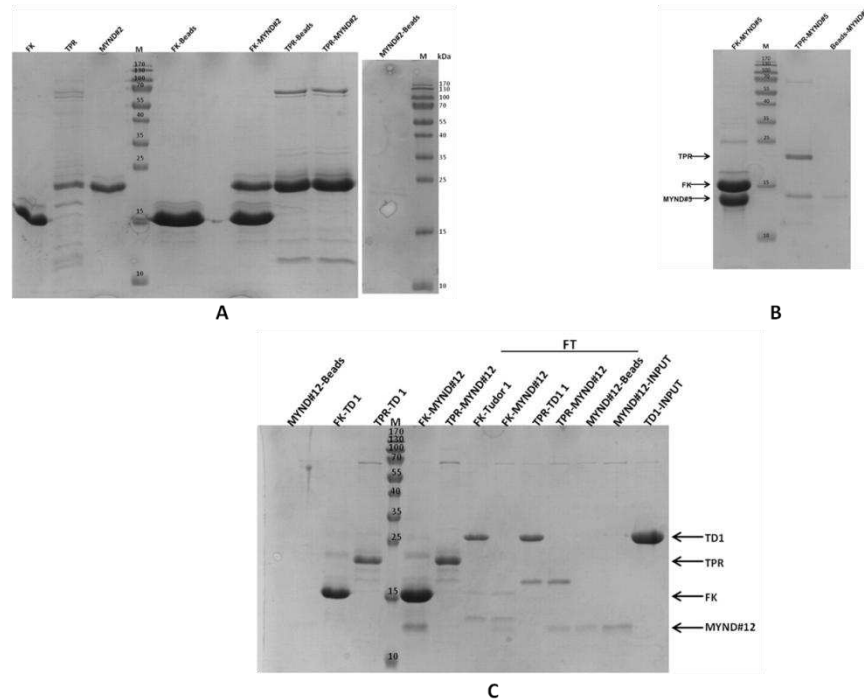


Figure 3.2.8. The FK domain of FKBP6 interacts with the MYND domain of TDRD1. A. Pull-down assay of MYND#2 (56-232 aa) interacting with the FK domain and probably with the TPR domain (the two proteins have a similar molecular weight). **B.** Pull-down assay of MYND#5 (127-232 aa) interacting with the FK domain and to a less extent with the TPR domain. **C.** Pull-down assay of MYND#12 (152-232 aa) specifically interacting with the FK domain, which does not bind TD1 (232-476 aa). Fractions corresponding to FT (flow through) were washed away (unbound proteins) after incubation of the proteins. All binding assays were performed with N-terminal His-tag FKBP6 constructs and non-tagged MYND domain fragments. Beads with cleaved MYND domain were used as control.

Overall, preliminary biochemical experiments show that there is a direct interaction of TDRD1 with FKBP6 through their MYND and FK domains respectively. Constructs comprising only the MYND domain are able to show specific association with the N-terminal region of FKBP6, even though longer constructs also show a weak binding to the TPR domain. Improving the solubility of MYND-containing protein fragments either by optimizing the buffer composition or by cloning new constructs is indispensable in order to further characterize its association with the FKBP6 protein.

4 – DISCUSSION AND FUTURE PERSPECTIVES

Discussion and Future Perspectives

PIWI proteins are a subset of the Argonaute protein superfamily and are expressed predominantly in the germline of a variety of organisms, including *Drosophila* and mammals. PIWI proteins associate with PIWI-interacting RNAs (piRNAs), small RNAs that are also expressed in the germline and silence transposable DNA elements and other genes showing complementarities to the sequences of associated piRNAs. Transposon silencing in mice is initiated by *de novo* methylation of transposon promoters in the embryonic germline. Repeat-derived piRNAs associating with the nuclear PIWI protein Miwi2 are implicated in specifying this event (Aravin and Bourc'his, 2008; Kuramochi-Miyagawa et al., 2008). Primary biogenesis feeds many of the piRNAs entering Mili, and some of these can initiate generation of secondary piRNAs entering Miwi2 via the ping-pong pathway. Briefly, this involves the transfer of a target RNA endonucleolytic cleavage fragment from Mili RNPs to Miwi2, where it would mature as a new secondary piRNA. Two additional factors, TDRD1 and MVH, are also implicated in production of Miwi2-bound piRNAs. In fact, *Mili*, *Mili^{DAH}* (Mili slicer inactive), *TDRD1* and *Mvh* mouse mutants all display a failure to load Miwi2, leading to it being mislocalized to the cytoplasm (Aravin and Bourc'his, 2008; De Fazio et al., 2011; Kuramochi-Miyagawa et al., 2008; Reuter et al., 2009; Vagin et al., 2009). This leads to transposon derepression and hypomethylation of promoter elements in the mutants.

The PIWI clade of Argonaute proteins is distinct from the Ago clade since it has been shown that at their N-termini they contain symmetrical dimethyl arginines (sDMAs) and this modification is mediated by the methyltransferase PRMT5. sDMA is one of the various methyl group modifications found on specific arginines in protein molecules (Bedford and Clarke, 2009) and is known to modify the ability of a protein to perform its biological activities. It was then proven that multiple members of the Tudor family of proteins necessary for gametogenesis, associate with PIWI proteins specifically through sDMAs in various combinations. Proteomic studies from different research groups mapped the methylation status of different PIWI proteins both in *Drosophila* and in mouse (Chen et al., 2009; Kirino et al., 2009a; Nishida et al., 2009; Vagin et al.,

2009). Arginine is found to be symmetrically dimethylated and monomethylated (MMAs) in the context of arginine-alanine (RA) and arginine-glycine (RG) motifs. Mili presents the most diverse methylation pattern, with two sDMAs found in prototypic FR₇₄G and FR₈₃P sequences, while Miwi, Aubergine (*Aub*, *Drosophila*) and AGO3 (*Drosophila*) display fewer sDMAs at their N-termini compared to Mili. Mouse protein Miwi2 and fly Piwi contain RG clusters, however their methylation status is not confirmed yet. The identification of these specific post-translational modifications in various studies differ, mainly due to the sensitivity of the technique used but overall, these proteomic studies are more complementary than contradictory. Finally, these differences may indicate that PIWI methylation can be regulated dynamically through spermatogenesis and may determine the identities of Tudor proteins that are recruited at distinct sub-cellular compartments (Chen et al., 2011; Siomi et al., 2010)

Tudor domain containing protein 1 (TDRD1) encompasses four Tudor domains (TDs) and is implicated in the recognition of symmetrical di-methylarginines of Mili (Reuter et al., 2009; Vagin et al., 2009). The occurrence of tandem Tudor domains and the variety of arginine methylation marks on Mili poses the question of how interactions between the two could help mediate the formation of functional complexes in the piRNA pathway. To provide insight into this interaction network, we identified soluble and well expressed extended Tudor domains (eTuds), either single or in tandem, and we measured the binding of each of the domains (TD1, TD2, TD3, TD4) to three symmetrically dimethylated peptides of Mili (denoted R45me₂, R74me₂, R83me₂ in this study). Using Isothermal Titration Calorimetry (ITC), we show that whereas TD1 binds very poorly to all peptides, TD2, 3 and 4 bind with moderate μ M affinity to the three peptides *in vitro*. Our results are consistent with other studies showing that one Tudor domain is able to recognize many methylated peptides with high affinity (Liu et al., 2010a; Liu et al., 2010b). Our measurements did however reveal a preference for binding of isolated TD2 and TD3 to consecutive Mili peptides, R45me₂ and R74me₂ respectively. Therefore we examined whether TDRD1 can integrate inputs from multiple Tudor domains by measurements of a longer peptide combining both these methylation marks to the tandem TD2-TD3

domains (TD2-3). No significantly enhanced affinity was observed suggesting that binding to each domain is largely independent, probably due to peptide and inter-domain flexibility. Furthermore, titrations of the doubly methylated longer peptide with single TD2 and TD3 also showed same levels of μM affinity, fitted to one set of binding sites. Similar studies of the extended Tudor domain of SDN1 protein with Miwi peptide harboring three methylation sites reveals same affinity as for one methylation site (Liu et al., 2010b). Conversely, tandem Tudor domains 7-11 of *Drosophila* protein TUDOR and single Tudor domain 11 (eTud11) display similar binding to an Aub peptide with one sDMA (Liu et al., 2010a). These observations suggest that Tudor and PIWI's sDMA-peptides display no-cooperativity and that there is no 1:1 exclusive interaction. This is in contrast to the recognition of the Sm-methylated arginines by the prototypic Tudor domain of SMN protein, where external cooperativity effects promote high affinity when multiple methylated arginines are present (Tripsianes et al., 2011). Overall, these results show that Tudor domains in different cellular pathways recognize their methylated binding partners through distinct mechanisms.

Since different eTuds of TDRD1 bind various methylated Mili peptides, we wanted to investigate in the atomic level how these interactions occur. For that reason, we tried to co-crystallize single and tandem eTuds with the three peptides. Our extensive efforts resulted in solving the structure of the Tudor domain 3 (TD3) in complex with the R45me2 peptide. The crystal structure reveals a negatively charged binding groove where TD3 is comprised of a barrel-like Tudor domain, commonly found in many Tudor domain containing proteins, fused to a SN-like domain. The overall fold extremely resembles the previously identified Tudor domain structures of fly eTud11 and the human SND1 (TDRD11) that recognize methylated arginine-peptides (Liu et al., 2010a; Liu et al., 2010b). However, the TD3-R45me2 complex unexpectedly reveals a completely different orientation of the peptide and the sDMA accommodated inside the aromatic cage. Interestingly, even though the aromatic residues forming the binding cage are conserved in all eTudor domain structures, the orientation of the symmetrically dimethylated arginine bound differs in our structure compared to others. More precisely, one of the methylated guanidine

groups of the sDMA is rotated 120 degrees about the CZ atom. Additionally, while sDMAs bound to extended Tudor domains are in the *anti-syn* conformation (one methyl group is on the opposite side of the guanidine-N ω bond while the other is on the same side), this is not the case for the cage of SMN Tudor domains where the *anti-anti* conformation is found (Tripsianes et al., 2011). It has been proposed that this difference is due to the systematic occurrence in the aromatic cage of single Tudor domains of a tryptophan in place of the phenylalanine most commonly found in eTud domains (F767 in TD3) and of the different rotameric state of an asparagine (N796 in TD3) (Tripsianes et al., 2011). Finally, this aromatic cage is also present in methyllysine Tudor domains, but the narrow arrangement of aromatic residues in methylarginine Tudors can only accommodate the planar methyl-guanidinium group (Liu et al., 2010b). These observations depict the first remarkable plasticity of the Tudor domain cage found not only in Tudor domains implicated in the piRNA mechanism but also in other methylarginine-dependent interaction modules.

The second evidence for Tudor domain plasticity resides in the analysis of the interaction network between the residues flanking the sDMA and the eTud domain structure. Others and we show that extensive interactions involving the SN-like domain stabilize the orientation of the peptide. Comparing other eTud-methylated peptide complexes, we observe that different protein residues contribute to binding to the flanking peptide amino acids, thus confirming the novel orientation observed; in TD3, the R45me2 peptide enters from the 'top' of the protein, while in other structures the peptides bind from the 'bottom'. Furthermore, our structure provides for the first time evidence for interactions that are formed with eTud residues outside the RG/RA motif that was not observed before due to crystal disorder. The large hydrophobic residues just downstream from the methylated arginine, as present in the R45me2 peptide, highlights the potential of similar orientations upon binding of similar peptides found in the N-terminus of Mili (e.g. R74me2). However, the low conservation of the N-terminal regions of PIWI proteins and the residues at the SN-like domain does not allow us to predict how other sDMA-containing peptides bind to eTud folds. This probably also reflects the different affinities present in ITC

measurements, suggesting that regional electrostatic and hydrophobic interactions may determine how well a specific peptide will interact with a specific eTudor domain. One possible way to explore peptide orientation would be by NMR titrations, in which the shift of specific amino acids upon binding can be detected, thus predicting the orientation of the bound-peptide. Finally, using ITC and mutagenesis on the SN-like domain residues responsible for making peptide contacts, informative affinities could be obtained.

The multiple contacts of methylated peptides with the extended Tudor domains prompted us to explore if other Tudor domains of the TDRD proteins in mouse have also the extended domain architecture. In a multiple sequence alignment we saw that most of the TDRD proteins display conservation of the amino acids located at the SN-like domain, the inter-domain surface and the Tudor core domain. This holds true also for eTud domains in *Drosophila* as seen from similar studies (Liu et al., 2010a). More specifically, the absolutely conserved R775 and D793 (residues in TD3) play key roles in stabilizing the aromatic cage and keeping the integrity of the eTud fold. Indeed, R774 mutation in Tudor domain 10 of fly TUDOR impairs the function of the domain (Arkov et al., 2006). On the other hand, both residues are missing in prototypical Tudor domains of SMN and SPF30, as well as in methyl-lysine-binding Tudor domains (Huang et al., 2006). Moreover, this extended nature is also indispensable for methylarginine binding; the first crystal structure of a Tudor domain involved in the piRNA pathway in TDRD2 protein (Chen et al., 2009) included only the prototypical Tudor domain (~80 residues) and later proved to be insufficient for binding sDMAs in the context of a peptide or as single residue (Liu et al., 2010b; Tripsianes et al., 2011). When a longer construct including the predicted SN-like domain (~200 residues) was titrated with Miwi's methylated peptide, the K_D measured was 4 μ M. Overall, these observations suggest that the eTud domains in the piRNA pathway are evolutionarily conserved in recognizing sDMAs in a context of a peptide while the prototypical Tudor domains SMN and SPF30 involved in pre-mRNA splicing bind only sDMAs of Sm proteins independently of flanking residues.

Tudor domain containing proteins in fly and murine piRNA pathways have proved to have important roles for the normal progression of oogenesis and spermatogenesis respectively. The varying number of extended Tudor domains they possess and the smaller number of methylated arginines raises the question of how many actually are active in binding sDMA-containing peptides. Our ITC data clearly suggest that TD2, 3 and 4 recognized methylated peptides with different affinities but TD1 showed always weak binding to the peptides tested. For that reason, we wanted to explore the methylarginine capacity of each eTud of TDRD1 by measuring the affinity to single sDMA residue. Our results further proved that TD1 is probably not binding single or in a peptide context sDMAs. Furthermore, TD2 and TD4 showed better binding to single dimethylarginines than TD3 (172 and 275 compared to 353 μ M respectively). This is probably due to the fact that TD2 and TD4 possess a tyrosine instead of a phenylalanine in TD3 in the aromatic cage (F791 in TD3) that can form hydrogen bond with the conserved glutamic acid placed adjacent to the pocket (E798 in TD3) as seen in the SMN prototypic Tudor domain (Tripsianes et al., 2011). Next, we titrated eTuds with single arginine unmodified (Runmethylated), asymmetrically dimethylated (aDMA) and monomethylated (MMA). As expected, we observed no affinity to Runmethylated, and very small binding with aDMAs and MMAs. This is in accordance with similar studies, further validating that Tudor domains are preferentially recognizing sDMAs over other arginine modifications (Liu et al., 2010a; Liu et al., 2010b; Tripsianes et al., 2011). This is also true in the context of a peptide, where the same results were obtained (Liu et al., 2010b). In addition, it is likely that monomethylation is not involved in the binding of any Tudor domain, as it is an intermediate step before the double methylation by type II methyltransferases, such as PRMT5 (Bedford and Clarke, 2009). However, even though we observed no binding of TD3 with R74 peptide, Liu and colleagues (2010) observed 94 μ M affinity for unmethylated peptide. One possible way to retrieve information about the contribution of the residues outside the aromatic pocket would be to mutate simultaneously all the aromatic residues of a Tudor domain and study binding kinetics with methylated peptides. Moreover, eTud of SND1 bound with high affinity to a histone peptide harboring sDMA (Liu et al., 2010b). It would be interesting to explore how specific are the associations of

eTuds with methylated PIWI peptides, by measuring their affinity for instance to Sm methylated tails. In addition, high-throughput approaches including Tudor domains from various proteins and peptide array screening in different methylation state can contribute to the understanding of Tudor-peptide association. Overall, we observe that much of the binding affinity derives from recognition of the methylated arginine itself and there is a possibility that unmethylated peptides are also recognized by Tudor domains.

Multiple sequence alignment between eTuds of TDRD1 revealed that the aromatic cage of TD1 possess N325 found in the place of the conserved tyrosine in the rest of the Tudor domains (Y774 in TD3). Mutagenesis analysis and ITC titrations revealed that indeed this single amino acid substitution (N325Y) was responsible for the low affinity of TD1 with either R45me2 peptide or single sDMAs, rendering it 'inactive'. Placing an asparagine in the place of tyrosine also impaired the methylarginine activity of TD3, suggesting that this probably also happens in TD4 or other TDRD proteins. Human and bovine TDRD1 maintain the asparagine at their TD1, however zebrafish displays an active TD1. In fact, fish single eTud domains of TDRD1 recognize sDMAs on Zili's N-terminus (Huang et al., 2011). One can speculate that the presence of only two PIWI proteins (Zili and Ziwi) requires the methyl-binding activity of all Tudor domains in order to operate efficiently. Hence, taking advantage of similar mutagenic analysis from other studies ((Liu et al., 2010b; Tripsianes et al., 2011)), we predict that the signature motif of an 'active' eTud domain is FXnW(F)YRXnF(Y)XDY(F)GN, consistent with the observation that all reported crystal structures have these residues. Based on this motif, we predict that more than half of the eTuds in TDRD proteins are unable to recognize methylated arginines. Our prediction is validated by studies focused mainly in fly eTud proteins. Firstly, *Drosophila* Tudor can bind sufficiently Aub and promote germ cell formation only with its five C-terminal Tudor domains (eTud7-11) (Liu et al., 2010a), which are all having the 'active' motif and bind sDMAs. On the other hand, the N-terminal six domains show significant deviations from it but they can still bind Aubergine, even if the need for methylation is unknown so far (Creed et al., 2010). Secondly, the fly Tudor domain protein Tejas which interacts with Aub in a sDMA-

independent manner displays an 'inactive' motif (Patil and Kai, 2010). Moreover, the predicted 'inactive' Spindle-E has one Tudor domain and also did not show sDMA-dependence neither on Aub nor on AGO3 (Nishida et al., 2009). On the other hand, other Tudor domains with a predicted 'active' motif display methylarginine dependence upon binding to their partners. Thus, we conclude that some Tudor domains are required for proper efficiency of the piRNA pathway by recognizing methylated arginine tails of PIWI proteins, while others are 'inactive' and have an unknown role or an unidentified so far binding partner. It is possible that they can recognize PIWI proteins in a methylarginine-independent manner like some members of the PHD and chromodomain families (Taverna et al., 2007), although a more detailed experimentation is required to draw safe conclusions.

Our prediction that TD1 is an 'inactive' Tudor domain is also confirmed by pull-down assays with endogenous PIWI proteins and MVH from adult mouse testis. While TD2, 3, 4 show strong association, TD1 bound very weakly only to Mili, suggesting that it may be involved in recognizing other N-terminal, non-methylated fragments that have not yet been tested by ITC. However, TDRD1 is still a binding partner of Mili, Miwi and MVH since the tandem four Tudor domains (TD1-4) displayed binding. At this point it should be pointed that in pull-down experiments, direct interactions between proteins cannot be confirmed and further experiments exploring direct protein-protein interactors are necessary to understand target specificity. In the same notion, similar experiments showed that MVH interacts with TDRD1 and TDRD6, as well as Mili and Miwi (Kirino et al., 2010). These observations suggest that TDRD proteins may be part of various heteromeric complexes recognizing sDMAs and aDMAs (present in MVH) during different stages of spermatogenesis. In this interaction network, direct binding of Mili to TDRD1 and Miwi to TDRD6 have already been established (Kirino et al., 2009b; Reuter et al., 2009) Furthermore, the 'active' TD1 N325Y mutant did not bind Mili sufficiently, suggesting that even though an 'active' aromatic cage is formed, it is possible to interact with another piRNA-associated factor in a sDMA-independent manner. In fact, treatment with MTA (methylation inhibitor) did not affect the interactions of MVH with TDRDs and

PIWI proteins (Kirino et al., 2010). Also, Mili's sDMA interaction with TDRD1 is abolished upon treatment with MTA but the N-terminal region was not detected by Y12 (antibody recognizing sDMAs) unlike full-length Mili (Reuter et al., 2009), implying that methylation may be dispensable for TDRD1-Mili interaction. Moreover, even though we did not see tight binding of eTuds with aDMAs, this interaction is still possible as shown by the recent crystal structure of SMN with a single aDMA (Tripsianes et al., 2011). Interestingly, the aDMA side-chain is bound in SMN in a rotated configuration which is much closer to that observed for sDMA bound to TD3 than to eTud11 and SDN1. However, the significance of this correspondance with aDMA binding to core Tudor domains is not clear. This further suggests that Tudor domains could potentially bind aDMA in the context of a peptide. MVH is the only piRNA factor harboring aDMAs and one could measure affinities of Tudor domains with aDMA-peptides to draw conclusions about direct association of the two proteins.

As noted before, multiple Tudor domains recognizing methylation marks on Mili, Miwi and MVH are likely to promote the formation of a complex interaction network. It has been proposed that PIWI or TDRD proteins can have a scaffolding role, by simultaneously engaging multiple sites for ligand recognition. In one scenario, Mili or Miwi could use their methylated arginines as docking sites for multiple Tudor domains of one or more TDRD proteins. Conversely, the aromatic cage and the extended eTud fold could serve as a molecular platform that accommodates multiple methylated peptides from one PIWI protein or cross-link distinct PIWI proteins and/or MVH. The elongated conformation of the SAXS model of TDRD1, with each Tudor domain poised to receive its interacting partner, along with its overall and binding pocket plasticities, provides the strongest evidence to date of a TDRD protein acting as a molecular scaffold. This means that TDRD1 can indeed act as a structural module capable of engaging one partner through a set of composite contacts or bind multiple partners simultaneously. These different options may be appropriate at different times to create unique complexes or interaction networks dynamically, depending on partner availability at different developmental stages and granule composition. Indeed, the universal role of Tudor domains as binders of methylated arginines

or lysines and the fact that most of TDRD proteins are composed mainly by extended Tudor domains makes them ideal candidates for a docking station, since PIWI proteins have been proven to be responsible for RNA binding and act as effective slicer entities of the piRNA pathway. Furthermore, this scenario is strengthened by us and others showing that there is no structural perturbation upon methylarginine binding and that Tudor domains bind in a dynamic equilibrium to multiple methylated sites. It would be interesting to see if structural rearrangements are taking place in the level of multiple binding sites, by calculating for instance the SAXS model of TDRD1 in the presence of a long methylated peptide derived from Mili or other proteins. Overall, we speculate that Tudor domains are able to read-out a methylarginine code and stabilize their interaction partners in analogy to the histone code where post-translational modifications are responsible for the recruitment of binding protein modules.

TDRD1 is an indispensable factor for the normal development of spermatogenesis, since its depletion leads to spermatogenic arrest at the pachytene stage, a phenotype similar to *Mili* mutants. (Kuramochi-Miyagawa et al., 2004). At the piRNA pathway TDRD1 was shown to participate in transposon silencing by piRNAs (Reuter et al., 2009) and have an impact both in PIWI protein localization and piRNA biogenesis. More specifically, the piRNA profile is altered in *Tdrd1* mutant mouse, with an increased amount of exon-derived piRNAs in pachytene stages. Interestingly, ping-pong signatures are lost in embryonic stages, leading to a depletion of secondary piRNAs targeting Line1 sequences and subsequent derepression of Line1 elements (Vagin et al., 2009). The amount of nuclear Miwi2 is reduced in the mutant, while in the cytoplasm it fails to localize in the specific granules where it is found in wild-type conditions. TDRD1 does not show involvement in the primary biogenesis of piRNAs but has a functional role in the amplification loop in a RNA-independent manner where it seems to specify the piRNA content of Mili and the efficient operation of the ping-pong cycle between Mili and Miwi2. It is not clear yet if it interacts with Miwi2 directly and the sub-cellular localization indicates that Mili-TDRD1 complex operates in the pi-bodies whereas Miwi2 is found in piP-bodies. These two sub-

organelles are highly dynamic and found in a close proximity, where an exchange of piRNA-associated factors may take place between them.

Taking into account various genetic and biochemical studies, it is speculated that during the pre-pachytene stage of spermatogenesis, in the first place, the N-terminus of Mili is methylated by PRMT5 and subsequently TDRD1 is recruited to interact with the methylated arginines. At this step, it is possible that PRMT5 can be a part of Mili-TDRD1 assembly and act synergistically for the formation of a functional complex. TDRD1 is able to bind methylarginine peptides independently and thus stabilize Mili in order to be functionally active to take part into the amplification loop with Miwi2. Another possibility is that TDRD1 can be the bridging molecule between the two PIWI proteins, in order to efficiently exchange sense and antisense piRNAs and this has been proposed for TUDOR protein, Aubergine and AGO3 in *Drosophila* (Nishida et al., 2009). However, Miwi2 is interacting with TDRD9 protein in piP-bodies, which can be formed even in the absence of TDRD1. Afterwards, both proteins are translocated to the nucleus (Aravin and Bourc'his, 2008; Shoji et al., 2009) for the repression of transposable elements. TDRD9 is predicted to have 'active' Tudor domain but information about methylarginine recognition of Miwi2 is missing. Overall, one can say that TDRD9 complexes with Miwi2 and together cooperate with TDRD1-Mili in distinct sub-cellular compartments non-redundantly.

Distinct PIWI-TDRD interactions have also been established during the pachytene stage of spermatogenesis but their functional implication in the piRNA pathway is still unclear. TDRD1 was shown to interact with both PIWI proteins, Mili and Miwi, along with other TDRD proteins. TDRD6 binds Miwi in sDMA-dependent manner *in vivo* at its first RG/RA cluster rather than the second and third clusters (Kirino et al., 2010) and its absence causes disruption of Chromatoid body formation in round spermatids (Vasileva et al., 2009). Interestingly, TDRD2 was also proven to bind symmetrically dimethylated arginines of Miwi in the same region (Chen et al., 2009). Similarly, this is also present in flies, where TUDOR, the fly orthologue of TDRD6 binds Aub via sDMAs and AGO3 only at its second arginine-rich cluster (Nishida et al., 2009). It would

be interesting to elucidate if this preference for specific regions at the N-terminus of PIWI proteins is also present in Mili-TDRD1 interactions. It should be noted here that we also tried unsuccessfully to explore binding specificities of single eTuds with various Mili N-terminal deletion fragments in an *in vitro* translation system. In fact, it was shown that in a quadruple mutant of Mili's R9, R39, R45 and R74 into lysines, there was no detectable binding to TDRD1 in 293 cells (Vagin et al., 2009). This is in accordance with our ITC data showing that the R83me2 peptide, which follows R74me2, displays low affinity to all eTuds, and especially TD4, suggesting for a minimal Mili's methylated region that TDRD1 can bind. On the other hand, the 'active' TD4 could recognize other untested methylated peptides of Mili with higher affinity. One potential experiment would still be to mutate selectively all Mili's arginines into lysines and then co-express GFP-tagged TDRD1 and Flag-tagged Mili mutants in HEK293T cells and detect any binding by Western blot. Taken altogether, we observe that there is a highly dynamic network in Tudor-PIWI interactions that is formed according to the developmental stages and the requirements of the piRNA pathway.

Recently, a new piRNA-associated factor was discovered in Pillai's group and shown to be indispensable for spermatogenesis. FK506-binding protein 6 (FKBP6) acts downstream of Mili and TDRD1 and affects Miwi2's localization and its association with piRNAs (Xiol et al., submitted). The isomerase activity of the FK domain in FKBP6 protein is lost but is still possible to accommodate proline residues at its binding pocket that are abundantly found in Mili's N-terminus, a scenario that is strengthened by the specific association of Mili and FKBP6 (Xiol et al., unpublished). Additionally, we have shown that there is association of the inactive isomerase domain with the MYND domain potentially through a common recognition motif found at FK domains' surface. Moreover, the TPR domain of FKBP6 was shown to specifically associate with Hsp90 that directly involves the chaperone system in the piRNA pathway. Thus, as it has been suggested, FKBP6 can be the unknown piRNA-loading factor that in a first step is recruited by the MYND domain of TDRD1 and afterwards associate with Mili to ensure the correct entry of transcripts into a functional RNP assembly (Reuter et al., 2009). On the other hand, it is also possible that the MYND domain

interacts with Mili's N-terminus as it has been proposed that TDRD1's MYND domain and its first two Tudor domains are sufficient for its association with Mili (Wang et al., 2009). More experimentation is required in order to confirm MYND-FK domain interaction and future structural studies can reveal how this protein-protein interaction module is formed.

To conclude, taken together our data and published evidence, we propose a purely speculative model in which the dynamic formation of different macromolecular complexes among PIWI, TDRD proteins and yet unidentified protein factors drive the efficient operation of the piRNA pathway towards the repression of transposable elements and the normal progression of spermatogenesis. These ribonucleoprotein assemblies resemble the RNA-induced silencing complexes (RISC) in miRNA and siRNA pathways. In mouse, methyltransferases are the 'writers' of the arginine methylation marks on PIWI proteins and Tudor domains are the 'readers', and this system contributes to the formation of sub-cellular localizations. This mechanism is in analogy to the function of PRMT5 in RNA processing, where the methylation of Sm tails promotes their association with Tudor domains for proper assembly of UsnRNP. In pi bodies, TDRD1 acts as a scaffold protein and through its plasticity and flexibility can 'scan' the methylation marks on Mili, thus stabilizing it and making it functional in the secondary processing pathway ([fig.4.1](#)). TDRD9 can possibly have a similar role with Miwi2 in the proximal piP-bodies and altogether act synergistically in order the latter to efficiently promote methylation of transposon promoters leading to their repression. In the primary processing pathway and the round spermatid stage, where Mili and Miwi are involved, many TDRD proteins participate with complementary and non-redundant functions during development. It is highly likely that molecular platforms composed of Tudor domains can also be formed at different sub-cellular compartments and at different stages in order to facilitate PIWIs complete their functional role. Multi-Tudor domain-containing proteins may use distinct Tudor domains in the same protein to simultaneously recognize different arginine methylated PIWI members. This dynamic network becomes more complex with the versatility presented by a subset of TDRD proteins, which contain various domain

combinations, mainly known to be involved in RNA binding or metabolism. A new level of complexity is added by the discovery of possible multiple ‘inactive’ Tudor domains and there are still many questions to be answered. Do they bind PIWI members in a sDMA-independent way or have an additional binding factor? Are they important for the efficiency of the piRNA pathway and sub-cellular localization? Further exciting experiments can shed light on these new Tudor domain’s roles. Taken altogether, we conclude that the coordination of Tudor-PIWI interaction provides a highly efficient mechanism to quickly form a functional silencing complex to protect DNA of future generations from transposons.

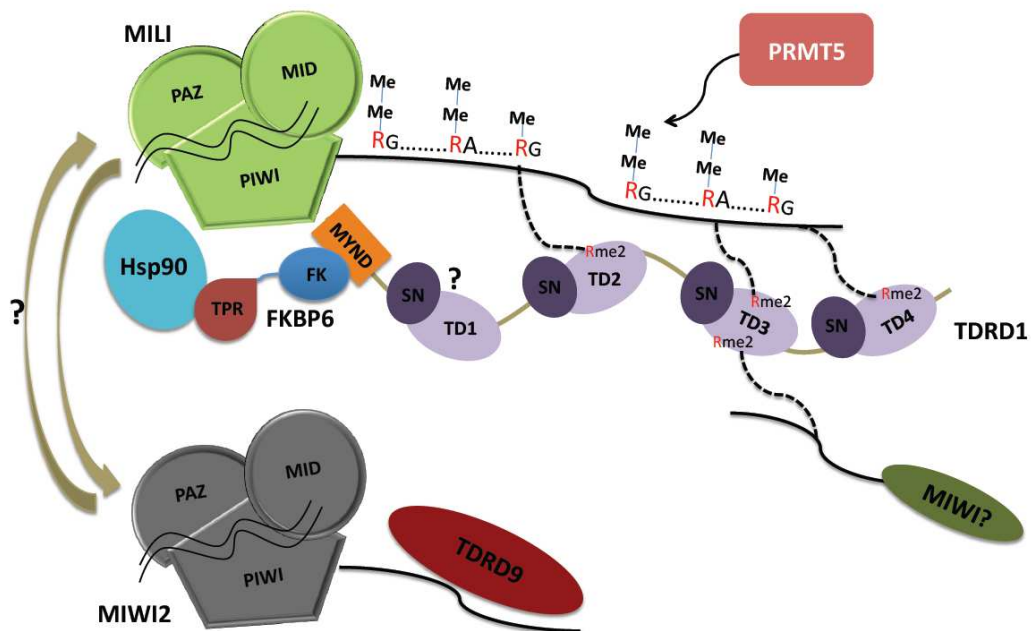


Figure 4.1. A model for the role of TDRD1 in the piRNA pathway. Mili contains clusters of RG/RA repeats that in a first step are symmetrically dimethylated by PRMT5. In the next step, TDRD1 is recruited and binds Mili’s methylated arginines at its N-terminal region independently (dashed lines), since there is no indication for 1:1 interaction. This association is mediated by both the Tudor core domain and the SN-like domain, with the latter recognizing proximal to the dimethylarginine residues. TD1 is unable to bind methylated arginines and thus has an unknown role. The MYND domain could possibly recruit FKBP6 via its FK domain resulting in the formation of a macromolecular complex. TDRD1 acts as a molecular scaffold and helps Mili to be functionally competent to allow the correct entry of the piRNA pool and start the amplification loop with Miwi2 (the question mark denotes the possibility for a linear transmission of piRNAs since Miwi2 is inactive as a slicer). Miwi2 interacts with TDRD9 and together are translocated to the nucleus where they promote the methylation of transposon sequences, leading to their repression. The ability of an extended Tudor domain to accommodate peptides from different orientations raises the possibility that additional factors harboring methylated arginines (Miwi, MVH) can also bind TDRD1.



RNA

A PUBLICATION OF THE RNA SOCIETY

The multiple Tudor domain-containing protein TDRD1 is a molecular scaffold for mouse Piwi proteins and piRNA biogenesis factors

Nikolas Mathioudakis, Andres Palencia, Jan Kadlec, et al.

RNA published online September 20, 2012

Access the most recent version at doi:[10.1261/rna.034181.112](https://doi.org/10.1261/rna.034181.112)

**Supplemental
Material**

<http://rnajournal.cshlp.org/content/suppl/2012/09/12/rna.034181.112.DC1.html>

P<P

Published online September 20, 2012 in advance of the print journal.

**Email alerting
service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)



Publish your data
using Ambion® products

ambion
by life technologies™

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to RNA go to:

<http://rnajournal.cshlp.org/subscriptions>

The multiple Tudor domain-containing protein TDRD1 is a molecular scaffold for mouse Piwi proteins and piRNA biogenesis factors

NIKOLAS MATHIOUDAKIS,¹ ANDRES PALENCIA,¹ JAN KADLEC,¹ ADAM ROUND,¹ KONSTANTINOS TRIPSANES,² MICHAEL SATTLER,² RAMESH S. PILLAI,¹ and STEPHEN CUSACK^{1,3}

¹European Molecular Biology Laboratory, Grenoble Outstation and Unit of Virus Host-Cell Interactions, UJF-EMBL-CNRS, UMI 3265, BP181, 38042 Grenoble Cedex 9, France

²Institute of Structural Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany

ABSTRACT

Piwi-interacting RNAs (piRNAs) are small noncoding RNAs expressed in the germline of animals. They associate with Argonaute proteins of the Piwi subfamily, forming ribonucleoprotein complexes that are involved in maintaining genome integrity. The N-terminal region of some Piwi proteins contains symmetrically dimethylated arginines. This modification is thought to enable recruitment of Tudor domain-containing proteins (TDRDs), which might serve as platforms mediating interactions between various proteins in the piRNA pathway. We measured the binding affinity of the four individual extended Tudor domains (TDs) of murine TDRD1 protein for three different methylarginine-containing peptides from murine Piwi protein MILI. The results show a preference of TD2 and TD3 for consecutive MILI peptides, whereas TD4 and TD1 have, respectively, lower and very weak affinity for any peptide. The affinity of TD1 for methylarginine peptides can be restored by a single-point mutation back to the consensus aromatic cage sequence. These observations were confirmed by pull-down experiments with endogenous Piwi and Piwi-associated proteins. The crystal structure of TD3 bound to a methylated MILI peptide shows an unexpected orientation of the bound peptide, with additional contacts of nonmethylated residues being made outside of the aromatic cage, consistent with solution NMR titration experiments. Finally, the molecular envelope of the four tandem Tudor domains of TDRD1, derived from small angle scattering data, reveals a flexible, elongated shape for the protein. Overall, the results show that TDRD1 can accommodate different peptides from different proteins, and can therefore act as a scaffold protein for complex assembly in the piRNA pathway.

Keywords: piRNA; Piwi protein; extended Tudor domain; methylated arginine; X-ray crystallography

INTRODUCTION

Small RNAs associate with Argonaute proteins to form the RNA-induced silencing complexes (RISC) that mediate transcriptional or post-transcriptional silencing of their nucleic acid targets (Filipowicz et al. 2008). Based on their sequence conservation Argonautes can be partitioned into the Ago and Piwi clades (Carmell et al. 2002). Ago members associate with ~21 nucleotide microRNAs (miRNAs) or small interfering RNAs (siRNAs) and are responsible for control of gene expression in all cell types. The Piwis, on the other hand, are restricted to the animal gonads, where they associate with

~24–31 nucleotide Piwi-interacting RNAs (piRNAs). Together, they are implicated in controlling transposon mobility in worms, flies, fishes, amphibians, and mammals (Ghildiyal and Zamore 2009; Malone et al. 2009; Siomi et al. 2011). Loss of Piwi proteins or other factors acting in the pathway leads to activation of transposons, resulting in DNA damage, and finally causing sterility of the individual.

Through a combination of genetic and computational studies, it is now clear that long single-stranded RNAs transcribed from genomic loci called piRNA clusters are major sources of piRNAs (Aravin et al. 2006; Girard et al. 2006; Brennecke et al. 2007). Currently, it is thought that piRNA biogenesis can occur through distinct primary or secondary processing pathways (Senti and Brennecke 2010). Primary processing describes the generation of piRNAs from single-stranded RNA precursors by the action of an unknown nuclease(s), since it has been shown that unlike miRNAs and

³Corresponding author

E-mail cusack@embl.fr

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.034181.112>.

siRNAs, which have double-stranded precursors, piRNA production does not depend on Dicer or Drosha (Vagin et al. 2006; Houwing et al. 2007). Secondary processing is an elegant mechanism to harness transposon cleavage fragments for new piRNA production. Here, primary piRNA-guided Piwi endonuclease (slicer) action defines the 5' end of sense secondary piRNAs, which can, in turn, generate the initiating antisense piRNA by cleaving cluster transcripts via the so-called ping-pong amplification cycle (Brennecke et al. 2007; Gunawardane et al. 2007). These piRNAs of opposing polarity usually enter distinct Piwi proteins, which in the case of flies is Aubergine (Aub) and Ago3. The ping-pong cycle in flies is essential to maintaining an overall antisense silencing potential to prevalent transposon populations (Li et al. 2009). In mice, Piwi proteins MILI and MIWI incorporate primary piRNAs, while MIWI2 depends on a functional engagement with MILI to generate its RNA partners via secondary biogenesis. Only piRNA-loaded MIWI2 is licensed for nuclear entry and transposon silencing by DNA methylation (Aravin et al. 2008; Kuramochi-Miyagawa et al. 2008).

Argonaute proteins can be structurally divided into the N-terminal region (N domain), PAZ, MID, and Piwi domains (Sashital and Doudna 2010). In all Argonautes, the small RNA 3' and 5' ends are anchored in the PAZ and MID domains, respectively. Also, in several Argonautes, the Piwi domain, which adopts an RNase H fold, harbors the slicer endonuclease activity (Song et al. 2003; Wang et al. 2009b). However, the Ago and Piwi clade members differ in their N domains. Piwi proteins uniquely carry several arginine-glycine (RG) and arginine-alanine (RA) dipeptides in their N domains, and several of these arginines have been shown to be post-translationally modified by methylation (Kirino et al. 2009; Reuter et al. 2009; Vagin et al. 2009). Although these are predominantly symmetrically dimethylated marks, some monomethylated arginines have also been found (Vagin et al. 2009). The symmetrically dimethylarginine (sDMA) marks, catalyzed by the methyltransferase PRMT5, act as affinity tags for members of the Tudor domain protein family (Kirino et al. 2009; Nishida et al. 2009; Reuter et al. 2009; Vagin et al. 2009; Wang et al. 2009a; Patil and Kai 2010; Handler et al. 2011).

The Tudor protein family is defined by the presence of the Tudor domain, which belongs to the “royal family” of modules that can specifically recognize methylated ligands to enable protein-protein interactions (Maurer-Stroh et al. 2003). They are implicated in RNA splicing, chromatin remodeling, and germline development. The prototypic Tudor domain is a ~60 amino acid β -barrel that builds an aromatic cage capable of accommodating either methylated lysines or arginines (Kim et al. 2006). Structural studies have revealed that there are two distinct groups of Tudor domain-containing proteins, both of which can recognize ligands with methylated arginines. The group defined by the Spinal Motor Neuron (SMN) proteins contains the prototypic

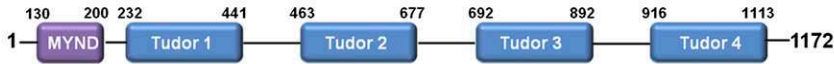
Tudor domain, sometimes in multiple copies, and recognizes only the methylated side chain, but not its peptide context (Selenko et al. 2001; Tripsianes et al. 2011). Only recently have the NMR structures of the minimal, prototypical Tudor domains of SMN and SPF30 (survival of motor neuron-related splicing factor 30) been determined in complex with single symmetrically and asymmetrically dimethylated arginine residues. The second group comprises so-called Tudor domain-containing (TDRD) proteins and is represented by the *Drosophila* TUDOR (Liu et al. 2010a) or the human Staphylococcal nuclease domain-containing 1 (SND1) (Liu et al. 2010b). These proteins contain one or more extended Tudor domains (eTud), typically of ~180 residues, in which the prototypic Tudor module is fused to a staphylococcal nuclease (SN) domain. The conserved aromatic cage of the Tudor domain is responsible for specifically binding the symmetrically dimethylated arginine side chain, but flanking residues make additional contacts with the SN domain, modulating the affinity of different methylated arginine peptides to the eTud domain (Liu et al. 2010a,b; Chen et al. 2011). The various eTud-peptide structures determined show that there is considerable plasticity in exactly how the flanking regions of the methylated arginine peptide are bound. Furthermore, it has recently been shown that there are subtle, yet systematic, conformational differences between the mode of binding of methylated arginine by the conserved aromatic cage in eTud compared with prototypical Tudor domains (Tripsianes et al. 2011).

Several TDRD proteins were detected in fly, fish, and mouse Piwi complexes, with the Piwi proteins often displaying distinct specificities for different TDRDs (Nishida et al. 2009; Reuter et al. 2009; Vagin et al. 2009; Wang et al. 2009a; Patil and Kai 2010; Handler et al. 2011). Consistent with the involvement of Tudor domains, many such interactions were shown to depend on the presence of sDMAs on Piwi proteins. Both Piwi proteins and TDRDs are colocalized in several cytoplasmic perinuclear granules called the nuage, which are signature features of germ cells (Arkov and Ramos 2010). The TDRD proteins carry varying numbers of Tudor domains and are frequently associated with other domains that impart additional functions such as helicase activity or RNA binding (Handler et al. 2011). All of these suggest the potential to generate an intricate and dynamic network of interactions and assemblies that carry unique functions. The importance of these proteins for germline development is highlighted by the sterility observed in TDRD mutants, which is often accompanied by transposon derepression and disrupted nuage (Siomi et al. 2011).

Tudor domain-containing 1 (TDRD1) is a multidomain protein with an N-terminal MYND (myeloid translocation protein 8, Nery, and DEAF-1) zinc finger domain, followed by four tandem extended Tudor domains, denoted TD1–4 (Fig. 1A; Chuma et al. 2003). Its expression tightly overlaps with that of Piwi proteins during mouse spermatogenesis,

A

TDRD1



MILI N-terminus

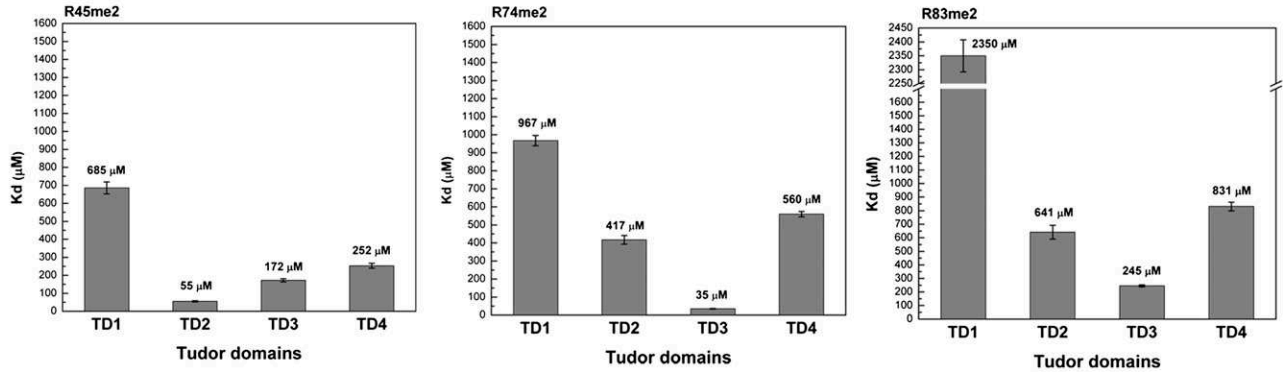


GRAGPAGRGLVFR

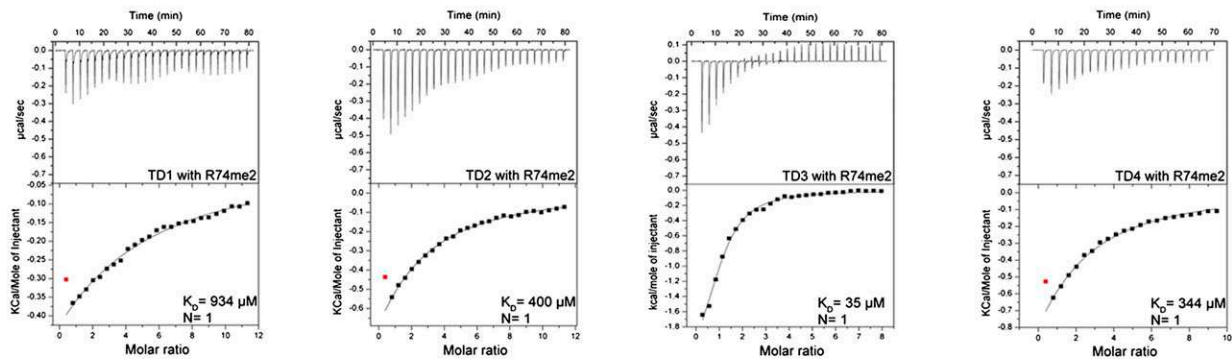
LVSMFRGMGLDT

MGLDTAFRPPSKR

B



C



D

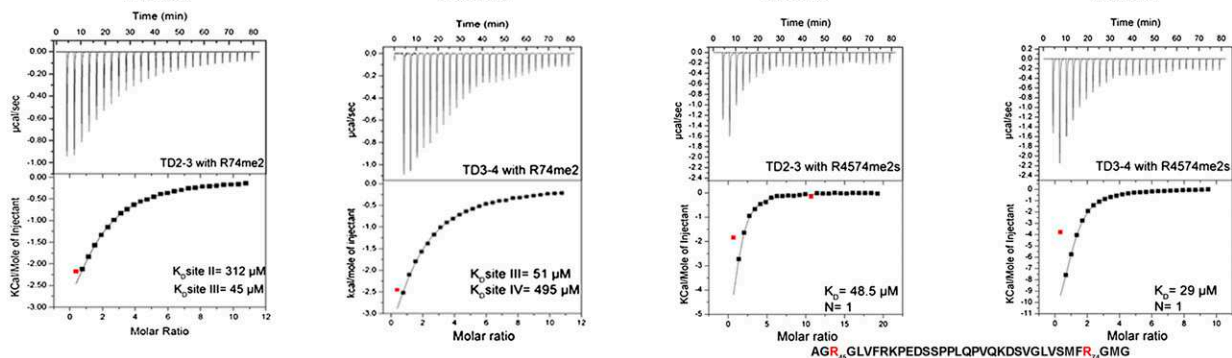


FIGURE 1. Binding affinity of sDMA containing peptides of MILI to individual TDRD1 eTud domains. (A) Schematic representation of mouse TDRD1 and the N terminus of MILI. TDRD1 has an N-terminal MYND domain and four tandem extended Tudor domains. The N terminus of MILI contains multiple arginine residues being mono- and symmetrically dimethylated (R highlighted in red). The three methylated peptides used in this study are indicated. (B) K_D s derived from ITC measurements for three sDMA-containing peptides of MILI binding to the four individual eTud domains of mouse TDRD1. Error bars indicate the standard deviation of two independent measurements. (C) Representative ITC experiments and fits to the data for the four single eTud domains (TD1, TD2, TD3, and TD4) of TDRD1 with the R74me2 peptide of MILI. The data were fitted to a single-site binding model. (D) ITC data for the double eTud domains TD2–3 and TD3–4 with the R74me2 peptide and with the R4574me2s doubly methylated peptide of MILI. The sequence of R4574me2s is indicated below the corresponding curves. The model for fitting the tandem TD2–3 and TD3–4 domains with the R74me2 ligand assumes two different binding sites as the peptide binds differently to each individual Tudor domain. For the R4574me2s peptide binding, a single-site binding model was used (one double methylated peptide to one tandem domain).

and it is reported to interact with all three Piwi proteins. In embryonic germ cells, TDRD1 associates with MILI and MIWI2, proteins that participate in secondary piRNA biogenesis (Vagin et al. 2009). Indeed, loss of *Tdrd1* results in impaired biogenesis of MIWI2-bound piRNAs, reduced transposon DNA methylation, and LINE1 retrotransposon derepression (Reuter et al. 2009; Vagin et al. 2009). Such male mice are infertile. The fish ortholog of TDRD1 is also reported to associate with ping-pong Piwi partners (Huang et al. 2011). Zebrafish lacking *Tdrd1* progressively lose germ cells and display transposon derepression, indicating a conserved role for TDRD1 in piRNA biogenesis.

Given that multiple sDMA residues in various sequence contexts decorate the N terminus of MILI, distinct specificities of the individual Tudor domains might drive complex formation. Here we used recombinant mouse TDRD1 Tudor domains, individually or in tandem, to examine their specificities for different sDMA-containing MILI peptides. Crystallographic studies of TD3 in its free and MILI sDMA-containing peptide-bound form reveal a unique orientation of the peptide as a whole and also of the sDMA in the aromatic cage. Interestingly, biochemical studies demonstrate that only TD2–4 of TDRD1 are functional in sDMA binding, while in TD1 a single amino acid substitution in the aromatic cage severely impairs sDMA binding. NMR experiments on the TD2 and TD3 domains confirm the crystallographic and thermodynamic findings of the interactions in solution. Finally, SAXS measurements on a construct containing all four tandem Tudor domains reveal a flexible elongated entity, lending support to a molecular scaffolding role for TDRD1 in the piRNA pathway.

RESULTS

Recognition of sDMA-peptides by single Tudor domains of TDRD1

The presence of multiple methylation marks at the N terminus of MILI prompted us to question how they interact with each eTud domain of TDRD1. To address this issue we first expressed and purified a construct spanning all four eTud domains of mouse TDRD1 (residues 232–1094). By partial proteolysis combined with bioinformatics analysis based on known eTud structure (Liu et al. 2010b), we identified well-behaved and soluble constructs of each of the four TDRD1 eTud domains separately (see Materials and Methods). We then measured by Isothermal Titration Calorimetry (ITC) the affinity of each single domain to three different MILI peptides previously shown to contain in vivo a symmetrical-dimethylated arginine (Vagin et al. 2009). The peptides tested were at successive positions and denoted R45me2 (38-GRAGPAGRme2GLVFR), R74me2 (69-LVSMFRme2GMGLDT), and R83me2 (76-MGLDTAFRme2PPSKR) (Fig. 1A). It is noteworthy that, whereas previous structural studies have used peptides in which the methylated arginines

have been flanked on both sides by glycine or alanine residues (Liu et al. 2010a,b), this is only true of the R45me2 MILI peptide.

The ITC results are summarized in Figure 1B with representative experimental data and fits shown in Figure 1C. The R45me2 peptide had the highest affinity for TD2 (K_D 55 μ M) and TD3 (172 μ M) domains, with TD4 having a slightly lower affinity (252 μ M) and the TD1 domain showing poor binding (685 μ M). The R74me2 peptide, in which the Rme2 is only flanked on one side by glycine, showed high affinity only for the TD3 domain (35 μ M), with the other domains showing substantially weaker binding. Indeed, when we titrated the R74me2 against the tandem domain constructs TD2–3 and TD3–4 the measured K_D of the major binding site in a two-site model is similar to that of the single TD3 domain (respectively, 45 and 51 μ M), with a binding to a second site being six- to 10-fold weaker (Fig. 1D). The R83me2 peptide, which has an unusual Phe–Rme2–Pro tripeptide, showed low affinity to all single Tudor domains: TD3 domain had the lowest K_D (245 μ M), with TD2 (536 μ M) and TD4 (831 μ M) following. TD1 domain, as for the rest of the peptides, showed a much higher K_D value (2.3 mM) (Fig. 1B). These results clearly indicate that the central TD2 and TD3 domains of TDRD1 have the highest affinities for the tested sDMA containing peptides of MILI, with TD4, and especially TD1, having significantly lower affinity.

TD2 and TD3 showed the highest affinity for consecutive singly arginine-methylated MILI peptides, respectively, R45me2 and R74me2. This prompted us to test the binding of tandem TD2–3 and TD3–4 domains to a doubly methylated peptide 43–77 of MILI (43-AGR45me2...FR74me2GMG-77, denoted R4574me2s), which encompasses both the original peptides, for indications of cooperativity. Satisfactory fits were achieved with one binding site (Fig. 1D) and gave a K_D of 48.5 μ M for TD2–3. This is comparable to the single TD2–R45me2 and TD3–R74me2 affinities (respectively, 55 and 35 μ M), showing that there is no enhancement of affinity. For TD3–4, titration of R4574me2s showed a K_D of 29 μ M, which is comparable to the TD3–R74me2 interaction (35 μ M) and consistent with, at most, weak cooperativity from TD4 binding (252 μ M for R45me2). Closer scrutiny of the thermodynamic binding parameters shows a favorable additive effect of the binding enthalpies from the single Tudor domains to the double-domain constructs, but that this is compensated for by unfavorable entropic contributions. Taken together, these results suggest that the central TD2 and TD3 domains bind methylarginine peptides essentially independently.

Crystal structure of the TD3–R45me2 peptide complex

In order to better understand the recognition of MILI-derived sDMA-peptides by the Tudor domains of TDRD1,

we tried to co-crystallize each of the domains with the different peptides. Only TD3 (residues 692–892), in the presence of the R45me2 peptide and without (residues 692–917), gave good quality crystals. Since molecular replacement with the known eTud structures did not work, the structure of the TD3–R45me2 peptide complex was solved de novo using selenomethionine-labeled protein and the MAD method and refined at 2.1 Å resolution (Table 1). Subsequently, the unbound TD3 domain was solved, in a different space group, by molecular replacement and refined at 2.8 Å resolution. Unexpectedly, the structure reveals a unique orientation of the bound peptide compared with the reported eTud–sDMA–peptide structures of fly TUDOR–Aub and human SND1–MIWI complexes (Liu et al. 2010a,b).

The TDRD1 TD3 domain folds into a bilobal α/β -extended Tudor (eTud) structure with 196/201 residues visible (697–892). It comprises a Tudor core domain (residues 697–702, 731–818) linked by three connections to a SN (staphylococcal nuclease)-like domain (703–730, 819–892) (Fig. 2A). The peptide is bound in a cleft between these two lobes and makes contact with both. The SN-like domain of TD3 is

comprised of two β strands ($\beta 1$ and $\beta 2$) exposed to the peptide-binding cleft, and three β strands ($\beta 8$ – $\beta 10$) and two α helices ($\alpha 2$ and $\alpha 3$) at the C-terminal part of the structure. The first two β -strands are connected to the Tudor core domain via a long α helix ($\alpha 1$). The Tudor core domain is a barrel comprised of four β -strands ($\beta 3$ – $\beta 6$), each of which contributes key residues to build the hydrophobic/aromatic cage that binds the methylated arginine side chain (Fig. 3A).

As expected, the overall structure of TD3 resembles the known structures of the 11th extended Tudor (eTud11) domain of *Drosophila* TUDOR (PDB 3NTI) and human SND1 (PDB 2OMC) with RMSDs of, respectively, 1.72 Å and 1.61 Å, for, respectively, 153 and 173 matched residues (Fig. 2B). The TD3 SN-like domain superposes well with staphylococcal nuclease (PDB 3T16, RMSD 1.70 Å for 84 residues aligned), but, like eTud11, lacks the long C-terminal helix, which is, however, found in the SND1 structure. Despite the overall similarity there are some notable differences between TD3 and the other known eTud domains. Firstly, TD3 has an N-terminal extension (residues 697–707) beyond the N-terminal end of the eTud11 and SND1 structures, which crosses back from the SN-like

TABLE 1. Crystallographic data and refinement statistics

	TD3 + R45me2			TD3
	Native	SelenoMet peak	SelenoMet inflection	Native
Space group	<i>P</i> 1	<i>P</i> 1	<i>P</i> 1	<i>P</i> 4 ₁ 2 ₁ 2
Cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å)	39.94, 53.96, 60.79	40.05, 53.63, 61.04	40.51, 54.29, 61.56	62.96, 62.96, 132.48
α β γ (°)	93.69, 98.56, 111.67	93.04, 98.42, 111.97	93.27, 98.22, 111.76	90.00, 90.00, 90.00
Resolution (Å)	50–2.1 (2.2–2.1) ^a	50–2.6 (2.7–2.6)	50–2.9 (2.8–2.7)	50–2.8 (2.9–2.8)
<i>R</i> _{merge}	11.3 (39.3)	11.7 (46.9)	12.3 (42.8)	7.8 (60.3)
<i>I</i> / σ <i>I</i>	7.28 (2.55)	10.05 (2.50)	10.42 (2.2)	14.80 (2.75)
Completeness (%)	95.5 (95.8)	98.5 (97.8)	88.2 (95)	98.8 (93.0)
Redundancy	2.4 (2.4)	3.32 (3.3)	3.51 (2.08)	5.87 (6.06)
Refinement				
Resolution (Å)	37–2.1			132.28–2.8
No. refls used (free)	24480 (1275)			6638 (330)
<i>R</i> _{work} / <i>R</i> _{free}	21.9/28.4			22.3/24.4
No. atoms total	3415			
Protein	3089			1544
Peptide	128			—
Water	186 + 2 glycerol			—
Average <i>B</i> -factors				
Protein	18.3			69.4
Peptide	19.4			—
Water	21.3			—
RMSDs				
Bond lengths (Å)	0.01			0.007
Bond angles (°)	1.38			1.011
Ramachandran				
Favored (%)	97.5			93.2
Allowed (%)	100			98.9

^aValues in parentheses are for the highest resolution shell.

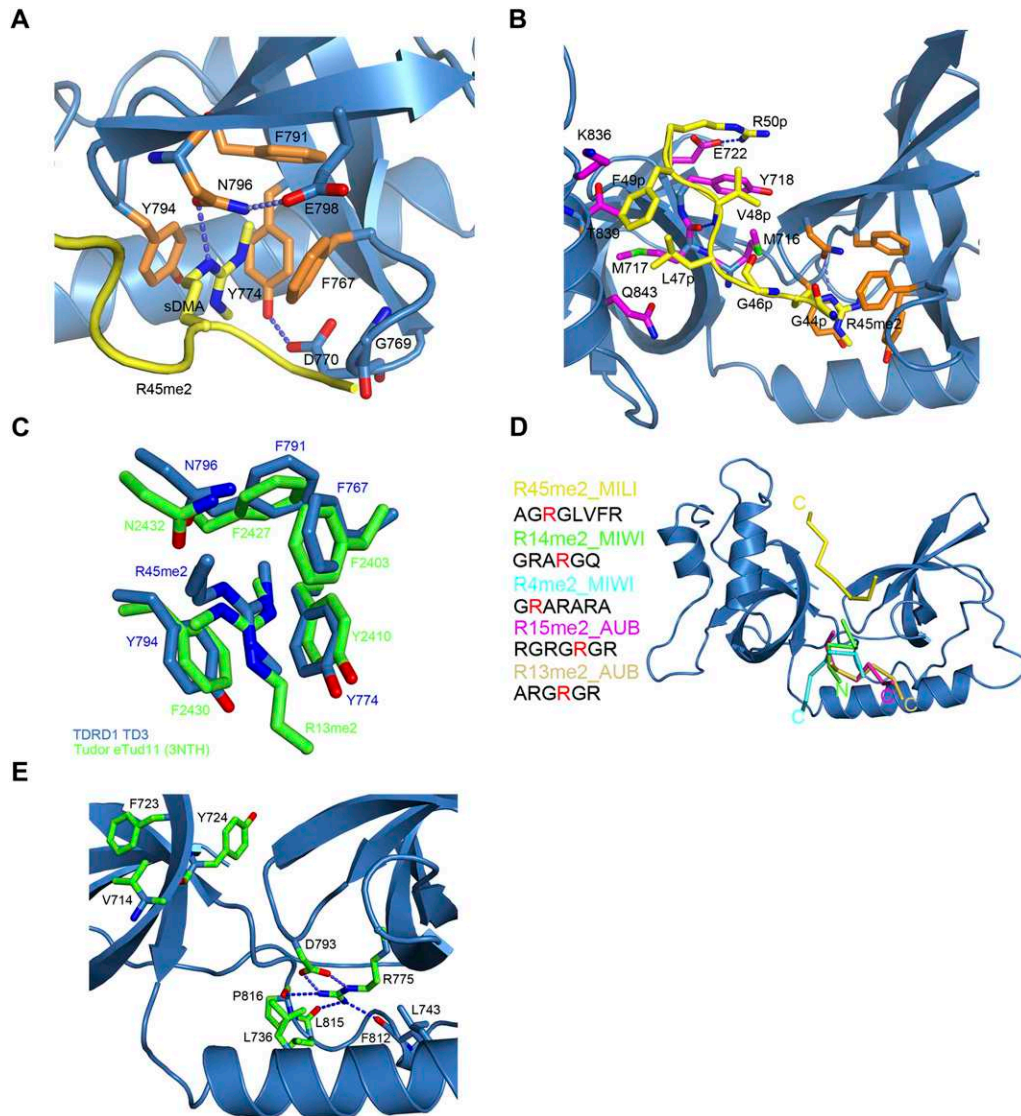


FIGURE 3. TD3 binds the R45me2 peptide in a different orientation from that previously observed in eTud domains–peptide complexes. (A) The sDMA in the aromatic cage of TD3 colored as in Figure 2A. The dipeptide GD stabilizing the cage and E798 interacting with N796 are also shown. (B) Details of the interactions between the R45me2 peptide and TD3. Colors are as in A with additionally residues of the SN-like domain interacting with the peptide being in magenta. Blue dashed lines indicate putative hydrogen bonds. (C) Structural comparison of the aromatic cages of TD3 (blue) and eTud11 (green) in complex with their corresponding peptides highlighting the different direction of entry of the two methylarginine groups. All other such structures exhibit the same conformation as eTud11/R13me2. (D) Diagram showing the orientation of various sDMA-containing peptides with reference to the TDRD1 TD3 structure (blue) obtained by superposition of eTud–peptide complexes of known structure. R14me2 (PDB 3OMG, green) and R4me2 (PDB 3OMC, cyan) were co-crystallized with the SND1 protein, while R15me2 (PDB 3NTI, magenta) and R13me2 (PDB 3NTH, wheat) were crystallized with eTud11 from *Drosophila* TUDOR. All of the above-mentioned peptides enter the aromatic pocket from the “bottom,” while the R45me2 peptide (yellow) enters from the “top” of the structure. The observed peptide residues in each structure are shown at the left with the sDMA highlighted in red. N- and C-termini of the peptides are also indicated. (E) Conserved residues in murine TDRD eTud domains. In a stick-model representation, the residues conserved in most of the eTud domains of TDRD proteins are shown (green). The absolutely conserved R775 is involved in extensive interactions with residues of the Tudor domain, thus stabilizing the aromatic cage. Blue dashed lines indicate hydrogen bonds.

cage are highly conserved in all Tudor domains that are known to bind sDMAs. The methylated side chain enters the cage through an opening from the top of the structure and is involved in extensive cation- π and hydrophobic interactions with the aromatic residues F767, Y774, Y794, and F791, respectively, which constitute the cage (Fig. 3A).

Y774 forms the back of the cage, with Y794 and F767 sandwiching the methylated side chain from each side. The hydroxyl group of Y794 hydrogen bonds to the of helix α 1 residue N740. F791 and the highly conserved asparagine N796 form the ceiling of the binding pocket. The dipeptide, 769-GD, reinforces the cage with G769

peptide stacking on F767 and D770 hydrogen bonding to Y774 (Fig. 3A).

By superimposing the TD3–R45me2 structure with the eTud11–R13me2 or SND1–peptide structures, a remarkable observation is that the methylarginine groups do not overlay, but the methylated guanidine group is rotated 120° about the CZ carbon (Fig. 3C). This is because the side chain enters the cage from a completely different direction. Thus, one methyl group of TD3–R45me2 (CQ2, which points toward F791) coincides with CQ1 in the other structures, whereas CQ1 of TD3–R45me2 coincides with the CD in the other structures (and, reciprocally, the CD of TD3–R45me2 coincides with CQ2 in the other structures) (Fig. 3C). A consequence of this is that the conserved asparagine N796 (which is held in position by an interaction with E798) makes a hydrogen bond to the N^ε atom of the arginine rather than to N^{η1/2} as in other Tudor domain structures. Nevertheless, N796 remains critical for methylarginine binding, as reported for related structures, since the single mutant N796A is enough to significantly weaken the interaction with the R45me2 peptide ($K_D \sim 1.4$ mM) (Fig. 4A).

Recently, solution NMR structures of the core Tudor domains of SMN and SPF30 bound to isolated sDMA or aDMA (asymmetrically dimethylarginine) residues have been described (Tripsianes et al. 2011). Interestingly, the orientation of the aliphatic part of sDMA bound to SMN (PDB 4A4E for SMN) is found to differ from that of aDMA (PDB 4A4G). sDMA is bound to SMN in a similar orientation to that observed in the previous eTud–sDMA peptide complexes (Liu et al. 2010b), except that the two methyl-groups are in the *anti-anti* conformation in SMN compared with the *anti-syn* conformation in the eTud complexes. It has been proposed that the latter difference is due to the systematic occurrence in the aromatic cage of single Tudor domain proteins of a tryptophan in place of the phenylalanine most commonly found in eTud domains (F767 in TD3) (Tripsianes et al. 2011). However, the aDMA side chain is bound in SMN in a rotated configuration, which is much closer to that observed for sDMA bound to TD3. In the TD3 complex, the sDMA nevertheless maintains the *anti-syn* conformation characteristic of eTud domains (Tripsianes et al. 2011). The significance of this correspondence with aDMA binding to core Tudor domains is not clear.

Orientation of the R45me2 peptide in the TD3 domain structure

The MILI R45me2 peptide in the crystal structure reveals seven ordered residues, 44-GRme2GLVFR-50, with A43 poorly ordered. As indicated above, it binds to TD3 in a unique orientation when compared with other known eTud/sDMA–peptide complexes. The C terminus of the peptide is at the “top” of the TD3 domain, whereas in the other structures, the peptide is bound at the “bottom” of the eTud domain (Fig. 3D). However, as in the other known structures, apart from the

sDMA, the peptide makes additional contacts with residues in the SN domain as well as the Tudor core domain (Fig. 3B).

The peptide plane of G44p (p denotes a peptide residue), which precedes the methylated arginine, is stabilized by hydrophobic contact with the tip of F767, part of the aromatic cage. This interaction is a consequence of the rotated position of the methylated arginine in the cage and requires the presence of G769 (mentioned above) to avoid steric clash. In SND1 and eTud11 this glycine is substituted by a valine or glutamate, respectively, which would hinder peptide binding (Fig. 2D). Residues C-terminal to the R45me2 form an extended chain and interact principally with residues 716–718 from the distal part of strand β 1 and residues 836–843 from the proximal part of helix α 2 of the SN-like domain. G46p and L47p make hydrophobic interactions with M716 (a hydrophilic E or Q in other structures) and M717, respectively, both from strand β 1, as well as Glu843 on helix α 2 (Fig. 3B). The amino group of V48p hydrogen bonds with the carbonyl oxygen of M717, and its side chain makes hydrophobic interactions with M716 and Y718 (Fig. 3B). Next, the benzene ring of F49p, which is stacked on L47p, makes van der Waals contacts with K836, which is on helix α 2. The last ordered amino acid of the peptide is R50p, which points out of the structure into the cleft between the SN-like domain and the Tudor core domain. Its amino group makes a hydrogen bond with the carbonyl oxygen of Y718 and its carbonyl group makes a water-mediated hydrogen bond with T839, which is also on helix α 2. Finally, the side chain of R50p stacks on Y718 (strand β 1) and makes a salt bridge with the carboxyl group of E722 (strand β 2) (Fig. 3B). It should be pointed out that there are additional interactions of the peptide with a crystallographic related TD3 molecule (denoted *); these involve two main-chain interactions with the main chain of I781* and a salt bridge of Glu722* with R50p. For reasons given immediately below we do not think that these interactions strongly influence the mode of peptide binding observed.

A significant difference between the TD3 structure and those of eTud11 and SND1 is the orientation of helix α 2 of the SN-like domain (Fig. 2B). For SND1 and eTud11 that accommodate their peptides from the bottom (Fig. 3D), the peptide is contacted by residues from the C-terminal (distal) end of this helix (e.g., Asn823 in SND1 equivalent to G848 in TD3) (Fig. 2D) as well as proximal parts of strand β 1 (e.g., F686 in SND1, equivalent to V713 in TD3). In TD3 the helix is tilted such that its N-terminal (proximal) end contacts the peptide together with the distal part of strand β 1 (as described above), but its distal end is shifted away from the binding cleft and does not contact the peptide (Figs. 2B, 3B,D). This appears to be an important structural adaptation together with, for instance, the presence of M716 and G769 (mentioned above), which permits and accommodates the unusual orientation of the peptide in the TD3 structure. In this respect, we also note that the bound R45me2 peptide is of the form **Rme2G Φ X Φ X**, where Φ is

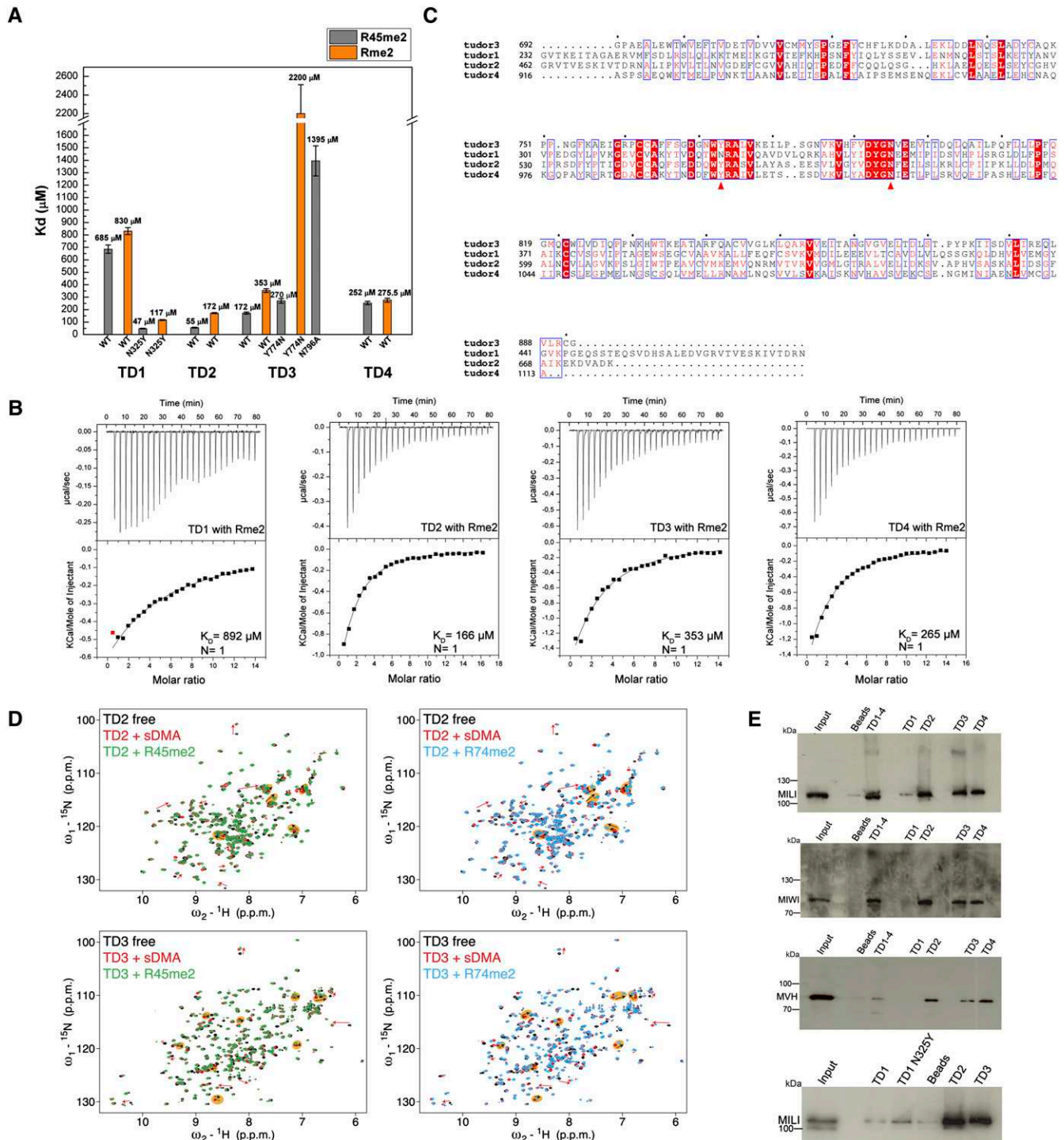


FIGURE 4. “Active” and “inactive” extended Tudor domains of TDRD1. (A) K_D s derived from ITC measurements for the binding of isolated symmetrically dimethylated arginine (Rme2) and R45me2 peptide to individual wild-type (wt) TDRD1 TDs and N325Y, Y774N, and N796A mutants. Error bars represent SD values from two experiments. (B) Representative ITC experiments and fits to the data. The four single eTud domains (TD1, TD2, TD3, and TD4) of TDRD1 with the isolated symmetrically dimethylated arginine (Rme2). (C) Multiple sequence alignment of the four individual Tudor domains of TDRD1. The first red arrow highlights the position of N325 of TD1, and the second arrow the Y774 of the aromatic cage of TD3. Mutations on both residues are critical for binding sDMAs (see A). (D) Binding of TD2 and TD3 domains of TDRD1 to sDMA-containing peptides monitored by NMR. Each panel shows an overlay of the ^1H , ^{15}N HSQC spectra of the respective domain when free (black), when saturated with an excess of naked sDMA (red), and when saturated with MILI derived sDMA-containing peptides (green or cyan). For well-resolved peaks, chemical-shift perturbations arising from sDMA contacts are annotated with red arrows and those induced by interactions with flanking residues with black arrows in orange background. Notice that other peaks are affected by both sDMA and flanking residues but are not labeled. (E) Pull-down assays of endogenous murine proteins MILI, MIWI, and mouse Vasa homolog (MVH) by individual and multiple TDs of TDRD1. His-tagged constructs with four Tudor domains (TD1–4) and single Tudor domains (TD1, TD1 N325Y mutant TD2, TD3, TD4) were used. Size markers in kilodaltons are indicated.

a largish hydrophobic residue (M, L, F), which is quite different from bound peptides in other known complex structures. Interestingly, the R74me2 peptide 74-Rme2GMGLD, which binds the best to TD3 (Fig. 1B), also has this pattern. This suggests that TD3 and the path the bound peptide takes are adapted to accommodate these hydrophobic residues. Indeed, some other nontested modified peptides of MILI conserve at least the first hydrophobic position, e.g., 95-Rme2GVLG, 100-Rme2GLSA, 163-Rme2GMDK. Thus, these peptides may all bind to TD3 in a similar way to that observed for the R45me2 peptide, although it is possible that other peptides without these special features might bind in a different manner.

To test the hypothesis of a similar binding mode for R45me2 and R74me2 to TD3, we recorded solution NMR ^1H , ^{15}N HSQC spectra to monitor the binding of sDMA and MILI sDMA-containing peptides to TD2 and TD3. Each domain binds the respective ligands (sDMA, R45me2, and R74me2) in the fast-exchange regime on the NMR chemical-shift time scale. In all cases the chemical-shift trajectories upon ligand titration are linear and occur at the same rate, indicating a single binding event (Supplemental Fig. 2). The apparent dissociation constants derived from the NMR titrations are generally in good agreement with those measured by ITC, except for an inversion of the relative affinities of TD3 for the two peptides (Table 2). By comparing the saturation points of each titration we can dissect chemical-shift perturbations originating solely by sDMA binding or, additionally, from contacts with the flanking residues present in the peptides (Fig. 4D). This analysis shows that for both TD2 and TD3 domains, the perturbations due to flanking residue interactions are the same irrespective of which peptide is bound (R45me2 or R74me2), indicating that a given eTud domain accommodates both peptides in the same orientation. For TD3 this is presumably the orientation observed in the crystal structure, although confirmation of this would require a full assignment of the NMR spectrum.

Structure of unliganded TDRD1 TD3

We solved the structure at 2.8 Å resolution of the TDRD1 TD3 domain in the absence of bound peptide. The construct

used was slightly longer (residues 692–917 instead of 692–892) and crystallized in a different space group. The structures are extremely similar (RMSD 0.65 Å for all 191 common visible residues), with small perturbations likely due to crystal contacts. The extra C-terminal residues are not observed, further suggesting that TD3 does not have the last C-terminal helix characteristic of SN domains, and is observed in SND1 but not eTud11. The peptide-binding site is unchanged except for a slight expansion of the aromatic cage, largely due to movement of the 767-FGD loop (Supplemental Fig. 1B). Thus, there is only a small element of induced fit in the peptide binding.

Conserved residues in eTud domains

The interactions of the R45me2 peptide, apart from that of the sDMA itself, clearly depend on the extended nature of TD3 and, indeed, it has previously been noted that the Tudor domains of the TDRD protein family are all found in the extended version (Liu et al. 2010a,b; Vourekas et al. 2010). We therefore used a multiple-sequence alignment of eTud domains from mouse TDRD family proteins to identify possible sequence signatures for such extended domains apart from the aromatic residues forming the cage (Supplemental Fig. 1A). Firstly, we identified highly conserved residues that seem important for establishing the eTud double-domain fold, and these include, in TD3, V714, 723-FY, L736, L743, R775, and 815-LP, whose location is indicated in Figure 3E. V714 and F723 are conserved residues from, respectively, strands $\beta 1$ and $\beta 2$ of the SN-like domain and form part of the hydrophobic core of the domain. Adjacent residue Y724 (W in several domains) makes van der Waals contacts across the interdomain interface with G795, a highly conserved feature of the aromatic cage. L736 and L743 are conserved hydrophobic residues found in the amphipathic helix $\alpha 1$ that connects the two domains and pack against the underside of the Tudor core domain, contributing to its hydrophobic core. Perhaps the most striking signature of the mouse eTud domains is the correlated conservation of R775, D793, and 815-LP (Supplemental Fig. 1A). The absolutely conserved R775 (adjacent to aromatic cage residue Y774) found in all eTud domains, but not in single Tudor domains, plays a key role in pinning together different parts of the structure and stabilizing formation of the aromatic cage. The side chain is largely buried, the aliphatic side chain makes hydrophobic contacts with L739, I808, and F817, and the guanidinium group makes a strong salt bridge with conserved D793 (adjacent to aromatic cage residue Y794), as well as three hydrogen bonds to the carbonyl oxygens of F812, L815, and P816 (Fig. 3E). The LP dipeptide is conserved in SND1, but not in eTud11. These considerations suggest that most of the Tudor domains of the mouse TDRD subfamily are, indeed, extended Tudor domains.

Secondly, we examined whether there are any conserved residues outside of the aromatic cage that are involved in

TABLE 2. Comparison of K_D and apparent K_D derived from, respectively, ITC and NMR data for TD2 and TD3

Tudor domain	Methylated ligand	K_D (μM) – NMR	K_D (μM) – ITC
TD2	sDMA	195 \pm 13	172 \pm 4
TD2	R45me2	75 \pm 12	55 \pm 3.6
TD2	R74me2	376 \pm 76	417 \pm 24
TD3	sDMA	795 \pm 37	353 \pm 14.5
TD3	R45me2	68 \pm 6	172 \pm 9
TD3	R74me2	230 \pm 67	35 \pm 1.8

extended peptide binding. Equivalent residues E2374 in eTud11 and E708 in SND1 (Fig. 2D), found in the connecting helix α 1, interact via hydrogen bonds with the nonmethylated R11 in the R13me2 peptide and R6 in the R4me2 peptide, respectively (Liu et al. 2010a,b). In TD3, the equivalent residue D737 does not interact with the peptide. Furthermore, Q2365 in eTud11 and Q699 in SND1, at equivalent positions on strand β 2, interact with R13 and A7 from, respectively, the R15me2 and R4me2 peptides. This residue is conserved as a glutamine in many mouse eTud domains, but is H726 in TD3 (Supplemental Fig. 1A). It does not bind the peptide, but preserves an important hydrogen-bond stabilizing residue, D765, one of the aromatic cage loop residues. The fact that these residues do not have conserved functions in TD3 with regard to peptide binding is not surprising given the very different mode of binding of the R45me2 peptide. Indeed, we have highlighted above adaptations in TD3 that appear to favor the mode of peptide interaction observed.

“Active” and “inactive” eTudor domains

Above, we have argued that all of the mouse TDRD domains are likely to have the eTud fold, but how many are actually active in binding sDMA-containing peptides? Our binding data clearly suggest that TD2, TD3, and TD4 can bind methylated arginine peptides of MILI, albeit with differing affinities. However, TD1 showed only weak binding to all of the peptides tested. To investigate this further we used ITC to measure the affinity of each separate TDRD1 eTud domain with free dimethylated (Rme2), monomethylated (MMA), or unmethylated arginines (Runmethylated) (Fig. 4A,B; Supplemental Fig. 3A). The single residues used were N-terminally acetylated and C-terminally amidated to mimic a peptide context. The isolated Rme2 bound to TD2, TD3, and TD4 with respective affinities of 172, 353, and 275.5 μ M, consistent with the results for the peptides and confirming that the non-sDMA parts of the various peptides tested contribute significantly to the affinity (Fig. 4A). The TD1 domain showed the weakest binding with \sim 830 μ M affinity (Fig. 4A), again suggesting that it may not be involved in methylation-dependent interactions. As expected, the monomethylated (MMA) and unmethylated arginine (R-unmethylated) showed very low and undetectable binding, respectively, to each eTud domain (Supplemental Fig. 3A). This holds true also in a peptide context, since the unmethylated R74 peptide displayed no binding to the TD3 domain, thus highlighting the indispensable role of the symmetrically dimethylated arginine for these interactions (data not shown).

Inspection of the TD1 sequence in comparison to the other three TDRD1 domains shows that N325 is found in place of a usually conserved tyrosine (Y774 in TD3) in the aromatic cage (Fig. 4C). An asparagine at this position is found in mammalian (human, mouse, and bovine) TDRD1,

but zebrafish and medaka TDRD1 maintain the tyrosine. To test whether this substitution was responsible for the observed low-binding affinity, we made the mutation N325Y in TD1. ITC measurements with an isolated Rme2 or with the R45me2 peptide show that this single substitution restores binding comparable to the best of the other eTud domains, with the affinity for Rme2 being 117 μ M and for the R45me2 peptide 47 μ M (Fig. 4A). On the other hand, when the reverse substitution, Y774N, was made in the TD3 domain, binding to the single Rme2 was very weak (2 mM), although the affinity to the R45me2 peptide was reduced from wild type, but not dramatically (270 μ M), showing that the additional affinity for the non-sDMA residues could maintain the binding (Fig. 4A). This was not the case, however, for the TD3 mutation Y774A, which failed to bind to either Rme2 or R45me2 (data not shown).

The results above show that substitutions of the conserved asparagine (e.g., N796A in TD3) or an aromatic cage tyrosine (e.g., N325 in TD1, Y774A in TD3) drastically reduce sDMA binding, rendering the domain effectively “inactive.” Indeed, the signature motif of “active” Tudor domains appears to be $\text{FX}_n\text{W}(\text{F})\text{YRX}_n\text{F}(\text{Y})\text{XDY}(\text{F})\text{GN}$, consistent with the fact that all reported crystal structures have these residues. On this basis we predict that at least half of the mouse TDRD protein eTud domains have very low or lack binding affinity for sDMA due to the presence of one or more significant deviations from the “active” motif (Supplemental Fig. 1A).

Interaction of endogenous murine proteins by TDRD1 eTud domains

In order to further investigate whether TD1 is “active” or not in the binding of methylated marks on piRNA pathway proteins, we performed pull-down assays from the lysate of adult mouse testes. We tested the binding of the four tandem eTud domains of TDRD1 (TD1–4), as well as each single eTud domain, to endogenous murine proteins MILI and MIWI, both of which are reported to contain sDMAs. Indeed, the presence of arginine modifications on these proteins was confirmed by Western blotting (Supplemental Fig. 3B). Our results show that both proteins can bind TD1 TD4 and single domains TD2, TD3, and TD4, presumably via the methylation marks, but only very weakly to TD1 (Fig. 4E). The germline-specific RNA helicase mouse Vasa homolog (MVH) is a secondary piRNA biogenesis factor (Kuramochi-Miyagawa et al. 2010) that is reported to carry both symmetrical and asymmetrical dimethylarginine modifications (Kirino et al. 2010); the presence of these marks was again verified by Western blotting (Supplemental Fig. 3B). MVH associates with TDRD1 and TDRD6 and is a component of MILI and MIWI complexes (Leroy et al. 1989; Tanaka et al. 2000; Kirino et al. 2010). As shown in Figure 4E, while MVH was retained on beads coated with TD2, TD3, or TD4, it failed to interact with TD1.

Furthermore, when pull-down assays were performed with the “reactivated” TD1 domain mutant N325Y and MILI, we observed an increase of the binding signal, although still less than for TD2 and TD3 (Fig. 4E). Thus, we conclude that the TD1 domain of TDRD1 is unlikely to be involved in binding components of the piRNA pathway that are known to be arginine methylated, but it cannot be ruled out that it binds a peptide from some other associated protein.

SAXS analysis of the tandem Tudor domains of TDRD1

We used small-angle X-ray scattering (SAXS) to determine the overall conformation of the tandem Tudor domains of TDRD1 in solution (Fig. 5). SAXS data were measured on both the four-Tudor domain construct (TD1–4) and the tandem TD1 and TD2 domains (TD1–2). Ab initio modeling was done with DAMMIF, and 40 individual models (for

both TD1–4 and TD1–2) were averaged using the program DAMAVER. The output of DAMAVER was refined using it as a starting model for the program DAMMIN to yield the final calculated envelope (see Materials and Methods). The envelope together with model independent parameters such as radius of gyration and Dmax (Table 3) show that TD1–4 has an elongated shape with four distinct lobes that are comparable in size to individual eTud domains and arranged end-to-end (Fig. 5A). The envelope derived from the TD1–2 data is consistent with a double-domain structure similar to either the first two or last two domains of TD1–4 (Fig. 5B).

To better characterize the molecular envelope of TD1–4, we performed rigid body modeling using the program CORAL (Petoukhov and Svergun 2005). This program translates and rotates the atomic models of individual domains against the SAXS data, with the missing flexible interdomain linkers being modeled as random chains. Models for TD1, TD2, and TD4 were derived by homology

modeling from the crystal structure of the *D. melanogaster* Tudor-SN protein (PDB code: 2WAC), while for TD3 the crystal structure reported here was used. Rigid body modeling simultaneously to both the TD1–4 and TD1–2 data gave χ^2 of 2.17 for the TD1–4 data and χ^2 of 1.96 and 1.82 for the TD1–2 data (respectively, modeled as TD1–2 or TD3–4) (Fig. 5A,C). The fit was not significantly improved, nor the four-domain model changed when only the TD1–4 data were used ($\chi^2 = 2.05$). Although the rigid body fits are of good quality, they are some systematic deviations that are likely due to flexibility of the interdomain linkages. Similarly, the CORAL model fits well within the independently derived ab initio TD1–4 average envelope, although there are some deviations, again consistent with flexibility, with respect to the most representative (filtered) envelope (Fig. 5A). These analyses show that the four-tandem eTud domains of TDRD1 are in an extended conformation in solution, with each domain being potentially accessible for peptide binding.

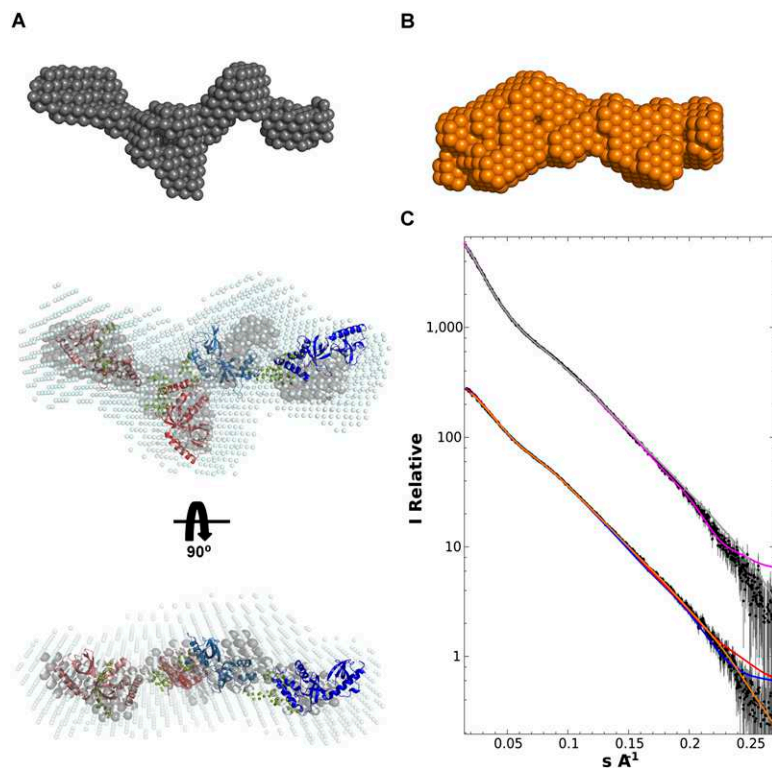


FIGURE 5. Overall architecture of the TD1–4 domains of TDRD1 as determined from small-angle scattering data. (A) TD1–4 ab initio models showing “most representative (filtered)” (gray) and “average” envelope (light blue) as given by DAMAVER. The rigid body model was produced using CORAL (Petoukhov and Svergun 2005). TD1 (salmon), TD2 (light red), and TD4 (deep blue) models were based on Tudor-SN (pdb code 2WAC). TD3 (light blue) was fitted using the crystal structure determined here. Flexible linkers connecting the eTud domains, positioned by CORAL, are depicted as spheres (olive green). (B) Ab initio model (orange) derived from the TD1–2 SAXS data. This is consistent with either half of the TD1–4 model. (C) SAXS data (black dots with error bars) and fits for TD1–4 and TD1–2 (solid lines). (Top) TD1–4 data. Fit of the ab initio model shown in A (gray) and the multidomain model created by CORAL (purple). (Bottom) TD1–2 data. Fit of the ab initio model shown in B (orange). Fits of the tandem TD1–2 and TD3–4 parts of the CORAL model are, respectively, shown in red and blue.

DISCUSSION

Transposon silencing in mice is initiated by de novo DNA methylation of transposon promoters in the embryonic germline. Repeat-derived piRNAs associating with the nuclear Piwi protein MIWI2 are implicated in specifying this event (Aravin et al. 2008; Kuramochi-

TABLE 3. Model independent parameters derived from SAXS data

Parameter	Source	TD1–2	TD1–4
R _g (Å)	Guinier	39.2 ± 3	54.2 ± 3
R _g (Å)	Porod	41 ± 4	54 ± 4
l _o	Guinier	31400 ± 600	69300 ± 600
l _o	Porod	32600 ± 2000	68700 ± 2000
Volume (Å ³)	Porod	69000 ± 4000	136000 ± 4000
Molecular weight (kD)	From amino acid sequence	49.2	95.5
	Calculated from l _o relative to BSA	49 ± 4	95 ± 5
D _{max} (Å)	GNOM	133 ± 5	200 ± 5

Miyagawa et al. 2008). Primary biogenesis feeds many of the piRNAs entering MILI, and some of these can initiate generation of secondary piRNAs entering MIWI2 via the ping-pong pathway. Briefly, this involves the transfer of a target RNA endonucleolytic cleavage fragment from MILI RNPs to MIWI2, where it would mature as a new secondary piRNA. Two additional factors, TDRD1 and MVH, are also implicated in production of MIWI2-bound piRNAs. Infact, *Mili*, *Mili*^{DAH} (*MILI* slicer inactive), *Tdrd1*, and *Mvh* mouse mutants all display a failure to load MIWI2, leading to it being mislocalized to the cytoplasm (Aravin et al. 2008; Kuramochi-Miyagawa et al. 2008; Reuter et al. 2009; Vagin et al. 2009; De Fazio et al. 2011). This leads to transposon derepression and hypomethylation of promoter elements in the mutants.

The occurrence of four-tandem extended Tudor domains in TDRD1 and a variety of arginine methylation marks on Piwi and Piwi-associated proteins poses the question of how interactions between the two could help mediate the formation of functional complexes in the piRNA pathway. One prevailing model proposes a molecular scaffold role for proteins like TDRD1, whereby each of the four Tudor domains might specify a unique interaction with a distinct binding partner, such as MILI and MIWI2, facilitating, for example, their interaction for ping-pong piRNA biogenesis. In this respect we note that although MIWI2 is detected in TDRD1 complexes (Vagin et al. 2009), it is unclear how it interacts, as arginine modification is not reported for MIWI2.

To provide further insight into the role of the tandem Tudor domains of TDRD1 we have measured the binding of each of the four domains separately to three different symmetrically dimethylated peptides of MILI and to isolated symmetrically dimethylated arginine. Our data show that whereas TD1 binds very poorly to all peptides, TD2, TD3, and TD4 bind with moderate micromolar affinity to the three peptides of MILI in vitro. More specifically, we find that whereas favorable flanking regions of a methylated peptide can enhance specificity only moderately upon binding of a single sDMA (e.g., factor of 1–3 for R45me2 or Rme2 binding to the four domains) (Fig. 4A), unfavor-

able flanking residues can decrease affinity, relative to Rme2. Consequently, different sDMA-containing peptides can modulate the affinity by, at most, an order of magnitude to a given domain, depending on the compatibility of these flanking residues with the binding site (Fig. 1B). Thus, there is some limited preference of a given domain for a given peptide, although it must be borne in mind that we have not tested all relevant methylated peptides. In particular, we suggest that TD3 preferentially recognizes peptides with

hydrophobic residues just C-terminal to the sDMA. Furthermore, our ITC measurements revealed an apparent preference for binding of isolated TD2 and TD3 to consecutive MILI peptides, R45me2 and R74me2, respectively. Therefore, we examined whether TDRD1 can integrate inputs from multiple Tudor domains by measuring affinity of a fused peptide combining both of these methylation marks to the tandem TD2–TD3 domains. Somewhat surprisingly, no significantly enhanced affinity was observed, suggesting that binding to each domain is largely independent, probably due to peptide and interdomain flexibility. However, it should be borne in mind that this conclusion is based only on one combination of dually methylated peptide and tandem domains. Our pull-down experiments clearly show that domains TD2, TD3, and TD4 are individually each capable of interacting with endogenous, arginine-methylated MILI, MIWI, and MVH. Taking these results together, we conclude that a 1:1 exclusive interaction of a Tudor domain with a particular sDMA-peptide is unlikely. We favor the hypothesis that TDRD1 can act as a general scaffold capable of engaging one partner through a set of composite contacts or binding multiple partners simultaneously. These various options may be appropriate at different times to create unique complexes or interaction networks dynamically, depending on partner availability.

The crystal structure of the TD3–R45me2 complex unexpectedly reveals a completely different orientation of the peptide compared with other such structures, even with the methylated arginine entering the aromatic cage from a different direction. This is perhaps an adaptation to preferentially bind peptides with large hydrophobic residues just downstream from the methylated arginine as present in the R45me2 and R74me2 peptides and highlights the potential of the extended Tudor domain architecture to adopt multiple binding modes.

The elongated conformation of the SAXS model of TDRD1, with each Tudor domain accessible to receive its interacting partner, provides the strongest evidence to date of the molecular scaffold model. This is further supported by our binding studies with the various Tudor domain constructs and full-length Piwi proteins (*MILI* and *MIWI*)

and biogenesis factors like MVH present in mouse testes lysates. Thus, it is likely that TDRD1 provides an environment in which the interacting proteins can engage in intermolecular interactions, driving biogenesis of piRNAs.

Our observation that TD1 of mouse TDRD1 is essentially unable to bind methylated arginine peptides is in concordance with previous predictions that many eTud domains in TDRD proteins lack the consensus aromatic cage residues (Liu et al. 2010a; Tripsianes et al. 2011). A single amino acid substitution (N325 instead of Y) within the aromatic cage of TD1 renders it ineffective. However, fishes (separated from humans by 450 million years) have conserved a tyrosine at the same position. As we demonstrated, changing N325 back to Y could restore methylated ligand-binding potential to the mouse TD1, suggesting that the “active” aromatic cage is the major determinant of eTud domain binding and, to a lesser extent, the flanking residues. The functional relevance of such “inactive” domains remains to be determined, but it is possible that they allow methylation-independent interaction with Piwi proteins or other piRNA pathway-associated factors. Indeed, the fly Tudor domain protein Yb lacks any “active” Tudor domains, but is still detected in Piwi complexes and is essential for primary piRNA biogenesis (Saito et al. 2010; Handler et al. 2011). All of these observations point to an evolving role for Tudor domains, some of which might be adopting new roles that go beyond recognition of methylated ligands.

MATERIALS AND METHODS

Protein expression and purification

TD1–4 (four tandem eTud domains of TDRD1)

The four tandem eTud domains of mouse TDRD1 (denoted TD1–4, residues 232–1094) were cloned into pETM11 vector to generate N-terminal His-tagged fusion protein. The recombinant protein was grown in *Escherichia coli* BL21 (DE3) cells at 37°C and induced with 1 mM IPTG at 20°C for 14 h. Purification was by affinity chromatography on nickel Sepharose fast-flow resin followed by TEV protease cleavage to remove the tag. A further purification step was made by size-exclusion chromatography. Finally, the protein was concentrated as required in buffer containing 20 mM Tris-HCl (pH 7.6), 0.15 M NaCl, and 1 mM DTT.

Identification of TDRD1 extended Tudor-domain pair and individual domain constructs

To experimentally define domain boundaries, multidomain constructs were iteratively subjected to limited proteolysis by trypsin with 1:1000 trypsin:protein ratio. Stable fragments were identified by N-terminal sequencing and mass spectrometry. Partial proteolysis of the TD1–4 construct led first to a stable fragment encompassing the first two Tudor domains (TD1–2, 232–677). Further limited proteolysis by trypsin on the TD1–2 construct gave rise to single domains TD1 (232–476) and TD2 (463–677). Digestion of a construct encompassing TD3–4 (692–1050) yielded

a stable fragment containing TD3 (692–892) that was eventually crystallized with the R45me2 peptide. The apo-TD3 structure was obtained by crystallizing a longer construct (Apo-TD3, 692–917), which along with TD4 (916–1113) was designed taking into account secondary structure predictions. Once all single domains had been defined, these were used to design tandem domains 2 and 3 (TD2–3, 463–917) and 3 and 4 (TD3–4, 692–1113).

All of these constructs were cloned, expressed, and purified as described above for the TD1–4 construct. TD2, TD3, apo-TD3, and TD4 were concentrated to 15 mg/mL in a buffer containing 20 mM Tris-HCl (pH 7.6), 150 mM NaCl, and 1 mM DTT for crystallization trials.

Isothermal titration calorimetry (ITC) measurements

ITC measurements were performed in duplicates at 25°C, using an ITC200 Micro-calorimeter (MicroCal, Inc). Experiments included 26 injections of 1.5 μ L of peptide solution (3–10 mM) into the sample cell containing 60–90 μ M of proteins, in 25 mM Tris-HCl (pH 7.6), 150 mM NaCl, and 5 mM β -mercaptoethanol buffer. All peptides (Peptide Specialty Laboratories GmbH) used for ITC experiments were dissolved or dialyzed into the same buffer as the protein and, if necessary, the pH adjusted to correspond with that of the protein solution. Phenylalanine-containing peptide concentrations were estimated with absorbance spectroscopy using the extinction molar coefficient at 257 nm (195 M⁻¹ cm⁻¹). Otherwise, peptide concentrations were estimated from the mass. Control experiments were performed under identical conditions to determine the heat signals that arise from injecting the peptide into the buffer. Binding isotherms were fit by nonlinear regression using Origin Software version 7.0 (MicroCal, Inc). The initial data point was routinely deleted. The ITC data were fit to a one-site binding model, or in the case of tandem Tudor domains, to the sequential binding site model, using software provided by MicroCal (Wiseman et al. 1989; Turnbull and Daranas 2003).

Crystallization trials

Extensive crystallization trials were performed of the various single- and multiple-domain constructs with and without the three MILI sDMA-containing peptides using the EMBL High Throughput Crystallization Facility. Typically, 576 conditions were screened for each protein in 96-well sitting-drop vapor diffusion format. Only apo-TD3 and TD3 bound to the R45me2 peptide ultimately gave good quality crystals.

TD3–R45me2 co-crystallization and structure determination

The TD3 (692–892) complex with R45me2 peptide was prepared by mixing protein and peptide in a 1:5 molar ratio, and co-crystals grew at 20°C in 0.2 M ammonium acetate, 0.05 M sodium cacodylate (pH 6.5), 30% PEG8000, 0.01 M magnesium acetate tetrahydrate. Crystals of the SeMet-labeled protein in complex with the peptide were grown at 20°C in slightly different conditions, 0.2 M NaCl, 0.1 M Bis-Tris (pH 7.0), and 29% PEG 3350.

Crystals belong to the space group of *P1* with two complexes per asymmetric unit. A complete native data set was collected to a resolution of 2.1 Å on beamline ID14-EH4 at the ESRF

(Grenoble) and processed with XDS (Kabsch 1993). The structure was solved by a SeMet MAD experiment. Data sets with a resolution of 2.7–2.8 Å were collected at the inflection point and peak wavelengths of the Se K-edge. Six SeMet sites were identified using SHARP (de La Fortelle 1997) and phases improved using RESOLVE. The initial model was built manually using COOT (Emsley and Cowtan 2004) and finally automatically built by ARP-wARP (Perrakis et al. 1999). The structure was refined using Refmac5 (with TLS refinement) (Murshudov et al. 1997) to a final R-factor of 21.9% and R_{free} of 28.4%, with all residues in allowed (97.7% in favored) regions of the Ramachandran plot, as analyzed by MOLPROBITY (Davis et al. 2004).

Apo-TD3 crystallization and structure determination

Unliganded TD3 (692–917) crystals were grown at 20°C in 10% PEG 1000 and 10% PEG 8000. They belong to space group $P4_12_12$ and diffract to 2.8 Å resolution. A complete data set was collected on beamline ID14–EH4, processed with XDS, and the structure solved by molecular replacement with PHASER (McCoy et al. 2005) using the TD3/R45me2 complex structure as a search model. The structure was refined with Refmac5 (including TLS) to a final R-factor of 22.7% and R_{free} of 25.5%. A total of 99.8% (95.8%) of residues are in allowed (favored) regions of the Ramachandran plot according to MOLPROBITY.

NMR spectroscopy

^{15}N -labeled samples of TD2 and TD3 were prepared by growing cells in minimal medium supplemented with $^{15}\text{NH}_4\text{Cl}$ (1 g liter $^{-1}$) as the sole nitrogen source and expressing and purifying as described above. In the final step of size exclusion chromatography, protein samples were exchanged into a buffer containing 20 mM sodium phosphate (pH 6.5), 50 mM NaCl, and 1 mM DTT.

^1H , ^{15}N HSQC titrations of ^{15}N -labeled TD2 or TD3 with successive addition of unlabeled ligands were performed on samples containing 200 μM protein to a final concentration ratio of 1:30 or 1:100 excess of the ligand (20 mM sodium phosphate at pH 6.5, 50 mM NaCl, and 7% (v/v) $^2\text{H}_2\text{O}$ added for the lock). Spectra were recorded at 298 K using an AV900 Bruker NMR spectrometer equipped with cryogenic triple resonance gradient probes.

SAXS data collection and analysis

Multidomain fragments containing the four Tudor domains (TD1–4) and Tudor domains 1 and 2 (TD1–2) were expressed and purified as described above. In the last purification step, Superdex 200 size exclusion column was used and 5 mM DTT was added before SAXS measurement. Data were collected at the ESRF BioSAXS station, ID14–EH3, which operates at a fixed energy (13.32 keV, $\lambda = 0.931$ Å) (Pernot et al. 2010). A total of 30 μL of protein solution was loaded into a 2-mm quartz capillary mounted in vacuum using an automated robotic system. The system enables the sample to flow through the beam during exposure to minimize the effect of radiation damage. A range of protein concentrations (10, 4.5, and 1.5 mg/mL) were measured to assess and account for interparticle effects. Two-dimensional scattering images were collected using a Vantec 2000 detector (Bruker) placed 1.83 m from the sample. For each sample, 10 frames of 30-sec duration were collected. Individual frames were processed automatically and independently using the software BsxCUBE, yielding individual

radially averaged curves of normalized intensity versus scattering angle $s = 4\pi\text{SIN}\theta/\lambda$. Time frames are combined, excluding any data points affected by aggregation due to radiation damage, to give the average scattering curve for each measurement. The average scattering from the buffer alone, measured before and after each sample measurement, was used for background subtraction using the program PRIMUS (Konarev et al. 2003). For both the TD1–4 and TD1–2 data, 40 ab initio models were created with DAMMIF (Franke and Svergun 2009), which were averaged and filtered using DAMAVER (Volkov and Svergun 2003). Rigid body models were produced with CORAL (Petoukhov and Svergun 2005), which allows the addition of missing linkers using a database of random self-avoiding chains. The plots of the 1D fits were produced with SAXSview (saxsview.sourceforge.net) and the figures of the 3D models were produced with PYMOL.

DATA DEPOSITION

Atomic coordinates and structure factors for the TD3-R45me2 peptide complex and apo-TD3 have been deposited in the wwPDB with codes 4B9W and 4B9X, respectively.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank the ESRF–EMBL Joint Structural Biology Group for access to ESRF beamlines and the EMBL–ESRF–ILL–IBS Partnership for Structural Biology for access to structural biology instrumentation, notably the high-throughput crystallization platform. Fruitful discussions with Jordi Xiol and Michael Reuter are also acknowledged.

Received May 4, 2012; accepted July 31, 2012.

REFERENCES

- Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, et al. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**: 203–207.
- Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ. 2008. A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. *Mol Cell* **31**: 785–799.
- Arkov AL, Ramos A. 2010. Building RNA-protein granules: Insight from the germline. *Trends Cell Biol* **20**: 482–490.
- Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**: 1089–1103.
- Carmell MA, Xuan Z, Zhang MQ, Hannon GJ. 2002. The Argonaute family: Tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev* **16**: 2733–2742.
- Chen C, Nott TJ, Jin J, Pawson T. 2011. Deciphering arginine methylation: Tudor tells the tale. *Nat Rev Mol Cell Biol* **12**: 629–642.
- Chuma S, Hiyoshi M, Yamamoto A, Hosokawa M, Takamune K, Nakatsuji N. 2003. Mouse Tudor Repeat-1 (MTR-1) is a novel

- component of chromatoid bodies/nuages in male germ cells and forms a complex with snRNPs. *Mech Dev* **120**: 979–990.
- Davis IW, Murray LW, Richardson JS, Richardson DC. 2004. MOLPROBITY: Structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* **32**: W615–W619.
- De Fazio S, Bartonicek N, Di Giacomo M, Abreu-Goodger C, Sankar A, Funaya C, Antony C, Moreira PN, Enright AJ, O'Carroll D. 2011. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* **480**: 259–263.
- de La Fortelle E, Bricogne G. 1997. Heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In *Methods in Enzymology*, Vol 276 (ed. CW Carter Jr, RM Sweet), pp. 472–494. Academic Press, Inc., San Diego, CA.
- Emsley P, Cowtan K. 2004. Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126–2132.
- Filipowicz W, Bhattacharyya SN, Sonenberg N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight? *Nat Rev Genet* **9**: 102–114.
- Franke D, Svergun DI. 2009. DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J Appl Crystallogr* **42**: 342–346.
- Ghildiyal M, Zamore PD. 2009. Small silencing RNAs: An expanding universe. *Nat Rev Genet* **10**: 94–108.
- Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**: 199–202.
- Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**: 1587–1590.
- Handler D, Olivieri D, Novatchkova M, Gruber FS, Meixner K, Mechtler K, Stark A, Sachidanandam R, Brennecke J. 2011. A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J* **30**: 3977–3993.
- Houwing S, Kamminga LM, Berezikov E, Cronenbold D, Girard A, van den Elst H, Filipov DV, Blaser H, Raz E, Moens CB, et al. 2007. A role for piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. *Cell* **129**: 69–82.
- Huang H, Gao Q, Peng X, Choi SY, Sarma K, Ren H, Morris AJ, Frohman MA. 2011. piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. *Dev Cell* **20**: 376–387.
- Kabsch W. 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr* **26**: 795–800.
- Kim J, Daniel J, Espejo A, Lake A, Krishna M, Xia L, Zhang Y, Bedford MT. 2006. Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep* **7**: 397–403.
- Kirino Y, Kim N, de Planell-Saguer M, Khandros E, Chiorean S, Klein PS, Rigoutsos I, Jongens TA, Mourelatos Z. 2009. Arginine methylation of Piwi proteins, catalyzed by dPRMT5, is required for Ago3 and Aub stability. *Nat Cell Biol* **11**: 652–658.
- Kirino Y, Vourekas A, Kim N, de Lima Alves F, Rappsilber J, Klein PS, Jongens TA, Mourelatos Z. 2010. Arginine methylation of Vasa protein is conserved across phyla. *J Biol Chem* **285**: 8148–8154.
- Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. 2003. PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* **36**: 1277–1282.
- Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, et al. 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22**: 908–917.
- Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Takamatsu K, Chuma S, Kojima-Kita K, Shiromoto Y, Asada N, Toyoda A, Fujiyama A, et al. 2010. MVH in piRNA processing and gene silencing of retrotransposons. *Genes Dev* **24**: 887–892.
- Leroy P, Alzari P, Sassoon D, Wolgemuth D, Fellous M. 1989. The protein encoded by a murine male germ cell-specific transcript is a putative ATP-dependent RNA helicase. *Cell* **57**: 549–559.
- Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM, et al. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137**: 509–521.
- Liu H, Wang JY, Huang Y, Li Z, Gong W, Lehmann R, Xu RM. 2010a. Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes Dev* **24**: 1876–1881.
- Liu K, Chen C, Guo Y, Lam R, Bian C, Xu C, Zhao DY, Jin J, MacKenzie F, Pawson T, et al. 2010b. Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. *Proc Natl Acad Sci* **107**: 18398–18403.
- Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* **137**: 522–535.
- Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP. 2003. The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem Sci* **28**: 69–74.
- McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. 2005. Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **61**: 458–464.
- Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240–255.
- Nishida KM, Okada TN, Kawamura T, Mituyama T, Kawamura Y, Inagaki S, Huang H, Chen D, Kodama T, Siomi H, et al. 2009. Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in *Drosophila* germlines. *EMBO J* **28**: 3820–3831.
- Patil VS, Kai T. 2010. Repression of retroelements in *Drosophila* germline via piRNA pathway by the Tudor domain protein Tejas. *Curr Biol* **20**: 724–730.
- Pernot P, Theveneau P, Giraud T, Fernandes R, Nurizzo D, Spruce D, Surr J, McSweeney S, Round A, Felisaz F, et al. 2010. New beamline dedicated to solution scattering from biological macromolecules at the ESRF. *J. Physics, Conference Series* **247**: doi: 10.1088/1742-6596/247/1/012009.
- Perrakis A, Morris R, Lamzin VS. 1999. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6**: 458–463.
- Petoukhov MV, Svergun DI. 2005. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* **89**: 1237–1250.
- Reuter M, Chuma S, Tanaka T, Franz T, Stark A, Pillai RS. 2009. Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nat Struct Mol Biol* **16**: 639–646.
- Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC. 2010. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes Dev* **24**: 2493–2498.
- Sashital DG, Doudna JA. 2010. Structural insights into RNA interference. *Curr Opin Struct Biol* **20**: 90–97.
- Selenko P, Sprangers R, Stier G, Buhler D, Fischer U, Sattler M. 2001. SMN Tudor domain structure and its interaction with the Sm proteins. *Nat Struct Biol* **8**: 27–31.
- Senti KA, Brennecke J. 2010. The piRNA pathway: A fly's perspective on the guardian of the genome. *Trends Genet* **26**: 499–509.
- Siomi MC, Sato K, Pezic D, Aravin AA. 2011. PIWI-interacting small RNAs: The vanguard of genome defence. *Nat Rev Mol Cell Biol* **12**: 246–258.

- Song JJ, Liu J, Tolia NH, Schneiderman J, Smith SK, Martienssen RA, Hannon GJ, Joshua-Tor L. 2003. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Biol* **10**: 1026–1032.
- Tanaka SS, Toyooka Y, Akasu R, Katoh-Fukui Y, Nakahara Y, Suzuki R, Yokoyama M, Noce T. 2000. The mouse homolog of *Drosophila Vasa* is required for the development of male germ cells. *Genes Dev* **14**: 841–853.
- Tripsianes K, Madl T, Machyna M, Fessas D, Englbrecht C, Fischer U, Neugebauer KM, Sattler M. 2011. Structural basis for dimethyl-arginine recognition by the Tudor domains of human SMN and SPF30 proteins. *Nat Struct Mol Biol* **18**: 1414–1420.
- Turnbull WB, Daranas AH. 2003. On the value of *c*: Can low affinity systems be studied by isothermal titration calorimetry? *J Am Chem Soc* **125**: 14859–14866.
- Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**: 320–324.
- Vagin VV, Wohlschlegel J, Qu J, Jonsson Z, Huang X, Chuma S, Girard A, Sachidanandam R, Hannon GJ, Aravin AA. 2009. Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev* **23**: 1749–1762.
- Volkov VV, Svergun DI. 2003. Uniqueness of *ab initio* shape determination in small-angle scattering. *J Appl Crystallogr* **36**: 860–864.
- Vourekas A, Kirino Y, Mourelatos Z. 2010. Elective affinities: A Tudor-Aubergine tale of germline partnership. *Genes Dev* **24**: 1963–1966.
- Wang J, Saxe JP, Tanaka T, Chuma S, Lin H. 2009a. Mili interacts with Tudor Domain Containing Protein 1 in regulating spermatogenesis. *Curr Biol* **19**: 640–644.
- Wang Y, Juranek S, Li H, Sheng G, Wardle GS, Tuschl T, Patel DJ. 2009b. Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461**: 754–761.
- Wiseman T, Williston S, Brandts JF, Lin LN. 1989. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* **179**: 131–137.

References

- Adams-Cioaba, M. A., Y. Guo, et al. (2010). "Structural studies of the tandem Tudor domains of fragile X mental retardation related proteins FXR1 and FXR2." PLoS One **5**(11): e13559.
- Ahearn, I. M., F. D. Tsai, et al. (2011). "FKBP12 binds to acylated H-ras and promotes depalmitoylation." Mol Cell **41**(2): 173-85.
- al-Mukhtar, K. A. and A. C. Webb (1971). "An ultrastructural study of primordial germ cells, oogonia and early oocytes in *Xenopus laevis*." J Embryol Exp Morphol **26**(2): 195-217.
- Anne, J. and B. M. Mechler (2005). "Valois, a component of the nuage and pole plasm, is involved in assembly of these structures, and binds to Tudor and the methyltransferase Capsuleen." Development **132**(9): 2167-77.
- Ansieau, S. and A. Leutz (2002). "The conserved Mynd domain of BS69 binds cellular and oncoviral proteins through a common PXLXP motif." J Biol Chem **277**(7): 4906-10.
- Aravin, A., D. Gaidatzis, et al. (2006). "A novel class of small RNAs bind to MILI protein in mouse testes." Nature **442**(7099): 203-7.
- Aravin, A. A. and D. Bourc'his (2008). "Small RNA guides for de novo DNA methylation in mammalian germ cells." Genes Dev **22**(8): 970-5.
- Aravin, A. A., G. J. Hannon, et al. (2007). "The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race." Science **318**(5851): 761-4.
- Aravin, A. A., M. Lagos-Quintana, et al. (2003). "The small RNA profile during *Drosophila melanogaster* development." Dev Cell **5**(2): 337-50.
- Aravin, A. A., N. M. Naumova, et al. (2001). "Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline." Curr Biol **11**(13): 1017-27.
- Aravin, A. A., R. Sachidanandam, et al. (2008). "A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice." Mol Cell **31**(6): 785-99.
- Aravin, A. A., R. Sachidanandam, et al. (2007). "Developmentally regulated piRNA clusters implicate MILI in transposon control." Science **316**(5825): 744-7.

- Aravin, A. A., G. W. van der Heijden, et al. (2009). "Cytoplasmic compartmentalization of the fetal piRNA pathway in mice." *PLoS Genet* **5**(12): e1000764.
- Arkov, A. L., J. Y. Wang, et al. (2006). "The role of Tudor domains in germline development and polar granule architecture." *Development* **133**(20): 4053-62.
- Aza-Blanc, P., C. L. Cooper, et al. (2003). "Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening." *Mol Cell* **12**(3): 627-37.
- Babiarz, J. E., J. G. Ruby, et al. (2008). "Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs." *Genes Dev* **22**(20): 2773-85.
- Baek, D., J. Villen, et al. (2008). "The impact of microRNAs on protein output." *Nature* **455**(7209): 64-71.
- Ball, L. J., N. V. Murzina, et al. (1997). "Structure of the chromatin binding (chromo) domain from mouse modifier protein 1." *Embo J* **16**(9): 2473-81.
- Bartel, D. P. (2004). "MicroRNAs: genomics, biogenesis, mechanism, and function." *Cell* **116**(2): 281-97.
- Bartel, D. P. (2009). "MicroRNAs: target recognition and regulatory functions." *Cell* **136**(2): 215-33.
- Bedford, M. T. and S. G. Clarke (2009). "Protein arginine methylation in mammals: who, what, and why." *Mol Cell* **33**(1): 1-13.
- Behm-Ansmant, I., J. Rehwinkel, et al. (2006). "mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes." *Genes Dev* **20**(14): 1885-98.
- Bernstein, E., A. A. Caudy, et al. (2001). "Role for a bidentate ribonuclease in the initiation step of RNA interference." *Nature* **409**(6818): 363-6.
- Bohnsack, M. T., K. Czaplinski, et al. (2004). "Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs." *Rna* **10**(2): 185-91.
- Boswell, R. E. and A. P. Mahowald (1985). "tudor, a gene required for assembly of the germ plasm in *Drosophila melanogaster*." *Cell* **43**(1): 97-104.

- Botuyan, M. V., J. Lee, et al. (2006). "Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair." *Cell* **127**(7): 1361-73.
- Brahms, H., L. Meheus, et al. (2001). "Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein." *Rna* **7**(11): 1531-42.
- Brahms, H., J. Raymackers, et al. (2000). "The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies." *J Biol Chem* **275**(22): 17122-9.
- Brennecke, J., A. A. Aravin, et al. (2007). "Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*." *Cell* **128**(6): 1089-103.
- Brennecke, J., C. D. Malone, et al. (2008). "An epigenetic role for maternally inherited piRNAs in transposon silencing." *Science* **322**(5906): 1387-92.
- Brennecke, J., A. Stark, et al. (2005). "Principles of microRNA-target recognition." *PLoS Biol* **3**(3): e85.
- Brodersen, P., L. Sakvarelidze-Achard, et al. (2008). "Widespread translational inhibition by plant miRNAs and siRNAs." *Science* **320**(5880): 1185-90.
- Brower-Toland, B., S. D. Findley, et al. (2007). "*Drosophila* PIWI associates with chromatin and interacts directly with HP1a." *Genes Dev* **21**(18): 2300-11.
- Brunger, A. T. (1993). "Assessment of phase accuracy by cross validation: the free R value. Methods and applications." *Acta Crystallogr D Biol Crystallogr* **49**(Pt 1): 24-36.
- Buhler, D., V. Raker, et al. (1999). "Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy." *Hum Mol Genet* **8**(13): 2351-7.
- Cai, X., C. H. Hagedorn, et al. (2004). "Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs." *Rna* **10**(12): 1957-66.
- Carmell, M. A., A. Girard, et al. (2007). "MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline." *Dev Cell* **12**(4): 503-14.
- Carthew, R. W. and E. J. Sontheimer (2009). "Origins and Mechanisms of miRNAs and siRNAs." *Cell* **136**(4): 642-55.

- Chan, S. W. (2008). "Inputs and outputs for chromatin-targeted RNAi." Trends Plant Sci **13**(7): 383-9.
- Chang, B., Y. Chen, et al. (2007). "JMJD6 is a histone arginine demethylase." Science **318**(5849): 444-7.
- Charier, G., J. Couprie, et al. (2004). "The Tudor tandem of 53BP1: a new structural motif involved in DNA and RG-rich peptide binding." Structure **12**(9): 1551-62.
- Chekulaeva, M. and W. Filipowicz (2009). "Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells." Curr Opin Cell Biol **21**(3): 452-60.
- Chen, C., J. Jin, et al. (2009). "Mouse Piwi interactome identifies binding mechanism of Tdrkh Tudor domain to arginine methylated Miwi." Proc Natl Acad Sci U S A **106**(48): 20336-41.
- Chen, C., T. J. Nott, et al. (2011). "Deciphering arginine methylation: Tudor tells the tale." Nat Rev Mol Cell Biol **12**(10): 629-42.
- Chen, V. B., W. B. Arendall, 3rd, et al. (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 12-21.
- Chendrimada, T. P., R. I. Gregory, et al. (2005). "TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing." Nature **436**(7051): 740-4.
- Cheng, D., J. Cote, et al. (2007). "The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing." Mol Cell **25**(1): 71-83.
- Choi, S. Y., P. Huang, et al. (2006). "A common lipid links Mfn-mediated mitochondrial fusion and SNARE-regulated exocytosis." Nat Cell Biol **8**(11): 1255-62.
- Chuma, S., M. Hiyoshi, et al. (2003). "Mouse Tudor Repeat-1 (MTR-1) is a novel component of chromatoid bodies/nuages in male germ cells and forms a complex with snRNPs." Mech Dev **120**(9): 979-90.
- Chuma, S., M. Hosokawa, et al. (2006). "Tdrd1/Mtr-1, a tudor-related gene, is essential for male germ-cell differentiation and nuage/germinal granule formation in mice." Proc Natl Acad Sci U S A **103**(43): 15894-9.
- Chung, W. J., K. Okamura, et al. (2008). "Endogenous RNA interference provides a somatic defense against Drosophila transposons." Curr Biol **18**(11): 795-802.

- Cogoni, C. and G. Macino (1999). "Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase." Nature **399**(6732): 166-9.
- Combet, C., C. Blanchet, et al. (2000). "NPS@: Network Protein Sequence Analysis." Trends in Biochemical Sciences **25**(3): 147-150.
- Cote, J. and S. Richard (2005). "Tudor domains bind symmetrical dimethylated arginines." J Biol Chem **280**(31): 28476-83.
- Cox, D. N., A. Chao, et al. (1998). "A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal." Genes Dev **12**(23): 3715-27.
- Cox, D. N., A. Chao, et al. (2000). "piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells." Development **127**(3): 503-14.
- Crackower, M. A., N. K. Kolas, et al. (2003). "Essential role of Fkbp6 in male fertility and homologous chromosome pairing in meiosis." Science **300**(5623): 1291-5.
- Creed, T. M., S. N. Loganathan, et al. (2010). "Novel role of specific Tudor domains in Tudor-Aubergine protein complex assembly and distribution during *Drosophila* oogenesis." Biochem Biophys Res Commun **402**(2): 384-9.
- Czech, B., C. D. Malone, et al. (2008). "An endogenous small interfering RNA pathway in *Drosophila*." Nature **453**(7196): 798-802.
- Davis, E., F. Caiment, et al. (2005). "RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus." Curr Biol **15**(8): 743-9.
- De Fazio, S., N. Bartonicek, et al. (2011). "The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements." Nature **480**(7376): 259-63.
- de La Fortelle, E., G. r. Bricogne, et al. (1997). [27] Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods in Enzymology, Academic Press. **Volume 276**: 472-494.
- Deng, W. and H. Lin (2002). "miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis." Dev Cell **2**(6): 819-30.
- Denli, A. M., B. B. Tops, et al. (2004). "Processing of primary microRNAs by the Microprocessor complex." Nature **432**(7014): 231-5.

- Denny, W. B., V. Prapapanich, et al. (2005). "Structure-function analysis of squirrel monkey FK506-binding protein 51, a potent inhibitor of glucocorticoid receptor activity." Endocrinology **146**(7): 3194-201.
- Desset, S., N. Buchon, et al. (2008). "In *Drosophila melanogaster* the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways." PLoS One **3**(2): e1526.
- Djupedal, I., M. Portoso, et al. (2005). "RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing." Genes Dev **19**(19): 2301-6.
- Du, T. and P. D. Zamore (2005). "microPrimer: the biogenesis and function of microRNA." Development **132**(21): 4645-52.
- Eddy, E. M. (1974). "Fine structural observations on the form and distribution of nuage in germ cells of the rat." Anat Rec **178**(4): 731-57.
- Eissenberg, J. C. (2001). "Molecular biology of the chromo domain: an ancient chromatin module comes of age." Gene **275**(1): 19-29.
- Elbashir, S. M., J. Harborth, et al. (2001). "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells." Nature **411**(6836): 494-8.
- Elbashir, S. M., W. Lendeckel, et al. (2001). "RNA interference is mediated by 21- and 22-nucleotide RNAs." Genes Dev **15**(2): 188-200.
- Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-32.
- Eulalio, A., I. Behm-Ansmant, et al. (2007). "P-body formation is a consequence, not the cause, of RNA-mediated gene silencing." Mol Cell Biol **27**(11): 3970-81.
- Eulalio, A., F. Triteschler, et al. (2009). "The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing." Rna **15**(8): 1433-42.
- Fabian, M. R., N. Sonenberg, et al. (2010). "Regulation of mRNA translation and stability by microRNAs." Annu Rev Biochem **79**: 351-79.
- Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*." Nature **391**(6669): 806-11.

- Fischer, G., B. Wittmann-Liebold, et al. (1989). "Cyclophilin and peptidyl-prolyl cis-trans isomerase are probably identical proteins." *Nature* **337**(6206): 476-8.
- Forstemann, K., M. D. Horwich, et al. (2007). "Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1." *Cell* **130**(2): 287-97.
- Frank, F., N. Sonenberg, et al. (2010). "Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2." *Nature* **465**(7299): 818-22.
- Franke, D. and D. I. Svergun (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography*. **42**: 342-346.
- Franks, T. M. and J. Lykke-Andersen (2008). "The control of mRNA decapping and P-body formation." *Mol Cell* **32**(5): 605-15.
- Friberg, A., L. Corsini, et al. (2009). "Structure and ligand binding of the extended Tudor domain of D. melanogaster Tudor-SN." *J Mol Biol* **387**(4): 921-34.
- Friesen, W. J. and G. Dreyfuss (2000). "Specific sequences of the Sm and Sm-like (Lsm) proteins mediate their interaction with the spinal muscular atrophy disease gene product (SMN)." *J Biol Chem* **275**(34): 26370-5.
- Friesen, W. J., S. Massenet, et al. (2001). "SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets." *Mol Cell* **7**(5): 1111-7.
- Frost, R. J., F. K. Hamra, et al. (2010). "MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs." *Proc Natl Acad Sci U S A* **107**(26): 11847-52.
- Fukuda, T., C. Hedinger, et al. (1975). "Ultrastructure of developing germ cells in the fetal human testis." *Cell Tissue Res* **161**(1): 55-70.
- Gangaraju, V. K., H. Yin, et al. (2011). "Drosophila Piwi functions in Hsp90-mediated suppression of phenotypic variation." *Nat Genet* **43**(2): 153-8.
- Ghildiyal, M., H. Seitz, et al. (2008). "Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells." *Science* **320**(5879): 1077-81.

- Ghildiyal, M. and P. D. Zamore (2009). "Small silencing RNAs: an expanding universe." Nat Rev Genet **10**(2): 94-108.
- Giot, L., J. S. Bader, et al. (2003). "A protein interaction map of Drosophila melanogaster." Science **302**(5651): 1727-36.
- Girard, A., R. Sachidanandam, et al. (2006). "A germline-specific class of small RNAs binds mammalian Piwi proteins." Nature **442**(7099): 199-202.
- Gouet, P., E. Courcelle, et al. (1999). "ESPrict: analysis of multiple sequence alignments in PostScript." Bioinformatics **15**(4): 305-8.
- Goulet, I., S. Boisvenue, et al. (2008). "TDRD3, a novel Tudor domain-containing protein, localizes to cytoplasmic stress granules." Hum Mol Genet **17**(19): 3055-74.
- Gray, S. G., A. H. Iglesias, et al. (2005). "Functional characterization of JMJD2A, a histone deacetylase- and retinoblastoma-binding protein." J Biol Chem **280**(31): 28507-18.
- Gregory, R. I., K. P. Yan, et al. (2004). "The Microprocessor complex mediates the genesis of microRNAs." Nature **432**(7014): 235-40.
- Grimson, A., K. K. Farh, et al. (2007). "MicroRNA targeting specificity in mammals: determinants beyond seed pairing." Mol Cell **27**(1): 91-105.
- Grimson, A., M. Srivastava, et al. (2008). "Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals." Nature **455**(7217): 1193-7.
- Grishok, A., A. E. Pasquinelli, et al. (2001). "Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing." Cell **106**(1): 23-34.
- Grivna, S. T., B. Pyhtila, et al. (2006). "MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis." Proc Natl Acad Sci U S A **103**(36): 13415-20.
- Gross, C. T. and W. McGinnis (1996). "DEAF-1, a novel protein that binds an essential region in a Deformed response element." Embo J **15**(8): 1961-70.
- Gunawardane, L. S., K. Saito, et al. (2007). "A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila." Science **315**(5818): 1587-90.

- Haase, A. D., S. Fenoglio, et al. (2010). "Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila." Genes Dev **24**(22): 2499-504.
- Haase, A. D., L. Jaskiewicz, et al. (2005). "TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing." EMBO Rep **6**(10): 961-7.
- Hammond, S. M., E. Bernstein, et al. (2000). "An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells." Nature **404**(6775): 293-6.
- Han, J., Y. Lee, et al. (2004). "The Drosha-DGCR8 complex in primary microRNA processing." Genes Dev **18**(24): 3016-27.
- Harding, M. W., A. Galat, et al. (1989). "A receptor for the immunosuppressant FK506 is a cis-trans peptidyl-prolyl isomerase." Nature **341**(6244): 758-60.
- Hay, B., L. Y. Jan, et al. (1988). "A protein component of Drosophila polar granules is encoded by vasa and has extensive sequence similarity to ATP-dependent helicases." Cell **55**(4): 577-87.
- Ho, S. N., H. D. Hunt, et al. (1989). "Site-directed mutagenesis by overlap extension using the polymerase chain reaction." Gene **77**(1): 51-9.
- Holdermann, I., N. H. Meyer, et al. (2012). "Chromodomains read the arginine code of post-translational targeting." Nat Struct Mol Biol **19**(2): 260-3.
- Horwich, M. D., C. Li, et al. (2007). "The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC." Curr Biol **17**(14): 1265-72.
- Hosokawa, M., M. Shoji, et al. (2007). "Tudor-related proteins TDRD1/MTR-1, TDRD6 and TDRD7/TRAP: domain composition, intracellular localization, and function in male germ cells in mice." Dev Biol **301**(1): 38-52.
- Houwing, S., E. Berezikov, et al. (2008). "Zili is required for germ cell differentiation and meiosis in zebrafish." Embo J **27**(20): 2702-11.
- Houwing, S., L. M. Kamminga, et al. (2007). "A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish." Cell **129**(1): 69-82.
- Huang, H., Q. Gao, et al. (2011). "piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling." Dev Cell **20**(3): 376-87.

- Huang, H. Y., S. Houwing, et al. (2011). "Tdrd1 acts as a molecular scaffold for Piwi proteins and piRNA targets in zebrafish." Embo J **30**(16): 3298-308.
- Huang, Y., J. Fang, et al. (2006). "Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A." Science **312**(5774): 748-51.
- Huang, Y., L. Ji, et al. (2009). "Structural insights into mechanisms of the small RNA methyltransferase HEN1." Nature **461**(7265): 823-7.
- Huggenvik, J. I., R. J. Michelson, et al. (1998). "Characterization of a nuclear deformed epidermal autoregulatory factor-1 (DEAF-1)-related (NUDR) transcriptional regulator protein." Mol Endocrinol **12**(10): 1619-39.
- Hughes-Davies, L., D. Huntsman, et al. (2003). "EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer." Cell **115**(5): 523-35.
- Hutvagner, G., J. McLachlan, et al. (2001). "A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA." Science **293**(5531): 834-8.
- Hutvagner, G. and M. J. Simard (2008). "Argonaute proteins: key players in RNA silencing." Nat Rev Mol Cell Biol **9**(1): 22-32.
- Hutvagner, G. and P. D. Zamore (2002). "A microRNA in a multiple-turnover RNAi enzyme complex." Science **297**(5589): 2056-60.
- Iki, T., M. Yoshikawa, et al. (2010). "In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90." Mol Cell **39**(2): 282-91.
- Ikura, T. and N. Ito (2007). "Requirements for peptidyl-prolyl isomerization activity: a comprehensive mutational analysis of the substrate-binding cavity of FK506-binding protein 12." Protein Sci **16**(12): 2618-25.
- Iwasaki, S., M. Kobayashi, et al. (2010). "Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes." Mol Cell **39**(2): 292-9.
- Jacobs, S. A. and S. Khorasanizadeh (2002). "Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail." Science **295**(5562): 2080-3.
- Jiang, F., X. Ye, et al. (2005). "Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila." Genes Dev **19**(14): 1674-9.

- Jinek, M. and J. A. Doudna (2009). "A three-dimensional view of the molecular machinery of RNA interference." Nature **457**(7228): 405-12.
- Jinek, M., M. R. Fabian, et al. (2010). "Structural insights into the human GW182-PABC interaction in microRNA-mediated deadenylation." Nat Struct Mol Biol **17**(2): 238-40.
- Johnston, M., M. C. Geoffroy, et al. (2010). "HSP90 protein stabilizes unloaded argonaute complexes and microscopic P-bodies in human cells." Mol Biol Cell **21**(9): 1462-9.
- Kabsch, W. (1993). "Xds." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-32.
- Kamminga, L. M., M. J. Luteijn, et al. (2010). "Hen1 is required for oocyte development and piRNA stability in zebrafish." Embo J **29**(21): 3688-700.
- Kamphausen, T., J. Fanghanel, et al. (2002). "Characterization of Arabidopsis thaliana AtFKBP42 that is membrane-bound and interacts with Hsp90." Plant J **32**(3): 263-76.
- Kaneda, M., M. Okano, et al. (2004). "Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting." Nature **429**(6994): 900-3.
- Kang, C. B., H. Ye, et al. (2008). "Solution structure of FK506 binding domain (FKBD) of Plasmodium falciparum FK506 binding protein 35 (PffFKBP35)." Proteins **70**(1): 300-2.
- Kato, Y., M. Kaneda, et al. (2007). "Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse." Hum Mol Genet **16**(19): 2272-80.
- Kawamura, Y., K. Saito, et al. (2008). "Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells." Nature **453**(7196): 793-7.
- Kawaoka, S., N. Izumi, et al. (2011). "3' end formation of PIWI-interacting RNAs in vitro." Mol Cell **43**(6): 1015-22.
- Kennerdell, J. R. and R. W. Carthew (1998). "Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway." Cell **95**(7): 1017-26.
- Ketting, R. F., S. E. Fischer, et al. (2001). "Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans." Genes Dev **15**(20): 2654-9.
- Khvorova, A., A. Lescoute, et al. (2003). "Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity." Nat Struct Biol **10**(9): 708-12.

- Kidwell, M. G. and H. M. Sang (1986). "Hybrid dysgenesis in *Drosophila melanogaster*: synthesis of RP strains by chromosomal contamination." Genet Res **47**(3): 181-5.
- Kim, D. H., L. M. Villeneuve, et al. (2006). "Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells." Nat Struct Mol Biol **13**(9): 793-7.
- Kim, V. N. (2005). "MicroRNA biogenesis: coordinated cropping and dicing." Nat Rev Mol Cell Biol **6**(5): 376-85.
- Kirino, Y., N. Kim, et al. (2009). "Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability." Nat Cell Biol **11**(5): 652-8.
- Kirino, Y. and Z. Mourelatos (2007). "2'-O-methyl modification in mouse piRNAs and its methylase." Nucleic Acids Symp Ser (Oxf)(51): 417-8.
- Kirino, Y., A. Vourekas, et al. (2010). "Arginine methylation of vasa protein is conserved across phyla." J Biol Chem **285**(11): 8148-54.
- Kirino, Y., A. Vourekas, et al. (2009). "Arginine methylation of Aubergine mediates Tudor binding and germ plasm localization." Rna **16**(1): 70-8.
- Knight, S. W. and B. L. Bass (2001). "A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*." Science **293**(5538): 2269-71.
- Kojima, K., S. Kuramochi-Miyagawa, et al. (2009). "Associations between PIWI proteins and TDRD1/MTR-1 are critical for integrated subcellular localization in murine male germ cells." Genes Cells **14**(10): 1155-65.
- Konarev, P. V., V. V. Volkov, et al. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography. **36**: 1277-1282.
- Kotaja, N., H. Lin, et al. (2006). "Interplay of PIWI/Argonaute protein MIWI and kinesin KIF17b in chromatoid bodies of male germ cells." J Cell Sci **119**(Pt 13): 2819-25.
- Kunzi, M. S. and P. M. Pitha (2005). "Interferon research: a brief history." Methods Mol Med **116**: 25-35.
- Kuramochi-Miyagawa, S., T. Kimura, et al. (2004). "Mili, a mammalian member of piwi family gene, is essential for spermatogenesis." Development **131**(4): 839-49.
- Kuramochi-Miyagawa, S., T. Kimura, et al. (2001). "Two mouse piwi-related genes: miwi and mili." Mech Dev **108**(1-2): 121-33.

- Kuramochi-Miyagawa, S., T. Watanabe, et al. (2010). "MVH in piRNA processing and gene silencing of retrotransposons." Genes Dev **24**(9): 887-92.
- Kuramochi-Miyagawa, S., T. Watanabe, et al. (2008). "DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes." Genes Dev **22**(7): 908-17.
- Kwak, P. B. and Y. Tomari (2012). "The N domain of Argonaute drives duplex unwinding during RISC assembly." Nat Struct Mol Biol **19**(2): 145-51.
- Lai, E. C., C. Wiel, et al. (2004). "Complementary miRNA pairs suggest a regulatory role for miRNA:miRNA duplexes." Rna **10**(2): 171-5.
- Landgraf, P., M. Rusu, et al. (2007). "A mammalian microRNA expression atlas based on small RNA library sequencing." Cell **129**(7): 1401-14.
- Landthaler, M., A. Yalcin, et al. (2004). "The human DiGeorge syndrome critical region gene 8 and its D. melanogaster homolog are required for miRNA biogenesis." Curr Biol **14**(23): 2162-7.
- Langer, G., S. X. Cohen, et al. (2008). "Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7." Nat Protoc **3**(7): 1171-9.
- Larkin, M. A., G. Blackshields, et al. (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-8.
- Lasko, P. F. and M. Ashburner (1988). "The product of the Drosophila gene vasa is very similar to eukaryotic initiation factor-4A." Nature **335**(6191): 611-7.
- Lau, N. C., N. Robine, et al. (2009). "Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line." Genome Res **19**(10): 1776-85.
- Lau, N. C., A. G. Seto, et al. (2006). "Characterization of the piRNA complex from rat testes." Science **313**(5785): 363-7.
- Lee, R. C., R. L. Feinbaum, et al. (1993). "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14." Cell **75**(5): 843-54.
- Lee, Y., C. Ahn, et al. (2003). "The nuclear RNase III Drosha initiates microRNA processing." Nature **425**(6956): 415-9.

- Lee, Y., I. Hur, et al. (2006). "The role of PACT in the RNA silencing pathway." Embo J **25**(3): 522-32.
- Lee, Y., K. Jeon, et al. (2002). "MicroRNA maturation: stepwise processing and subcellular localization." Embo J **21**(17): 4663-70.
- Lee, Y., M. Kim, et al. (2004). "MicroRNA genes are transcribed by RNA polymerase II." Embo J **23**(20): 4051-60.
- Lee, Y. S., K. Nakahara, et al. (2004). "Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways." Cell **117**(1): 69-81.
- Leroy, P., P. Alzari, et al. (1989). "The protein encoded by a murine male germ cell-specific transcript is a putative ATP-dependent RNA helicase." Cell **57**(4): 549-59.
- Leslie, A. G., H. R. Powell, et al. (2002). "Automation of the collection and processing of X-ray diffraction data -- a generic approach." Acta Crystallogr D Biol Crystallogr **58**(Pt 11): 1924-8.
- Leuschner, P. J., S. L. Ameres, et al. (2006). "Cleavage of the siRNA passenger strand during RISC assembly in human cells." EMBO Rep **7**(3): 314-20.
- Lewis, B. P., C. B. Burge, et al. (2005). "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets." Cell **120**(1): 15-20.
- Lewis, B. P., I. H. Shih, et al. (2003). "Prediction of mammalian microRNA targets." Cell **115**(7): 787-98.
- Li, C., V. V. Vagin, et al. (2009). "Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies." Cell **137**(3): 509-21.
- Li, M. A., J. D. Alls, et al. (2003). "The large Maf factor Traffic Jam controls gonad morphogenesis in Drosophila." Nat Cell Biol **5**(11): 994-1000.
- Lim, A. K. and T. Kai (2007). "Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster." Proc Natl Acad Sci U S A **104**(16): 6714-9.
- Lin, H. and A. C. Spradling (1997). "A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary." Development **124**(12): 2463-76.

- Linder, B., O. Plottner, et al. (2008). "Tdrd3 is a novel stress granule-associated protein interacting with the Fragile-X syndrome protein FMRP." Hum Mol Genet **17**(20): 3236-46.
- Lingel, A., B. Simon, et al. (2003). "Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain." Nature **426**(6965): 465-9.
- Lippman, Z. and R. Martienssen (2004). "The role of RNA interference in heterochromatic silencing." Nature **431**(7006): 364-70.
- Liu, H., J. Y. Wang, et al. (2010). "Structural basis for methylarginine-dependent recognition of Aubergine by Tudor." Genes Dev **24**(17): 1876-81.
- Liu, J., M. A. Carmell, et al. (2004). "Argonaute2 is the catalytic engine of mammalian RNAi." Science **305**(5689): 1437-41.
- Liu, K., C. Chen, et al. (2010). "Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain." Proc Natl Acad Sci U S A **107**(43): 18398-403.
- Liu, K., Y. Guo, et al. (2012). "Crystal structure of TDRD3 and methyl-arginine binding characterization of TDRD3, SMN and SPF30." PLoS One **7**(2): e30375.
- Liu, Q., T. A. Rand, et al. (2003). "R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway." Science **301**(5641): 1921-5.
- Liu, Y., W. Chen, et al. (2007). "Structural basis for recognition of SMRT/N-CoR by the MYND domain and its contribution to AML1/ETO's activity." Cancer Cell **11**(6): 483-97.
- Llave, C., Z. Xie, et al. (2002). "Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA." Science **297**(5589): 2053-6.
- Lu, K. P., G. Finn, et al. (2007). "Prolyl cis-trans isomerization as a molecular timer." Nat Chem Biol **3**(10): 619-29.
- Lund, E., S. Guttinger, et al. (2004). "Nuclear export of microRNA precursors." Science **303**(5654): 95-8.
- Lytle, J. R., T. A. Yario, et al. (2007). "Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR." Proc Natl Acad Sci U S A **104**(23): 9667-72.
- Ma, J. B., Y. R. Yuan, et al. (2005). "Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein." Nature **434**(7033): 666-70.

- Ma, L., G. M. Buchold, et al. (2009). "GASZ is essential for male meiosis and suppression of retrotransposon expression in the male germline." PLoS Genet **5**(9): e1000635.
- Ma, Q. and J. P. Whitlock, Jr. (1997). "A novel cytoplasmic protein that interacts with the Ah receptor, contains tetratricopeptide repeat motifs, and augments the transcriptional response to 2,3,7,8-tetrachlorodibenzo-p-dioxin." J Biol Chem **272**(14): 8878-84.
- Macrae, I. J., K. Zhou, et al. (2006). "Structural basis for double-stranded RNA processing by Dicer." Science **311**(5758): 195-8.
- Malone, C. D. and G. J. Hannon (2009). "Small RNAs as guardians of the genome." Cell **136**(4): 656-68.
- Mansfield, J. H., B. D. Harfe, et al. (2004). "MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression." Nat Genet **36**(10): 1079-83.
- Matranga, C., Y. Tomari, et al. (2005). "Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes." Cell **123**(4): 607-20.
- Matthews, B. W. (1968). "Solvent content of protein crystals." Journal of Molecular Biology **33**(2): 491-497.
- Maurer-Stroh, S., N. J. Dickens, et al. (2003). "The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains." Trends Biochem Sci **28**(2): 69-74.
- McCoy, A. J., R. W. Grosse-Kunstleve, et al. (2005). "Likelihood-enhanced fast translation functions." Acta Crystallogr D Biol Crystallogr **61**(Pt 4): 458-64.
- Meister, G., C. Eggert, et al. (2001). "Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln." Curr Biol **11**(24): 1990-4.
- Meister, G., M. Landthaler, et al. (2004). "Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs." Mol Cell **15**(2): 185-97.
- Meyer, B. K., M. G. Pray-Grant, et al. (1998). "Hepatitis B virus X-associated protein 2 is a subunit of the unliganded aryl hydrocarbon receptor core complex and exhibits transcriptional enhancer activity." Mol Cell Biol **18**(2): 978-88.
- Min, J., A. Allali-Hassani, et al. (2007). "L3MBTL1 recognition of mono- and dimethylated histones." Nat Struct Mol Biol **14**(12): 1229-30.

- Miyoshi, K., H. Tsukumo, et al. (2005). "Slicer function of Drosophila Argonautes and its involvement in RISC formation." Genes Dev **19**(23): 2837-48.
- Miyoshi, T., A. Takeuchi, et al. (2010). "A direct role for Hsp90 in pre-RISC formation in Drosophila." Nat Struct Mol Biol **17**(8): 1024-6.
- Moazed, D. (2009). "Small RNAs in transcriptional gene silencing and genome defence." Nature **457**(7228): 413-20.
- Mochizuki, K. and Y. Matsui (2010). "Epigenetic profiles in primordial germ cells: global modulation and fine tuning of the epigenome for acquisition of totipotency." Dev Growth Differ **52**(6): 517-25.
- Moshkovich, N. and E. P. Lei (2010). "HP1 recruitment in the absence of argonaute proteins in Drosophila." PLoS Genet **6**(3): e1000880.
- Mueller, G. A., M. T. Miller, et al. "Solution structure of the Drosha double-stranded RNA-binding domain." Silence **1**(1): 2.
- Munn, K. and R. Steward (2000). "The shut-down gene of Drosophila melanogaster encodes a novel FK506-binding protein essential for the formation of germline cysts during oogenesis." Genetics **156**(1): 245-56.
- Murshudov, G. N., A. A. Vagin, et al. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. Acta Crystallographica Section D. **53**: 240-255.
- Nagao, A., T. Mituyama, et al. (2010). "Biogenesis pathways of piRNAs loaded onto AGO3 in the Drosophila testis." Rna **16**(12): 2503-15.
- Najbauer, J., B. A. Johnson, et al. (1993). "Peptides with sequences similar to glycine, arginine-rich motifs in proteins interacting with RNA are efficiently recognized by methyltransferase(s) modifying arginine in numerous proteins." J Biol Chem **268**(14): 10501-9.
- Nelson, C. J., H. Santos-Rosa, et al. (2006). "Proline isomerization of histone H3 regulates lysine methylation and gene expression." Cell **126**(5): 905-16.
- Nielsen, P. R., D. Nietlispach, et al. (2002). "Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9." Nature **416**(6876): 103-7.

- Niki, Y., T. Yamaguchi, et al. (2006). "Establishment of stable cell lines of Drosophila germ-line stem cells." Proc Natl Acad Sci U S A **103**(44): 16325-30.
- Nishida, K. M., T. N. Okada, et al. (2009). "Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in Drosophila germlines." Embo J **28**(24): 3820-31.
- Nishida, K. M., K. Saito, et al. (2007). "Gene silencing mechanisms mediated by Aubergine piRNA complexes in Drosophila male gonad." Rna **13**(11): 1911-22.
- Nowotny, M., S. A. Gaidamakov, et al. (2005). "Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis." Cell **121**(7): 1005-16.
- Ohara, T., Y. Sakaguchi, et al. (2007). "The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated." Nat Struct Mol Biol **14**(4): 349-50.
- Okada, C., E. Yamashita, et al. (2009). "A high-resolution structure of the pre-microRNA nuclear export machinery." Science **326**(5957): 1275-9.
- Okamura, K., S. Balla, et al. (2008). "Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster." Nat Struct Mol Biol **15**(6): 581-90.
- Olivieri, D., M. M. Sykora, et al. (2010). "An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila." Embo J **29**(19): 3301-17.
- Ooi, S. K., C. Qiu, et al. (2007). "DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA." Nature **448**(7154): 714-7.
- Pahlich, S., R. P. Zakaryan, et al. (2006). "Protein arginine methylation: Cellular functions and methods of analysis." Biochim Biophys Acta **1764**(12): 1890-903.
- Paik, W. K., J. Z. Farooqui, et al. (1980). "Protein methylation: cytochrome c methylation as a model system." Adv Enzyme Regul **19**: 471-86.
- Pak, J. and A. Fire (2007). "Distinct populations of primary and secondary effectors during RNAi in C. elegans." Science **315**(5809): 241-4.
- Pal-Bhadra, M., B. A. Leibovitch, et al. (2004). "Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery." Science **303**(5658): 669-72.

- Papp, I., M. F. Mette, et al. (2003). "Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors." Plant Physiol **132**(3): 1382-90.
- Pare, J. M., N. Tahbaz, et al. (2009). "Hsp90 regulates the function of argonaute 2 and its recruitment to stress granules and P-bodies." Mol Biol Cell **20**(14): 3273-84.
- Park, W., J. Li, et al. (2002). "CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana." Curr Biol **12**(17): 1484-95.
- Parker, J. S., S. M. Roe, et al. (2004). "Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity." Embo J **23**(24): 4727-37.
- Patil, V. S. and T. Kai (2010). "Repression of retroelements in Drosophila germline via piRNA pathway by the Tudor domain protein Tejas." Curr Biol **20**(8): 724-30.
- Pelisson, A., E. Sarot, et al. (2007). "A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the Drosophila ovary." J Virol **81**(4): 1951-60.
- Pelisson, A., S. U. Song, et al. (1994). "Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene." Embo J **13**(18): 4401-11.
- Pellizzoni, L., N. Kataoka, et al. (1998). "A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing." Cell **95**(5): 615-24.
- Pellizzoni, L., J. Yong, et al. (2002). "Essential role for the SMN complex in the specificity of snRNP assembly." Science **298**(5599): 1775-9.
- Pernot, G., M. Stoffel, et al. (2010). "Precise control of thermal conductivity at the nanoscale through individual phonon-scattering barriers." Nat Mater **9**(6): 491-5.
- Perrakis, A., T. K. Sixma, et al. (1997). "wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models." Acta Crystallogr D Biol Crystallogr **53**(Pt 4): 448-55.
- Petersen, C. P., M. E. Bordeleau, et al. (2006). "Short RNAs repress translation after initiation in mammalian cells." Mol Cell **21**(4): 533-42.
- Petoukhov, M. V. and D. I. Svergun (2005). "Global rigid body modeling of macromolecular complexes against small-angle scattering data." Biophys J **89**(2): 1237-50.

- Pfeffer, S., M. Zavolan, et al. (2004). "Identification of virus-encoded microRNAs." Science **304**(5671): 734-6.
- Pillai, R. S. (2005). "MicroRNA function: multiple mechanisms for a tiny RNA?" Rna **11**(12): 1753-61.
- Pillai, R. S. and S. Chuma (2012). "piRNAs and their involvement in male germline development in mice." Dev Growth Differ.
- Ponting, C. P. (1997). "Tudor domains in proteins that interact with RNA." Trends Biochem Sci **22**(2): 51-2.
- Qi, H., T. Watanabe, et al. (2011). "The Yb body, a major site for Piwi-associated RNA biogenesis and a gateway for Piwi expression and transport to the nucleus in somatic cells." J Biol Chem **286**(5): 3789-97.
- Qu, F., X. Ye, et al. (2005). "RDR6 has a broad-spectrum but temperature-dependent antiviral defense role in *Nicotiana benthamiana*." J Virol **79**(24): 15209-17.
- Ramos, A., D. Hollingworth, et al. (2006). "The structure of the N-terminal domain of the fragile X mental retardation protein: a platform for protein-protein interaction." Structure **14**(1): 21-31.
- Rand, T. A., S. Petersen, et al. (2005). "Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation." Cell **123**(4): 621-9.
- Rappsilber, J., P. Ajuh, et al. (2001). "SPF30 is an essential human splicing factor required for assembly of the U4/U5/U6 tri-small nuclear ribonucleoprotein into the spliceosome." J Biol Chem **276**(33): 31142-50.
- Rehwinkel, J., I. Behm-Ansmant, et al. (2005). "A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing." Rna **11**(11): 1640-7.
- Reinhart, B. J., E. G. Weinstein, et al. (2002). "MicroRNAs in plants." Genes Dev **16**(13): 1616-26.
- Reuter, M., P. Berninger, et al. (2011). "Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing." Nature **480**(7376): 264-7.
- Reuter, M., S. Chuma, et al. (2009). "Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile." Nat Struct Mol Biol **16**(6): 639-46.

- Rhoades, M. W., B. J. Reinhart, et al. (2002). "Prediction of plant microRNA targets." Cell **110**(4): 513-20.
- Riggs, D. L., P. J. Roberts, et al. (2003). "The Hsp90-binding peptidylprolyl isomerase FKBP52 potentiates glucocorticoid signaling in vivo." Embo J **22**(5): 1158-67.
- Rivas, F. V., N. H. Tolia, et al. (2005). "Purified Argonaute2 and an siRNA form recombinant human RISC." Nat Struct Mol Biol **12**(4): 340-9.
- Robine, N., N. C. Lau, et al. (2009). "A broadly conserved pathway generates 3'UTR-directed primary piRNAs." Curr Biol **19**(24): 2066-76.
- Rossmann, M. G. (1990). "The molecular replacement method." Acta Crystallogr A **46 (Pt 2)**: 73-82.
- Rouget, C., C. Papin, et al. (2010). "Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo." Nature **467**(7319): 1128-32.
- Ruby, J. G., C. Jan, et al. (2006). "Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans." Cell **127**(6): 1193-207.
- Ruby, J. G., C. H. Jan, et al. (2007). "Intronic microRNA precursors that bypass Drosha processing." Nature **448**(7149): 83-6.
- Saito, K., S. Inagaki, et al. (2009). "A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila." Nature **461**(7268): 1296-9.
- Saito, K., H. Ishizu, et al. (2010). "Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila." Genes Dev **24**(22): 2493-8.
- Saito, K., A. Ishizuka, et al. (2005). "Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells." PLoS Biol **3**(7): e235.
- Saito, K., K. M. Nishida, et al. (2006). "Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome." Genes Dev **20**(16): 2214-22.
- Saito, K., Y. Sakaguchi, et al. (2007). "Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends." Genes Dev **21**(13): 1603-8.
- Sanchez Alvarado, A. and P. A. Newmark (1999). "Double-stranded RNA specifically disrupts gene expression during planarian regeneration." Proc Natl Acad Sci U S A **96**(9): 5049-54.

- Scheufler, C., A. Brinker, et al. (2000). "Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine." Cell **101**(2): 199-210.
- Schirle, N. T. and I. J. MacRae (2012). "The crystal structure of human Argonaute2." Science **336**(6084): 1037-40.
- Schmidt, A., G. Palumbo, et al. (1999). "Genetic and molecular characterization of sting, a gene involved in crystal formation and meiotic drive in the male germ line of *Drosophila melanogaster*." Genetics **151**(2): 749-60.
- Schwarz, D. S., G. Hutvagner, et al. (2003). "Asymmetry in the assembly of the RNAi enzyme complex." Cell **115**(2): 199-208.
- Selbach, M., B. Schwanhausser, et al. (2008). "Widespread changes in protein synthesis induced by microRNAs." Nature **455**(7209): 58-63.
- Selenko, P., R. Sprangers, et al. (2001). "SMN tudor domain structure and its interaction with the Sm proteins." Nat Struct Biol **8**(1): 27-31.
- Sengoku, T., O. Nureki, et al. (2006). "Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa." Cell **125**(2): 287-300.
- Senti, K. A. and J. Brennecke (2010). "The piRNA pathway: a fly's perspective on the guardian of the genome." Trends Genet **26**(12): 499-509.
- Shen, B. and H. M. Goodman (2004). "Uridine addition after microRNA-directed cleavage." Science **306**(5698): 997.
- Shoji, M., T. Tanaka, et al. (2009). "The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline." Dev Cell **17**(6): 775-87.
- Siekierka, J. J., S. H. Hung, et al. (1989). "A cytosolic binding protein for the immunosuppressant FK506 has peptidyl-prolyl isomerase activity but is distinct from cyclophilin." Nature **341**(6244): 755-7.
- Sijen, T., F. A. Steiner, et al. (2007). "Secondary siRNAs result from unprimed RNA synthesis and form a distinct class." Science **315**(5809): 244-7.
- Sims, R. J., 3rd, L. A. Rojas, et al. (2011). "The C-terminal domain of RNA polymerase II is modified by site-specific methylation." Science **332**(6025): 99-103.

- Sinars, C. R., J. Cheung-Flynn, et al. (2003). "Structure of the large FK506-binding protein FKBP51, an Hsp90-binding protein and a component of steroid receptor complexes." Proc Natl Acad Sci U S A **100**(3): 868-73.
- Siomi, M. C., T. Mannen, et al. (2010). "How does the royal family of Tudor rule the PIWI-interacting RNA pathway?" Genes Dev **24**(7): 636-46.
- Siomi, M. C., K. Sato, et al. (2011). "PIWI-interacting small RNAs: the vanguard of genome defence." Nat Rev Mol Cell Biol **12**(4): 246-58.
- Slotkin, R. K. and R. Martienssen (2007). "Transposable elements and the epigenetic regulation of the genome." Nat Rev Genet **8**(4): 272-85.
- Sohn, S. Y., W. J. Bae, et al. (2007). "Crystal structure of human DGCR8 core." Nat Struct Mol Biol **14**(9): 847-53.
- Song, J. J. and L. Joshua-Tor (2006). "Argonaute and RNA--getting into the groove." Curr Opin Struct Biol **16**(1): 5-11.
- Soper, S. F., G. W. van der Heijden, et al. (2008). "Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis." Dev Cell **15**(2): 285-97.
- Specchia, V., L. Piacentini, et al. (2010). "Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons." Nature **463**(7281): 662-5.
- Stark, A., J. Brennecke, et al. (2005). "Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution." Cell **123**(6): 1133-46.
- Stec, I., S. B. Nagl, et al. (2000). "The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation?" FEBS Lett **473**(1): 1-5.
- Stefani, G. and F. J. Slack (2008). "Small non-coding RNAs in animal development." Nat Rev Mol Cell Biol **9**(3): 219-30.
- Sugawara, K., K. Morita, et al. (2001). "BIP, a BRAM-interacting protein involved in TGF-beta signalling, regulates body length in *Caenorhabditis elegans*." Genes Cells **6**(7): 599-606.
- Sugiyama, T., H. Cam, et al. (2005). "RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production." Proc Natl Acad Sci U S A **102**(1): 152-7.

- Sullivan, C. S., A. T. Grundhoff, et al. (2005). "SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells." Nature **435**(7042): 682-6.
- Suzuki, Y., M. Minami, et al. (2009). "The Hsp90 inhibitor geldanamycin abrogates colocalization of eIF4E and eIF4E-transporter into stress granules and association of eIF4E with eIF4G." J Biol Chem **284**(51): 35597-604.
- Svergun, D., C. Barberato, et al. (1995). CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. Journal of Applied Crystallography. **28**: 768-773.
- Takahashi, N., T. Hayano, et al. (1989). "Peptidyl-prolyl cis-trans isomerase is the cyclosporin A-binding protein cyclophilin." Nature **337**(6206): 473-5.
- Tam, O. H., A. A. Aravin, et al. (2008). "Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes." Nature **453**(7194): 534-8.
- Tanaka, S. S., Y. Toyooka, et al. (2000). "The mouse homolog of Drosophila Vasa is required for the development of male germ cells." Genes Dev **14**(7): 841-53.
- Tanaka, T., M. Hosokawa, et al. (2011). "Tudor domain containing 7 (Tdrd7) is essential for dynamic ribonucleoprotein (RNP) remodeling of chromatoid bodies during spermatogenesis." Proc Natl Acad Sci U S A **108**(26): 10579-84.
- Tang, G., B. J. Reinhart, et al. (2003). "A biochemical framework for RNA silencing in plants." Genes Dev **17**(1): 49-63.
- Taverna, S. D., H. Li, et al. (2007). "How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers." Nat Struct Mol Biol **14**(11): 1025-40.
- Thomson, T. and P. Lasko (2004). "Drosophila tudor is essential for polar granule assembly and pole cell specification, but not for posterior patterning." Genesis **40**(3): 164-70.
- Thomson, T. and H. Lin (2009). "The biogenesis and function of PIWI proteins and piRNAs: progress and prospect." Annu Rev Cell Dev Biol **25**: 355-76.
- Timerman, A. P., H. Onoue, et al. (1996). "Selective binding of FKBP12.6 by the cardiac ryanodine receptor." J Biol Chem **271**(34): 20385-91.

- Timmerman, A. P., G. Wiederrecht, et al. (1995). "Characterization of an exchange reaction between soluble FKBP-12 and the FKBP-ryanodine receptor complex. Modulation by FKBP mutants deficient in peptidyl-prolyl isomerase activity." J Biol Chem **270**(6): 2451-9.
- Tomari, Y., C. Matranga, et al. (2004). "A protein sensor for siRNA asymmetry." Science **306**(5700): 1377-80.
- Tomari, Y. and P. D. Zamore (2005). "Perspective: machines for RNAi." Genes Dev **19**(5): 517-29.
- Toyooka, Y., N. Tsunekawa, et al. (2000). "Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development." Mech Dev **93**(1-2): 139-49.
- Tripsianes, K., T. Madl, et al. (2011). "Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins." Nat Struct Mol Biol **18**(12): 1414-20.
- Trojer, P., G. Li, et al. (2007). "L3MBTL1, a histone-methylation-dependent chromatin lock." Cell **129**(5): 915-28.
- Turnbull, W. B. and A. H. Daranas (2003). "On the value of c: can low affinity systems be studied by isothermal titration calorimetry?" J Am Chem Soc **125**(48): 14859-66.
- Unhavaithaya, Y., Y. Hao, et al. (2009). "MILI, a PIWI-interacting RNA-binding protein, is required for germ line stem cell self-renewal and appears to positively regulate translation." J Biol Chem **284**(10): 6507-19.
- Vagin, V. V., A. Sigova, et al. (2006). "A distinct small RNA pathway silences selfish genetic elements in the germline." Science **313**(5785): 320-4.
- Vagin, V. V., J. Wohlschlegel, et al. (2009). "Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members." Genes Dev **23**(15): 1749-62.
- Van Duyne, G. D., R. F. Standaert, et al. (1991). "Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex." Science **252**(5007): 839-42.
- Vasileva, A., D. Tiedau, et al. (2009). "Tdrd6 is required for spermiogenesis, chromatoid body architecture, and regulation of miRNA expression." Curr Biol **19**(8): 630-9.

- Vezzoli, A., N. Bonadies, et al. (2010). "Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1." Nat Struct Mol Biol **17**(5): 617-9.
- Volkov, V. V. and D. I. Svergun (2003). Uniqueness of ab initio shape determination in small-angle scattering. Journal of Applied Crystallography. **36**: 860-864.
- Vourekas, A., Y. Kirino, et al. (2010). "Elective affinities: a Tudor-Aubergine tale of germline partnership." Genes Dev **24**(18): 1963-6.
- Wang, H. W., C. Noland, et al. (2009). "Structural insights into RNA processing by the human RISC-loading complex." Nat Struct Mol Biol **16**(11): 1148-53.
- Wang, J., J. P. Saxe, et al. (2009). "Mili interacts with tudor domain-containing protein 1 in regulating spermatogenesis." Curr Biol **19**(8): 640-4.
- Wang, L. and P. Ligoxygakis (2006). "Pathogen recognition and signalling in the Drosophila innate immune response." Immunobiology **211**(4): 251-61.
- Wang, Y., S. Juranek, et al. (2008). "Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex." Nature **456**(7224): 921-6.
- Wang, Y., S. Juranek, et al. (2009). "Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes." Nature **461**(7265): 754-61.
- Wang, Z., J. Song, et al. (2010). "Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to Cyp33 and HDAC-mediated repression." Cell **141**(7): 1183-94.
- Watanabe, T., S. Chuma, et al. (2011). "MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline." Dev Cell **20**(3): 364-75.
- Watanabe, T., Y. Totoki, et al. (2008). "Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes." Nature **453**(7194): 539-43.
- Waterhouse, P. M., M. W. Graham, et al. (1998). "Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA." Proc Natl Acad Sci U S A **95**(23): 13959-64.
- Wightman, B., I. Ha, et al. (1993). "Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans." Cell **75**(5): 855-62.
- Wiseman, T., S. Williston, et al. (1989). "Rapid measurement of binding constants and heats of binding using a new titration calorimeter." Analytical Biochemistry **179**(1): 131-137.

- Wong, R. L., B. J. Wlodarczyk, et al. (2008). "Mouse Fkbp8 activity is required to inhibit cell death and establish dorso-ventral patterning in the posterior neural tube." Hum Mol Genet **17**(4): 587-601.
- Xie, X., J. Lu, et al. (2005). "Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals." Nature **434**(7031): 338-45.
- Xie, Z., L. K. Johansen, et al. (2004). "Genetic and functional diversification of small RNA pathways in plants." PLoS Biol **2**(5): E104.
- Yabuta, Y., H. Ohta, et al. (2011). "TDRD5 is required for retrotransposon silencing, chromatoid body assembly, and spermiogenesis in mice." J Cell Biol **192**(5): 781-95.
- Yan, K. S., S. Yan, et al. (2003). "Structure and conserved RNA binding of the PAZ domain." Nature **426**(6965): 468-74.
- Yang, X. J. (2005). "Multisite protein modification and intramolecular signaling." Oncogene **24**(10): 1653-62.
- Yang, Y., Y. Lu, et al. (2010). "TDRD3 is an effector molecule for arginine-methylated histone marks." Mol Cell **40**(6): 1016-23.
- Yang, Z. R., R. Thomson, et al. (2005). "RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins." Bioinformatics **21**(16): 3369-76.
- Yekta, S., I. H. Shih, et al. (2004). "MicroRNA-directed cleavage of HOXB8 mRNA." Science **304**(5670): 594-6.
- Yekta, S., C. J. Tabin, et al. (2008). "MicroRNAs in the Hox network: an apparent link to posterior prevalence." Nat Rev Genet **9**(10): 789-96.
- Yi, R., Y. Qin, et al. (2003). "Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs." Genes Dev **17**(24): 3011-6.
- Yin, H. and H. Lin (2007). "An epigenetic activation role of Piwi and a Piwi-associated piRNA in *Drosophila melanogaster*." Nature **450**(7167): 304-8.
- Young, J. C., W. M. Obermann, et al. (1998). "Specific binding of tetratricopeptide repeat proteins to the C-terminal 12-kDa domain of hsp90." J Biol Chem **273**(29): 18007-10.

- Zamore, P. D., T. Tuschl, et al. (2000). "RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals." Cell **101**(1): 25-33.
- Zeng, Y. and B. R. Cullen (2004). "Structural requirements for pre-microRNA binding and nuclear export by Exportin 5." Nucleic Acids Res **32**(16): 4776-85.
- Zeytuni, N. and R. Zarivach (2012). "Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module." Structure **20**(3): 397-405.
- Zhang, D., H. G. Yoon, et al. (2005). "JMJD2A is a novel N-CoR-interacting protein and is involved in repression of the human transcription factor achaete scute-like homologue 2 (ASCL2/Hash2)." Mol Cell Biol **25**(15): 6404-14.
- Zhao, Q., L. Qin, et al. (2007). "Structure of human spindlin1. Tandem tudor-like domains for cell cycle regulation." J Biol Chem **282**(1): 647-56.
- Zheng, K., J. Xiol, et al. (2010). "Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway." Proc Natl Acad Sci U S A **107**(26): 11841-6.
- Zhou, X. Z., O. Kops, et al. (2000). "Pin1-dependent prolyl isomerization regulates dephosphorylation of Cdc25C and tau proteins." Mol Cell **6**(4): 873-83.

ACKNOWLEDGEMENTS

Watching the beautiful sea view from Santorini Island, I cannot think of a better place to finish writing this PhD thesis and thank all the people that contributed to the PhD project and to my life in general in Grenoble.

My first thanks goes to Stephen Cusack for giving me the opportunity to work in his lab. His experience, knowledge patience and loyalty to me cannot be described with few words. He provided everything I needed for accomplishing the goals of this project and gave me the independence and freedom to develop my own ideas and skills, in order to learn how to overcome many difficulties.

Next, I would like to thank my PhD jury members for accepting my invitation to examine and read my PhD work. Carlo Petosa, Winfried Weissenhorn and the two rapporteurs Andres Ramos and Michael Sattler. Also my TAC members Ramesh Pillai, Winfried Weissenhorn and Christoph Mueller for their guidance, help and contribution during the first 3 years of PhD.

Also, I could not ask for a better lab environment to work, since all the members of the group helped me in many ways to feel welcome and comfortable. Delphine, Juan, the new members Ivam, Emiko, Jorge, Piotr, made my life easier and happier at the lab and at the office. I would like especially to thank Thomas for helping me with everything I ever needed in the lab and of course the past members Eva and Marcello, for all their help in my first steps as a PhD student. Of course, Jan and Andres were my tutors in X-ray crystallography and ITC respectively, and without them I would never be able to overcome technical difficulties and understand basic experimental principles.

Moreover, I would like to thank the EMBL staff for providing all the things I even needed and the collaborators for the TDRD1 project. EMBL Core Facilities and IBS platforms for all the analyses of protein samples, Adam Round for SAXS measurements, analysis and teaching and Kostas Tripsianes and Michael Sattler for their time and effort to design and perform NMR experiments.

Furthermore, I would like to say a big thank you to Ramesh Pillai for all his support during these 4 years. Our useful discussions, his ideas and knowledge in the piRNA field were more than necessary and helpful for me to learn how to think scientifically and act accordingly to all the new challenges. The members of his group Micahel and Jordi, for teaching me new techniques and Elisa, Pietro and Raman for very useful discussions about the piRNA pathway.

In addition, I consider that nothing would have been achieved without the company of good friends inside and outside of EMBL. A big thanks to Emily, Carmen, Elisa, Cristina, Michael, Matthias, Miriam, Pietro, Giovanna and especially Jordi for offering great moments of laughter and relaxation (usually with the help of some beers!). Also, my very good friends outside work: Christos, Evi, Alexandra, Sofia, Kostis, Manu, Alexandros and all the rest Greek mafia for beautiful times and holiday trips and excursions.

Finally, a very big thanks you to the most important people in my life. My parents, who were always next to me in all the difficult stages of the PhD and of the life in Grenoble. They were supporting and encouraging me all the time and I would not have achieved my goals without their help. Next, my best friends Giannis D, Giannis K, Sofia and my cousin Nantia, for making me forget all the stress and adjusting their life according to my busy schedule in order to make my vacations and time in Greece always better and better. Also, a big big thank you to Anastasia, who is always supporting me and encouraging me during the late stages of the PhD, and making me feel all the time optimistic and happy. Lastly, I thank b.a.s for all the support and help during my first steps in Grenoble.

Σας ευχαριστώ όλους!!

Thank you all!