Le complexe MILI/mHEN1 et études fonctionnelles des protéines DrTDRD1 et DrMOV10L

Stephanie Eckhardt

EMBL

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité: **Biologie cellulaire**

Arrêté ministériel: 7 août 2006

Présentée par

Stephanie ECKHARDT

Thèse dirigée par **Ramesh PILLAI**

Préparée au sein du **Laboratoire de Régulation de l'expression des gènes, Laboratoire Européen de Biologie Moléculaire (EMBL)** et dans **l'École Doctorale Chimie et Science du vivant**

Le complexe MILI/mHEN1 et études fonctionnelles des protéines *Dr*TDRD1 et *Dr*MOV10L

Thèse soutenue publiquement le **12.04.2011** devant le jury composé de :

Dr Winfried WEISSENHORN	
Professeur d'Université Joseph Fourier,	Président
Dr Oliver MÜHLEMANN	
Professeur d'Université de Berne,	Rapporteur
Dr Marc BILLAUD	
Directeur de recherche IAB,	Rapporteur
Dr Donal O'CARROLL	
Chef d'équipe, EMBL Monterotondo,	Examinateur
Dr Stephen CUSACK	
Directeur de l'antenne EMBL Grenoble,	Examinateur
Dr Ramesh PILLAI	
Chef d'équipe, EMBL Grenoble,	Directeur de thèse

EMBL

UNIVERSITÉ DE GRENOBLE

THESIS

for obtaining the degree

DOCTOR OF THE UNIVERSITY GRENOBLE

Field: **Cell biology**

Ministerial Decree: 7 august 2006

Submitted by

Stephanie ECKHARDT

Thesis directed by **Ramesh PILLAI**

Prepared in the **Regulation of Gene Expression Laboratory, European Molecular Biology Laboratory (EMBL)** and the **Graduate School Chemistry and Life Sciences**

The MILI/mHEN1 complex and functional studies of *Dr*TDRD1 and *Dr*MOV10L

Thesis publicly defended on the **12.04.2011** in front of the following committee:

Dr Winfried WEISSENHORN	
Professor at the University Joseph Fourier,	Chair
Dr Oliver MÜHLEMANN	
Professor at the University Bern,	Reviewer
Dr Marc BILLAUD	
Research Director IAB,	Reviewer
Dr Donal O'CARROLL	
Groupleader, EMBL Monterotondo,	Examiner
Dr Stephen CUSACK	
Head of Outstation EMBL Grenoble,	Examiner
Dr Ramesh PILLAI	
Groupleader, EMBL Grenoble,	Thesis Director

Mots clés

piARNs, protéines Piwi, mHEN1, HEN1 body, *Dr*TDRD1, *Dr*MOV10L

Key words

piRNAs, Piwi proteins, mHEN1, HEN1 body, *Dr*TDRD1, *Dr*MOV10L

“There are in fact two things, science and opinion; the former begets knowledge, the latter ignorance.”

Hippocrates

Law, Bk IV, circa 395 B.C

“Il y a en effet deux choses, la science et l’opinion ; celle-là conduit au savoir, celle-ci à l’ignorance.”

Hippocrate

La Loi, Livre IV, environ 395 AJC

Argonaute proteins associate with small RNAs to participate in gene regulatory processes. Piwi proteins are a sub-clade of Argonaute that are mainly expressed in the animal germline. They bind to Piwi-interacting RNAs (piRNAs) and contribute to genome integrity through transposon silencing. To investigate factors involved in the piRNA pathway, mice and *Danio rerio* (zebrafish) were exploited as model systems to investigate the piRNA methyltransferase mHEN1, the Tudor domain-containing protein TDRD1 and the helicase MOV10L.

One defining feature of piRNAs is the presence of a 2`-O-methyl group on the 3` terminal nucleotide which is catalyzed by the animal homologs of the plant RNA methyltransferase HEN1. The mouse homolog, mHEN1, was shown to methylate RNA substrates *in vitro*. The connection of mHEN1 to the piRNA pathway was shown in this study by identifying an interaction between the C-terminal region of mHEN1 and the N-terminal region of MILI adjacent to the PAZ domain, but distinct from that used for methylation-dependent interaction with Tudor proteins. These results suggest the presence of a composite interaction domain formed by the PAZ domain and the N-terminal region proximal to it. In purified mouse germ cells and testis sections mHEN1 was detected, using specific antisera, in a cytoplasmic granule distinct from the chromatoid body that contains the Piwi proteins. This study implicates the HEN1 body as a potential site of piRNA biogenesis.

As a result of challenges encountered with mice for manipulation of the germline, zebrafish embryos were developed as an alternate model system for studying piRNA pathway components in our laboratory. The role of two piRNA pathway factors, *Dr*TDRD1 and *Dr*MOV10L were investigated, *in vivo*. In mice MOV10L is thought to be involved in the stabilization or loading of piRNAs into Piwi proteins. The role of TDRD1 was found to be an essential component of the piRNA silencing complex. The expression profile of zebrafish *Dr*TDRD1 and *Dr*MOV10L was determined from the one-cell egg stage to the adult fish along with the role of the 3`UTR of *Dr*TDRD1 in restricting expression to the germ cells. Using morpholino knockdowns, *Dr*MOV10L was functionally linked to the transposon repression pathway. Taken together, these studies establish the mechanism of recruitment of the piRNA methyltransferase and take the first steps towards providing a manipulatable *in vivo* system for studying piRNA factors.

Les protéines argonautes, en s'associant a des petits ARNs, participent à la régulation génique. Les protéines Piwi appartenant à la sous-classe des argonautes sont principalement exprimées dans les cellules de lignée germinale. Elles se lient aux petits ARNs interagissant avec les protéines Piwi (piARN) et contribuent au maintien de l'intégrité génique grâce à des transposons responsables de la mise sous silence génique. Afin d'étudier les différents facteurs impliqués dans la voie de piARN, les modèles souris (*Mus musculus*) et danio zébré ou poisson zèbre (*Danio rerio*) ont été utilisés pour explorer les méthyltransférases de piARN, mHEN1, la protéine contenant le domaine Tudor 1, TDRD1 et l'hélicase MOV10L.

Les piARN se caractérisent par la présence d'un groupe 2`-O-methyl sur le dernier nucléotide en extrémité 3`, ajouté par un homologue mammifère de la méthyltransférase d'ARN de plantes HEN1. Il a été montré que l'homologue chez la souris, mHEN1, pouvait méthyler des ARNs *in vitro*. Dans cette étude il a été démontré en particulier que cette mHEN1 était impliquée dans la voie de piARN grâce à l'identification d'une interaction entre la région C-terminale de mHEN1 et la région N-terminale de MILI adjacente au domaine PAZ, région d'interaction différente de celle observée pour les protéines Tudor. Ces résultats suggèrent la présence d'un domaine composite d'interaction formé par le domaine PAZ et la région proximale de son extrémité N-terminale. La protéine mHEN1 a été détectée dans des granules cytoplasmiques à partir de cellules germinales purifiées et de sections de testicules en utilisant des anti-sera spécifiques, alors que les protéines Piwi ont été détectées dans des corps chromatoïdes. Les résultats de cette étude impliquent que HEN1 serait un site potentiel pour la biogenèse des piARN.

La manipulation de lignées germinales de souris s'avère particulièrement difficile. Un système modèle alternatif, à partir d'embryons de poisson zèbre a été développé dans le laboratoire pour l'étude de la voie des piARN. Le rôle de deux facteurs, *Dr*TDRD1 et *Dr*MOV10L, impliqués dans cette voie ont été étudiés *in vivo*. Dans les souris, MOV10L serait impliquée dans la stabilisation ou dans le recrutement des piARN vers les protéines Piwi. Le rôle de TDRD1 s'est révélé essentiel pour le complexe de silence des piARN. Le profile d'expression des protéines *Dr*TDRD1 et *Dr*MOV10L de danio zèbre a été déterminé à différents phases de développement, du stade œuf unicellulaire au stade adulte. La spécificité du profile d'expression dans les cellules germinales est dépendante de la région 3` non traduite de *Dr*TDRD1. En outre il a été montré, grâce à l'utilisation de morpholinos, que

*Dr*MOV10L était fonctionnellement impliquée dans la voie de répression des transposons. L'ensemble de ces études a permis de comprendre le mécanisme de recrutement d'une méthyltransférase de piARN et de mettre au point un système modèle permettant leur étude *in vivo*.

This thesis would not have been possible without the support of the following people. I am eternally grateful to my **parents, grandparents, future parents in law** and lastly **my fiancé Louis Hutin** for their loving help.

I would like to thank **Prof. Dr. Wolfgang Nellen** from the University Kassel who has had a big influence in my career and who introduced to me the world of research. **Dr. Kriton Kalantidis,** from the IMBB FORTH, Heraklion, Greece, **Dr. Fredrik Söderbom,** from the University Uppsala and BMC, Sweden, both formed my view of science in a way that encouraged me to initiate PhD studies.

To my thesis advisor, **Dr. Ramesh Pillai,** who accepted me as his first PhD student. I thank him for giving me the opportunity to work independently on many different projects.

Thanks to my Thesis advisory committee: **Dr. Stephen Cusack, Dr. Saadi Khochbin** and **Dr. Donal O'Carroll** for their good advice and support.

Special thanks to the members of our laboratory **Dr. Michael Reuter, Jordi Xiol, and Elisa Cora** for all our scientific discussions and helpful advice.

I am deeply grateful to my wonderful lab-mates who helped me not only with fruitful scientific discussions, but also kept my motivation going. Special thanks to my close colleagues, **Dr. Sebastien Muller, Dr. Rodrigo Louro, Dr. Philipp Berninger** and **Dr. Radha Raman Pandey** for keeping my spirit focused on the true meaning of life. In the same breath, I have to thank **Dr. Adam Round**.

While at the EMBL many trainees joined the lab. I would like to thank **Anisa, Aurélien, Elsa, Emerens, Emilia, Hivin, Kirsten, Maartje, Magdalena, Ricardo, Sylvain, Zhaolin, Anna** and **Margorzata**.

I have not forgotten **Jérôme Boudin** a former technician in the lab for his strength and power to do well. To the EMBL administration staff **Sylviane Troger**, **Dominique Lancon**, **Virginie Bertholet, Franscois Tronel**, and **Mary-Jane Villot** for making daily life so much easier.

A big thanks to **Nicolas Martinelli**, **Nicolas Martinez, Dr. Danielle Desravines** and **Philippe Mas** for helping with the translations in the thesis and **Dr. Andrew McCarthy**, **Dr. Max Nanao**, **Dr. Chloe Zubieta** and **Dr. Simon Conn** for proofreading of the manuscript.

Last but not least thanks to all members of the **EMBL community**, whose paths I have crossed the last four years.

Ce travail de thèse n'aurait pas aboutit sans le support des personnes suivantes : je serais éternellement reconnaissante a **mes parents**, **grands parents**, **futurs beaux parents** et **mon fiance Dr Louis Hutin** pour leur aide bienveillante.

Je souhaiterais remercier le **Dr Wolfgang Nellen** de l'Université de Kassel qui a représenté dans ma carrière une grande influence et aussi m'a initiée à l'univers de la Recherche. Ensemble, les **Dr Kriton Kalantidis,** du IMBB FORTH, Heraklion, Grèce, et **Dr Fredrik Söderbom**, de l'Université d' Uppsala et BMC, Suède, m'ont incitée à poursuivre une thèse par leur vision respective du domaine scientifique.

A mon responsable de thèse, le **Dr Ramesh Pillai,** qui m'a accueillie comme premier étudiant en thèse au sein de son laboratoire. Je le remercie pour l'opportunité de travailler de façon indépendante sur des projets varies.

Je remercie mon jury de thèse : **Dr Stephen Cusack, Dr Saadi Khochbin** and **Dr Donal O'Carroll** pour leurs conseils pertinents et leurs encouragements.

Je remercie particulièrement les membres de notre laboratoire : **Dr Michael Reuter, Jordi Xiol** et **Elisa Cora** pour toutes les discussions scientifiques et les conseils justes que nous avons partagés.

Je voudrais également exprimer ma profonde gratitude a mes formidables collègues qui ne m'ont pas seulement aidée a travers de fructueuses discussions scientifiques, mais aussi ont su trouver les mots pour garder ma motivation intacte.Des remerciements particuliers a mes collègues proches: **Dr Sebastien Muller, Dr Rodrigo Louro, Dr Philipp Berninger** et **Dr Radha Raman Pandey** qui m'ont aidée à garder l'esprit concentré sur les véritables valeurs de l'existence. Dans la même veine, je pense également au **Dr Adam Round**.

Pendant ce travail de thèse, beaucoup de stagiaires ont intégré les activités du laboratoire. Je voudrais donc remercier **Anisa, Aurélien, Elsa, Emerens, Emilia, Hivin, Kirsten, Maartje, Magdalena, Ricardo, Sylvain, Zhaolin, Anna** et **Margorzata**.

Je n'ai pas oublié **Jérôme Boudin**, technicien de l'équipe pour son énergie et sa disponibilité. A l'équipe administrative de l'EMBL : **Sylviane Troger**, **Dominique Lancon**, **Virginie Bertholet, Franscois Tronel**, et **Mary-Jane Villot** pour rendre mes activités quotidienne beaucoup plus aisées.

Un grand merci a **Nicolas Martinelli**, **Nicolas Martinez, Dr Danielle Desravines** et **Philippe Mas** pour leur aide sur les traductions inhérentes a mon mémoire de thèse; et **Dr Andrew McCarthy**, **Dr Max Nanao**, **Dr Chloe Zubieta** et **Dr Simon Conn** pour les corrections qu'ils ont apportes a celui ci.

Pour finir, merci a tous les membres de **la communauté EMBL** de Grenoble avec qui nos chemins se sont croises durant ces quartes dernières années.

3`		3 prime
5`		5 prime
7mG		7-methylguanine
A		Adenine
AGO		Argonaute protein
Ala	(A)	Alanine
Arg	(R)	Arginine
Asn	(N)	Asparagine
Asp	(D)	Aspartic acid
ATP		Adenosinetriphosphate
AUB		Aubergine
c		Control
casiRNA		Cis-acting siRNAs
CB		Chromatoid body
cDNA		Copy DNA
C. elegans		*Caenorhabditis elegans*
CTD		C-terminal domain
C-terminus		Carboxyl-terminus
Cys	(C)	Cysteine
DCL		Dicer-like
Dcr		Dicer
D. melanogaster		*Drosophila melanogaster*
DNA		Deoxyribonucleic acid
DNase		Deoxyribonuclease
DND1		Dead end protein1
Dnmt		*de novo* DNA methyltransferase
dpf		Days post fertilization
D. rerio (Dr)		*Danio rerio*
Drosha		The Hebrew and Yiddish word for sermon in Judaism.
dsDNA		Double stranded DNA
dsRBD		Double stranded RNA binding domain
dsRNA		Double-stranded RNA
E. coli		*Escherichia coli*
e. g.		*Exempli gratia*
endo-siRNA		Endogenous siRNAs
et al.		And co-workers
G		Guanine
Gcl		germ cell less
GFP		Green fluorescent protein
Gln	(Q)	Glutamine
Glu	(E)	Glutamic acid
Gly	(G)	Glycine
GST		Glutathione-S-transferase
GW		Glycine (G)-tryptophan (W)
HeLa cells		Henrietta Lacks cells
HEN		HUA ENHANCER
HILI		human Piwi-like 2
His	(H)	Histidine
HIV		Human immunodeficiency virus
HIWI		Human Piwi-like 1
HIWI2		Human Piwi-like 4
hPIWIl3		Human Piwi-like 3
H. sapiens		*Homo sapiens*
IAP		Intracisternal-A particle
i. e.		*id est*
Ile	(I)	Isoleucine

Leu (L)	Leucine
LINE	Long interspersed nuclear elements
lsiRNA	Long siRNAs
LTR	Long terminal repeats
Lys (K)	Lysine
Met (M)	Methionine
MILI	Piwi-like protein 2 (Mus musculus)
MIWI	Piwi-like protein 1 (Mus musculus)
MIWI2	Piwi-like protein 4 (Mus musculus)
miRNA	microRNA
miRNP(s)	miRNA-containing effector complex
Mirtron	miRNA from intron
MMA	Monomethylated Arginine
M. musculus (m)	*Mus musculus*
MOV	Moloney leukemia virus homolog
mRNA	Messenger RNA
MTR1	Mouse Tudor repeat 1
MVH	Mouse Vasa homolog
natsiRNA	Natural antisense siRNAs
ncRNA	Non-coding RNA
N-terminus	Amino terminus
nt	Nucleotide(s)
OD	Optical density
OH	Hydroxyl
ORF	Open reading frame
Pasha	Partner of Drosha
PAZ	PIWI-Argonaute-Zwille domain
P-body	Processing-body
PCR	Polymerase chain reaction
PGC	Primordial germ cells
pH	*Pondus Hydrogenii* or *potentia Hydrogenii*
Phe (F)	Phenylalanine
piRNA	Piwi-interacting RNAs
Piwi	P-element induced wimpy testis
Pol	Polymerase
pre-miRNA	Precursor miRNA
pri-miRNA	Primary miRNA
PRMT	Protein arginine methyltransferase
Pro (P)	Proline
R2D2	Two dsRNA-binding domains (R2), associated with DCR-2
RanBPM	Ran-GTP binding protein
RanGTP	Ran guanosine triphosphate
rasiRNA	Repeat-associated short interfering RNA
RBD	RNA binding domain
RDE	RNAi defective family member
RDR	RNA-dependent polymerase
RdRP	RNA-dependent polymerase
RecQ	ATP-dependent DNA helicase
RISC	RNA-induced silencing complex
RIWI	Rat PIWI
RLC	RISC-loading complex
RNA	Ribonucleic acid
RNAi	RNA interference
RNase	Ribonuclease
RNP	Ribonucleoprotein particle
rRNA	Ribosomal RNA
RT-PCR	Reverse transcription polymerase chain reaction

SAH	S-adenosylhomocysteine
SAM	S-Adenosyl methionine
scanRNA	Scanning RNA
SD	Standard deviation
sDMA	Symmetrically di-methylated arginines
Ser (S)	Serine
*Sf*21	(IPLB-Sf21AE) *Spodoptera frugiperda* cell line
SINE	short intersperse nuclear element
siRNA	Small interfering RNA
SIWI	Silk worm Aubergine protein
Sm proteins	Smith antigene proteins
snoRNA	Small nucleolar RNA
snRNPs	Small nuclear ribonucleoproteins
SPR	Surface Plasmon Resonance
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
Ste	Stellate
Su(ste)	Suppressor of Stellate
T	Thymine
TAC	Thesis advisory committee
tasiRNA	Trans-acting siRNAs
TDRD	Tudor domain-containing
Thr (T)	Threonine
TNP	Transition protein
tRNA	Transfer RNA
Trp (W)	Tryptophan
Tyr (Y)	Tyrosine
U	Uracil
UTR	Untranslated region
UV	Ultraviolet
Val (V)	Valine
Vls	*Valois*
Wago	Worm-specific Argonaute subfamily
WDR77	WD repeat domain 77
WT	Wild type
Yb helicase	female sterile (1) helicase
ZILI	Zebrafish Piwi-like 2
ZIWI	Zebrafish Piwi-like 1
Zuc	Zucchini

%	Percent
Å	Ångström
bp	Base pair(s)
°C	Degrees centigrade
dpc	Days post coitum
dpp	Days post partum
fmol	Femtomole
g	Gramm(s)
h	Hour(s)
hpf	Hours post fertilization
kb	Kilo base(s)
kDa	Kilo Dalton(s)
L	Liter(s)
M	Molar
mA	Milliampere(s)
mg	Microgramme(s)
min	Minute(s)
mJ	Millijoule(s)
mL	Millilitre(s)
mm	Millimetre(s)
mM	Millimolar
µCi	Microcurie(s)
µg	Microgramme(s)
µJ	Microjoule(s)
µL	Microlitre(s)
ng	Nanogramme(s)
pmol	Picomole(s)
s	Second(s)
U	Unit(s)
V	Volt(s)
w/v	Weight/volume
W	Watt(s)

Chapter 1 : Introduction **1**

1.1 Overview introduction **2**

1.2 Introduction Aperçu **2**

1.3 Introduction to the field of small RNAs **3**

1.3.1 Discovery of small RNA gene silencing 3
1.3.2 The Argonaute proteins 4
1.3.3 The miRNA pathway 7
1.3.4 The siRNA pathway 13
1.3.5 Piwi-interacting RNAs, a germline-specific small RNA class 17
1.3.6 piRNA biogenesis 20
1.3.7 piRNA function 25

1.4 Factors involved in the piRNA pathway **30**

1.4.1 The methyltransferase HEN1 adds methylation marks on the 3` end of small RNAs 30
1.4.2 Piwi proteins carry methylated Arginines at their N-terminal region 34
1.4.3 Tudor domain-containing proteins recognize post-transcriptional modifications on Piwi proteins 37
1.4.4 Helicases in the piRNA pathway 40

1.5 Project description **43**

1.5.1 The role of the piRNA methylation and the methyltransferase mHEN1 43
1.5.2 TDRD1 and MOV10L, two proteins involved in the piRNA pathway 44

1.6 Description du project **46**

1.6.1 Le rôle de la méthylation des piRNA et de la méthyltransférase mHEN1 46
1.6.2 TDRD1 et MOV10L, deux protéines impliquées dans la voie des piARN 47

Chapter 2: Direct recruitment of the piRNA methyltransferase mHEN1 via interaction with the Piwi protein MILI **49**

2.1 Contribution remark **50**

2.2 Remarques concernant ma contribution **50**

2.3 Direct recruitment of the piRNA methyltransferase mHEN1 via interaction with the Piwi protein MILI **51**

2.3.1 Abstract 52
2.3.2 Resume 53
2.3.3 Introduction 54
2.3.4 Results 57
2.3.5 Discussion 70
2.3.6 Acknowledgements 75
2.3.7 Methods 75
2.3.8 Supplementary data 78

2.4 Additional experiments for the characterization of mHEN1/MILI complex **80**

2.4.1 Attempt to express the mHEN1/ MILI complex in insect cells **80**

2.4.1.1 Aim 80
2.4.1.2 Results 81
2.4.1.3 Discussion 84
2.4.1.4 Method *Sf*21 expression of MILI and mHEN1 84

2.4.2 RNF17 is does not influence the 3'end processing of piRNAs **86**

2.4.2.1 Aim ... 86
2.4.2.2 Results ... 87
2.4.2.3 Discussion ... 89
2.4.2.4 Methods ... 89

2.4.3 Structural insight of binding of the methylated piRNAs by the MIWI PAZ domain ... 90

2.4.3.1 Contribution remarks ... 90
2.4.3.2 Remarques concernant ma contribution ... 90
2.4.3.3 Aim ... 91
2.4.3.4 Results ... 92
2.4.3.5 Discussion ... 97
2.4.3.6 Methods ... 99

Chapter 3: Analysis of the MOV10L and TDRD1 homologs in zebrafish ... 101

3.1 Contribution remark ... 102

3.2 Remarques concernant ma contribution ... 102

3.3 Analysis of the MOV10L and TDRD1 homologs in zebrafish ... 103

3.3.1 Abstract ... 104
3.3.2 Resume ... 104
3.3.3 Introduction ... 105
3.3.4 Results ... 112
3.3.5 Discussion *Dr*MOV10L and *Dr*TDRD1 ... 127
3.3.6 Methods ... 130

3.4 Additional experiments ... 133

3.4.1 Development of a reporter assay for functional studies of piRNAs ... 133

3.4.1.1 Aim ... 133
3.4.1.2 Results ... 134
3.4.1.3 Discussion ... 138
3.4.1.4 Method ... 139

3.4.2 Identification of possible interaction partners of the MYND domain of TDRD1 using a yeast two hybrid ... 140

3.4.2.1 Aim ... 140
3.4.2.2 Results ... 141
3.4.2.3 Discussion ... 142
3.4.2.4 Method ... 144

Chapter 4: General conclusions and future perspectives ... 145

4.1 General conclusions and future perspectives ... 146

4.2 Conclusions générales et perspectives ... 151

Supplements ... 157

References ... 186

Figure 1: Domains of the Argonaute proteins 6
Figure 2: miRNA pathways 12
Figure 3: siRNA pathway 16
Figure 4: Germ cell development and spermatogenesis in mouse 19
Figure 5: The Ping-pong cycle in *Drosophila* 24
Figure 6: The methyltransferase of HEN1 is highly conserved. 33
Figure 7: Symmetrical dimethylation at the N-termini of mouse and *Drosophila* Piwi proteins 36
Figure 8: Structure of the Tudor domain 11 of *Drosophila* TUDOR 39
Figure 9:Endogenous mHEN1 in testes extracts is an RNA methyltransferase 58
Figure 10: mHEN1 interacts with MILI 60
Figure 11: mHEN1 interacts with MILI independent of RNAs and Piwi methylation status 62
Figure 12: The mHEN1 interaction site on MILI is close to the PAZ domain 64
Figure 13: MILI interacts with the N-terminus of mHEN1 close to methyltransferase domain 66
Figure 14: Cytoplasmic localization of mHEN1 68
Figure 15: Model 74
Figure 16: MILI/mHEN1 complex studies in *Sf*21 82
Figure 17: RNF17 and mHEN1 show a similar localization pattern. 86
Figure 18: RNF17 does not affect the 2'O methylation of the 3'end of piRNAs 88
Figure 19: Purified PAZ domains and mutants 93
Figure 20: UV-cross-linking experiments of PAZ domain and mutants to RNAs. 96
Figure 21: Piwi expression during zebrafish development 107
Figure 22: *Dr*MOV10L alignment 113
Figure 23: *Dr*TDRD1 alignment 114
Figure 24: RT-PCRs to determine the expression of fish piRNA-pathway related mRNAs. 115
Figure 25: Tests of α-*Dr*TDRD1 and α-*Dr*MOV10L antibodies 117
Figure 26: Immunoprecipitations of zebrafish piRNA pathway componments 119
Figure 27: Binding studies of *Dr*TDRD1 and *Dr*MOV10L 121
Figure 28: Localization of *Dr*TDRD1 123
Figure 29: Morpholino knockdowns 126
Figure 30: Reporter constructs and image processing 136
Figure 31: Analysis of the reporter constructs 137
Figure 32: Structure of the ETO MYND domain 140
Figure 33: Yeast-two-hybrid 141

Figure S 1: Alignment of HEN1 from different animals 78
Figure S 2: Endogenous mHEN1 from testes extracts is devoid of any associated small RNAs 79

Chapter 1 : Introduction

1.1 Overview introduction

The introductory part of this thesis first concerns the discovery of small RNA mediated gene silencing. To date, three main groups of small non-coding RNAs (ncRNAs) are the focus of interest: microRNAs (miRNAs), small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs). The enzymatic core in all small ncRNA mechanisms is the Argonaute protein family, which is summarized before the miRNA and siRNA pathways are discussed. The focus then turns to piRNAs, their biogenesis and functions. Finally, some known factors of the piRNA pathway are highlighted, in particular the RNA methyltransferase mHEN1, the Tudor domain-containing family and Helicases.

1.2 Introduction Aperçu

L'introduction de cette thèse concerne tout d'abord la répression de gènes médiée par les petits ARNs. A ce jour, trois groupes d'ARNs non-codants (ncARNs) sont privilégiés : les microARNs (miARNs), les petits ARNs interférents (siARNs) et les ARNs interagissant avec Piwi (piARNs). Les enzymes appartenant à la famille des protéines Argonautes sont à l'origine de tous les mécanismes incluant les petits ncARNs. La famille des protéines Argonautes sera introduite, puis nous discuterons des voies de signalisation des miARNs et siARNs. Nous aborderons ensuite le sujet des piARNs, leur biogénèse ainsi que leurs fonctions. Enfin, quelques facteurs connus de la voie des piARNs seront développés, en particulier la ARN methyltransferase mHEN1, la famille des protéines à domaine Tudor et les hélicases.

1.3 Introduction to the field of small RNAs

1.3.1 Discovery of small RNA gene silencing

Genomes contain the entire hereditary information of an organism and are encoded in either DNA or for many types of virus in RNA. Human DNA is packed in 22 autosomal chromosome pairs and one sex-determining pair. The human genome contains about 6 billion DNA base pairs. Remarkably, this only codes for 23,000 proteins, which is approximately 1.5 % of the genome, the remaining 98.5% consists of non-coding (nc)RNA genes, regulatory sequences, introns and regions of unknown function which contain evolutionary artifacts that have no present-day usage (Abdellah 2004; Lander *et al.* 2001).

The huge amount of genomic information in a cell has to be controlled and regulated. The diverse mechanisms that cells and viruses use to control the flow of genetic information or gene production are collectively referred to as gene regulation. Known mechanisms include regulation of the DNA-RNA transcription up to and including post-translational modification of a protein. However, not only the sequence and regulatory processes dictate how the cell functions, changes in the phenotype or gene expression can be caused by other factors. The study of these factors and underlying mechanisms is called epigenetics, one part of which concerns the study of small non-coding RNAs that regulate gene expression.

In 1998, Fire and Mello demonstrated that double stranded RNA (dsRNA) triggers gene silencing in *C. elegans* much better than single stranded RNA (Fire *et al.* 1998). The Nobel Prize for Physiology or Medicine in 2006 was awarded for their discovery that small interfering RNAs (siRNAs) caused the suppression of gene activity in a homology-dependent manner, the RNA interference (RNAi) pathway.

In 1993, the first microRNA (miRNA, or small temporal RNA: stRNAs), named lin-4 (abnormal cell linage 4), was discovered as a negative regulator of LIN14, a protein which builds a temporal gradient by its decreasing expression during the larvae development in the nematode *C. elegans*. This 22 nt miRNA is partially complementary to the 3`untranslated region (UTR) of its target mRNA, repressing lin-14 translation and influencing its stability (Lee *et al.* 1993; Wightman *et al.* 1993). After several years of research a second miRNA, let-

7, was discovered in *C. elegans* (Reinhart *et al.* 2000). Let-7 was found to be highly conserved (Pasquinelli *et al.* 2000) and a search for more regulatory small RNA sequences started.

To date, three main groups of small RNAs are in the main focus of interest: microRNAs, small interfering RNAs (siRNAs) and since 2006 the Piwi-interacting RNAs (piRNAs). They all have the ability to regulate a broad variety of biological processed (Choudhuri 2009).

1.3.2 The Argonaute proteins

At the core of the small ncRNA pathways is the Argonaute protein family. These proteins were first discovered due to their function for stem cell self-renewal and germ cell development and maintenance (Bohmert *et al.* 1998; Cox *et al.* 1998). This protein family is highly conserved evolutionarily and several members are found in many species: humans 8, mice 7, *Drosophila melanogaster* 5 and *Arabidopsis thaliana* 10 Argonaute genes (Hock & Meister 2008) (see Figure 1 A).

All Argonautes contain four domains: The N-terminus, a PIWI/Argonaut/Zwille (PAZ) domain, a MID domain and the PIWI domain (see Figure 1 B and C) (Hutvagner & Simard 2008). The PAZ domain shows similarities to an oligonucleotide and oligosaccharide binding (OB) -fold and is responsible for the specific recognition of the 3` end of small ncRNAs (Lingel *et al.* 2003, 2004; Ma *et al.* 2004; Song *et al.* 2003; Yan *et al.* 2003). The MID domain adopts a Rossmann-like fold, and contains a highly conserved 5`-phosphate-binding pocket for small RNAs (Boland *et al.* 2010; Frank *et al.* 2010; Wang *et al.* 2008b; Yuan *et al.* 2005). The Piwi domain displays an RNaseH fold and contains a conserved DDH/L motif, which is responsible for RNA cleavage activity of some Argonaute members (Parker *et al.* 2004; Song *et al.* 2003; Song *et al.* 2004; Wang *et al.* 2008a; Wang *et al.* 2009b; Yan *et al.* 2003; Yuan *et al.* 2005). It is presently unclear why other Argonautes with the conserved PIWI domain and DDH motifs are inactive.

In the Argonaute protein family three distinct subfamilies can be distinguished: the AGO, the Piwi and a worm-specific subclade of the Argonautes (Wago) (Sasaki *et al.* 2003; Tolia & Joshua-Tor 2007). The ubiquitously expressed AGO proteins bind to miRNAs and

siRNAs, and are involved in gene silencing events at both transcriptional and post-transcriptional levels (Buhler & Moazed 2007; Filipowicz 2005; Tomari & Zamore 2005). The mainly germline-expressed Piwi proteins bind to piRNAs, and are responsible for the germline development, gametogenesis and transposon silencing (Aravin *et al.* 2006; Girard *et al.* 2006; Grivna *et al.* 2006a; Lau *et al.* 2006; Watanabe *et al.* 2006) (see Figure 1 A, red indicated Piwi proteins).

The differentiation between the AGO and the Piwi subfamily is based on the similarity of the proteins to either the AGO proteins of *Arabidopsis* or to *Drosophila* PIWI. While the AGO proteins have been found in nearly all eukaryotes, Piwi proteins are restricted to animals and ciliates; organisms which have a sexual reproduction cycle (Cox *et al.* 1998). While the PAZ, MID and PIWI domains of Piwi and AGO are conserved, they sequentially differ mainly at their N-termini (Cox *et al.* 1998).

In the next part, the most studied miRNA and siRNA pathway shall be highlighted before the main focus of this thesis, the piRNA pathway, is described.

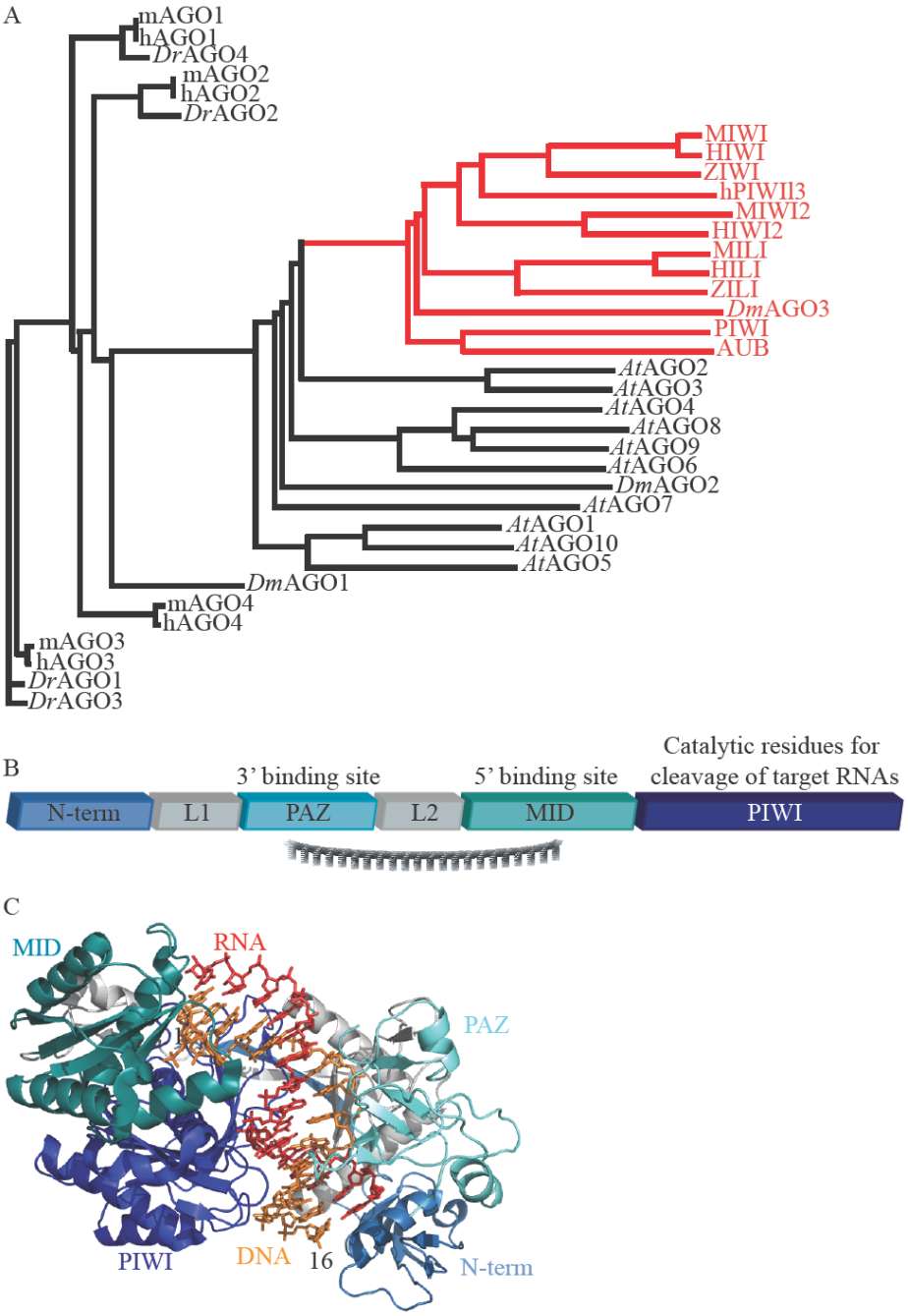


Figure 1: Domains of the Argonaute proteins

(A) Argonaute proteins of *Arabidopsis thaliana*, mouse, human, zebrafish and *Drosophila melanogaster* were aligned and used to create a phylogram with ClustalW (Chenna *et al.* 2003).

(B) The Argonaute proteins contain an N-terminal domain followed by the first linker (L1). In the centre is the PAZ domain, which binds the 3` end of small RNAs. The PAZ domain and the MID domain are separated by the second linker region (L2). The MID domain binds the 5` end of the small RNAs and orients it towards the PIWI domain, which contains catalytic residues (DDH) for target RNA cleavage.

(C) The 2.6 Å crystal structure of the Ago ternary complex of *Thermus thermophilus* Ago by Wang and colleagues. The N-terminus is indicated in sky-blue, the PAZ domain in cyan, the Mid domain in green-blue and the Piwi domain in dark blue. The linkers (L1 and L2) are displayed in grey. The Ago protein was co-crystallized with a bound 21-nucleotide guide DNA (orange) and a 19-nucleotide target RNA (red) (Wang *et al.* 2009b)

1.3.3 The miRNA pathway

Most miRNAs are evolutionally conserved and encoded as isolated genes or in closely spaced clusters of 2-7 genes, from intergenic or exonic or intronic regions (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee & Ambros 2001; Mourelatos et al. 2002). The canonical biogenesis of miRNAs in animals starts in the nucleus, where POL II transcribes a precursor, containing a 5` terminal 7-methyl guanylate cap and a 3` poly(A) tail, as for mRNAs (Cai et al. 2004; Lee et al. 2004a) (see Figure 2 A). This primary miRNA (pri-miRNA) (Lee *et al.* 2002) can encode several miRNAs (Lagos-Quintana *et al.* 2001; Lau *et al.* 2001; Mourelatos *et al.* 2002). The pri-miRNAs are cut into an intermediate form, the 60-70 nt pre-miRNA (Lagos-Quintana *et al.* 2001; Lee & Ambros 2001; Lee *et al.* 1993; Mourelatos *et al.* 2002), by the microprocessor protein complex. This contains the RNaseIII endonuclease Drosha (Lee *et al.* 2003) and the dsRNA binding protein Pascha (partner of Drosha) in flies and *C. elegans* (Denli *et al.* 2004) or DGCR8 (human DiGeorge Syndrome Critical Region Gene 8) in mammals (Gregory *et al.* 2004; Han *et al.* 2004a; Landthaler *et al.* 2004; Wang *et al.* 2007). Pasha/DGCR8 directly interacts with the stem of the pri-miRNA and serves as molecular anchor and positioning guide for Drosha (Han *et al.* 2006). Drosha creates the so called pre-miRNA, a hairpin structure, where the loop is flanked by base-paired arms that form a stem, which contain some unpaired nucleotides (Han *et al.* 2006; Lagos-Quintana *et al.* 2001; Lau *et al.* 2001; Mourelatos *et al.* 2002). The pre-miRNA stem carries a two nucleotide overhang at their 3` end, typical after the action of an RNase III enzyme, and a phosphate group at their 5` end (Lee *et al.* 2003).

The canonical miRNA pathway gives rise to most of the known miRNAs, but some miRNAs are produced by nuclear pre-mRNA splicing. This gives rise to pre-miRNA introns, which mimic the structure of pre-miRNAs, and are called mirtrons (Babiarz *et al.* 2008; Berezikov *et al.* 2007; Glazov *et al.* 2008; Okamura *et al.* 2007; Ruby *et al.* 2007) (see Figure 2 b). The spliced introns accumulate first as lariat shaped products and are then 2`-5` debranched by the lariat debranching enzyme (Ldbr). These products contain the 2 nucleotide 3` overhangs, depending on the splicing site and can enter the standard miRNA biogenesis pathway, where the 3` arm of the hairpin gives rise to mature miRNAs (Okamura *et al.* 2007; Ruby *et al.* 2007). Not all mirtrons define their 3` end by splicing and carry a longer 3` tail

(Ruby *et al.* 2007). In such cases the 3` end undergoes a trimming event in the nucleus that is mediated by the exosome 11 complex via Rrp6 (Flynt *et al.* 2010).

Another group of miRNAs in mouse embryonic stem cells derive from endogenously expressed short hairpins (shRNAs). These shRNAs are probably transcribed by POL III and are DGCR8 independent (Babiarz *et al.* 2008) (see Figure 2 c). Finally, some other sources of miRNAs are the small nucleolar RNAs (snoRNAs), the transfer RNAs (tRNAs), as well as the terminal hairpins of endo-siRNAs long stem loop precursors that can bypass the Drosha step (Cole *et al.* 2009; Ender *et al.* 2008; Miyoshi *et al.* 2010; Saraiya & Wang 2008).

The pre-miRNAs from the different pathways are exported from the nucleus to the cytoplasm by the Ran guanosine triphosphate (RanGTP)-dependent dsRNA-binding protein Exportin5 shuttle system (Bohnsack *et al.* 2004; Lund *et al.* 2004; Yi *et al.* 2003). Exportin5 recognizes the 3` overhangs of the pre-miRNAs and binds directly to the stem of at least 16 nt. Additionally, the binding of the RanGTP-Exportin5 complex to the pre-miRNA protects the RNA from exonucleolytic degradation in the nucleus (Okada *et al.* 2009; Zeng & Cullen 2004).

Once in the cytoplasm, the pre-miRNAs are cleaved by Dicer into dsRNA duplexes of 21-25 nt with 2nt overhangs at the 3` end (Bernstein *et al.* 2001; Grishok *et al.* 2001; Hutvagner *et al.* 2001; Ketting *et al.* 2001; Knight & Bass 2001; Lee & Ambros 2001) (Figure 2). Dicer prefers to cleave with an Uridine at the 5` position, which results in an 5`U preference in mature miRNAs (Aravin *et al.* 2003). Dicer is accompanied by its dsRNA binding protein TRBP (human immunodeficiency virus (HIV-1) trans-activating response (TAR) RNA-binding protein) in mammals or the BP isoforms of Loqs in flies (Forstemann *et al.* 2005; Haase *et al.* 2005; Jiang *et al.* 2005; Saito *et al.* 2005). TRBP and Loqs enhance the processing of pre-miRNAs to miRNA duplexes by increasing the substrate affinity of Dicer and stimulating the association of the complex with the Argonaute proteins (Chendrimada *et al.* 2005; Haase *et al.* 2005; Jiang *et al.* 2005; Saito *et al.* 2005).

Recently it was shown by two groups that a miRNA pathway exists, which requires Drosha, but not Dicer. Here the pre-miRNAs are directly loaded into AGO2. The AGO protein itself then cleaves the pre-miRNAs into miRNA (Cheloufi *et al.* 2010; Cifuentes *et al.* 2010) with a defined 5`end. The 3` end remains variable, because it is then uridilated and afterwards trimmed to create a functional miRNA (Cifuentes *et al.* 2010).

The miRNA duplex, created by Dicer, contains one miRNA (guide strand) and one passenger strand (miRNA*), corresponding to the two sides of the pre-miRNA stem. Only one strand of the duplex is incorporated into an Argonaute protein (Grishok *et al.* 2001; Mourelatos *et al.* 2002). The passenger strand is thought to be degraded. Only in some cases are both strands used to produce mature miRNAs (Ruby *et al.* 2006). The sorting is carried out by sensing the relative thermodynamic stability of the 5` ends of each strand to determine the guide strand, or the strand which directs the 5` silencing. The 5' end of the actual miRNA strand is less tightly paired to the complementary strand (Khvorova *et al.* 2003; Schwarz *et al.* 2003).

In most organisms, miRNAs are sorted to associate with specific Argonautes. For example, although that *Drosophila* contains two members of the Ago protein clade - AGO1 and AGO2-, the miRNAs are actively sorted into AGO1. This sorting of miRNAs into AGO1 requires DCR-1 and Loqs. In contrast, the association of siRNAs into AGO2 requires DCR-2/R2D2, which are components of the siRNA pathway. If the miRNA duplex has central mismatches, their binding to the DCR-2/R2D2 heterodimer is reduced and the small RNA is loaded in AGO1 (Forstemann *et al.* 2007; Tomari *et al.* 2007). Additionally, the different AGO associations of small RNAs itself is dependent on specific mismatches of the duplex. While miRNAs often show a bulge in the duplex at the positions 9 and 10, Watson-Crick base paring at position 9 and 10 promote miRNA* binding to AGO2. In *Drosophila*, the miRNA* loaded into AGO2 has a 3` end modified by the RNA methyltransferase HUA ENHANCER1 and behaves similar to siRNAs. The identity of the 5` nucleotide affects the sorting as well, but is not sufficient for the recognition by specific Ago proteins (Czech *et al.* 2009; Okamura *et al.* 2009).

In mammals the sorting of the miRNA/miRNA* is less clear. All four AGO proteins (hAGO1-4) in human show similar a preference towards the structure of the RNA duplexes in requiring a central mismatch for RISC loading (Yoda *et al.* 2010). In different tissues at different developmental stages, the use of miRNA and miRNA* strands can vary to a large degree (Hu *et al.* 2009; Landgraf *et al.* 2007). In mice AGO2 is the dominant miRNA partner. Although, mutant analysis showed, that the different AGO proteins (mouse AGO 1, 3 and 4) in the organism can substitute for AGO2 (Liu *et al.* 2004).

Plants lack a Drosha homolog, hence production of miRNAs in plants require cleavage by the nuclear Dicer-like 1 (DCL1) (see Figure 2). After POL II transcription of the pri-miRNA, DAWDLE (DDL), a FHA domain-containing RNA-binding protein, binds to the pri-miRNA and facilitates the binding of DCL1 (Yu *et al.* 2008). DCL1 is accompanied by the dsRNA binding protein HYL1 (Han *et al.* 2004b; Vazquez *et al.* 2004a), which binds to the Cys_2His_2 zinc finger protein SERRATE (SE). Together these proteins position DCL1 on the miRNA transcript precursor for correct processing (Kurihara *et al.* 2006; Yang *et al.* 2006a). Here, the pri-miRNA is first cleaved into the pre-miRNA (70-190 nt), and is then cut into the miRNA duplex (Kurihara & Watanabe 2004; Papp *et al.* 2003; Park *et al.* 2002; Reinhart *et al.* 2002; Xie *et al.* 2004). A putative mirtron discovered in rice grains, indicates that plants may use spliced introns to produce miRNAs similar to animals (Zhu *et al.* 2008).

Unlike animal miRNAs, both miRNA strands in plants are 2`-O-methylated at their 3` end by the RNA methyltransferase HEN1 (Park *et al.* 2002; Yang *et al.* 2006b; Yu *et al.* 2005). This protects the miRNAs from oligo-uridylation, a signal for degradation (Li *et al.* 2005). The miRNAs are exported by the Exportin5 homolog HASTY into the cytoplasm (Park *et al.* 2005; Peragine *et al.* 2004). The addition of the methyl group on the miRNAs could happen before loading into AGO1, as both strands are methylated (Yu *et al.* 2005). However it is not yet clear, if the methylation of the miRNAs and the binding to AGO1 is localized in the nucleus or in the cytoplasm, given that HEN1 and AGO1 localize in both compartments (Fang & Spector 2007).

In animals and plants the mechanism by which miRNAs regulate their targets is dependent on the Argonaute proteins they are bound to and the degree of complementarity the miRNA have to their target mRNA (Hutvagner & Zamore 2002; Liu *et al.* 2004; Meister *et al.* 2004). Although rare cases of miRNA target sites with nearly full complementarity to miRNA exist in animals (Davis *et al.* 2005; Hutvagner & Zamore 2002; Mansfield *et al.* 2004; Pfeffer *et al.* 2004; Song *et al.* 2004; Sullivan *et al.* 2005; Yekta *et al.* 2004), the target recognition by partial complementarity is more common in metazoan (Bartel 2009; Brennecke *et al.* 2005; Lai 2004; Lewis *et al.* 2005; Lewis *et al.* 2003; Stark *et al.* 2005; Xie *et al.* 2005) (see Figure 2). Therefore, the nucleotides 2-8 of miRNA, called the miRNA ‘seed’, guide the miRNA and Argonaut-containing miRISC to its targets by Watson-Crick base pairing. The relatively small size of the seed sequence provokes the possibility that one miRNA can address many different targets (Baek *et al.* 2008; Brennecke *et al.* 2005; Krek *et*

al. 2005; Lai 2004; Lewis *et al.* 2005; Lewis *et al.* 2003; Lim *et al.* 2005; Selbach *et al.* 2008; Stark *et al.* 2005; Xie *et al.* 2005). The 'seed' sequence contributes the most to the thermodynamic stability of the target binding (Ameres *et al.* 2007; Haley & Zamore 2004). So if mismatches occur in the seed sequence, then the intermolecular interaction between the miRNA and the AGO protein is disturbed (Parker *et al.* 2009; Wang *et al.* 2008a). The target sites are preferentially located in an AU-rich region and tend to be located towards the beginning or end of long 3' UTRs (Gaidatzis *et al.* 2007; Grimson *et al.* 2007). When several target sites are within one transcript, the miRNA guided miRISC acts independently (Doench *et al.* 2003), but if the target sites are within a short range they can act synergetically (Grimson *et al.* 2007). In general, the binding of the miRISC to a target site with partial complementarity leads to a translational block, a mRNA destabilization or both (Lee *et al.* 1993; Lim *et al.* 2005; Pillai *et al.* 2005; Wightman *et al.* 1993).

In plants a perfect or near-perfect complementarity to the mRNA target site is the norm. In most cases this leads to a cleavage of the target sequence, the so called Slicing reaction (Llave *et al.* 2002; Rhoades *et al.* 2002; Tang *et al.* 2003) (see Figure 2). It was shown that AGO1 has slicer activity (Baumberger & Baulcombe 2005). But AGO1 or AGO10 bound to miRNAs can also block translation, similar to animal miRNAs (Brodersen *et al.* 2008).

In general, miRISCs associated with translational blocked RNAs, are guided by the Glycine-tryptophan-rich 182 (GW182) protein to small cytoplasmic foci called processing (P)- bodies. Here the mRNA can be degraded by the decapping and deadenylation enzymes or stored for later release at a specific time point depending on cellular requirements (Behm-Ansmant *et al.* 2006; Eulalio *et al.* 2008; Liu *et al.* 2005a; Liu *et al.* 2005b; Pillai *et al.* 2005; Sen & Blau 2005).

This mechanism of miRNA-guided gene expression regulation is important in a broad variety of biological processes including development (Brennecke *et al.* 2003; Giraldez *et al.* 2005; Giraldez *et al.* 2006; Lee *et al.* 1993; Reinhart *et al.* 2000), cell cycle regulation (Linsley *et al.* 2007), metabolism (Krutzfeldt *et al.* 2005; Poy *et al.* 2004; Xu *et al.* 2003), immunoresponse effects (Azuma-Mukai *et al.* 2008; Yeung *et al.* 2009) and diseases such as cancer (Esquela-Kerscher & Slack 2006; Lu *et al.* 2005).

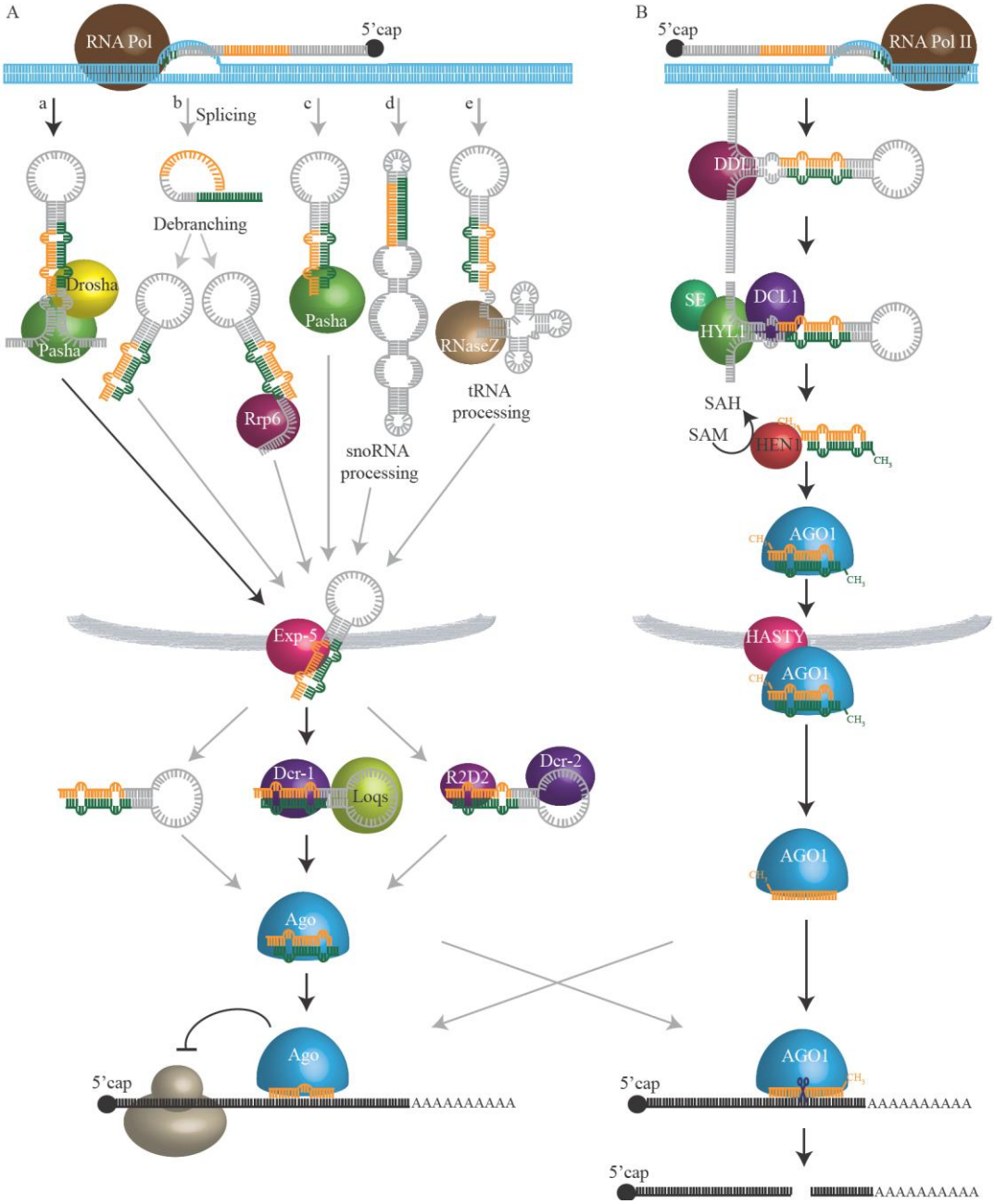


Figure 2: miRNA pathways

(A) miRNAs of animals can derive from (a) pri-miRNA which is processed into pre-miRNAs by the Drosha/Pasha (green/yellow) complex, (b) mitrons, which are processed by debranching enzymes and cleaved by Rrp6 (red) if they have a 3' tail, (c) hairpins, which are Pasha dependent, (d) snoRNA precursors or (e) tRNA precursors or virus induced tRNA like structures requiring specific processing enzymes. The pre-miRNA is exported into the nucleus by Exportin5 (pink) and directly associated with an Ago protein (blue), or cleaved by the Dcr-1/Loqs (purple/green) complex or the Dcr-2/R2D2 (purple/lilac) complex and then bound to an Ago protein. The miRISC can then perform translational inhibition and RNA destabilization (left side) or mRNA cleavage (right side). The main path of miRNAs is indicated with black arrows.

(B) The plant miRNA pathway. The pri-miRNA is bound by DDL1 (dark red), which recruits DCL1 (purple), HYL1 (yellow-green) and SE (green) to free the miRNA duplex. miRNAs may derive from mirtron in plants, too. However, it is not yet clear if the methylation of the miRNAs by HEN1 (red) and the binding of to AGO1 (blue) are located in the nucleus or cytoplasm. The miRNAs are exported by HASTY (pink). The most common effect by the plant miRISC is the cleavage of its target mRNA (right side) but translation inhibition has also been reported.

1.3.4 The siRNA pathway

Initially, the siRNA pathway was thought only to act as a defense mechanism for the cell to protect against selfish and invasive RNA elements. This was mainly due to the fact that disruptions of the pathway in *D. melanogaster* and *C. elegans* did not cause an obvious phenotype, except for a higher sensitivity to viral infections and increased activity of repetitive elements (Ding & Voinnet 2007; Lee *et al.* 2004b; Okamura *et al.* 2004; Tabara *et al.* 1999; Tabara *et al.* 2002).

A broad variety of endo-siRNAs have been described, deriving from hairpins formed by inverted repeats, natural antisense siRNAs (natsiRNAs) from complementary transcripts and secondary siRNAs, which are generated by RNA-dependent RNA polymerases (RdRPs). In plants, the predominant endo-siRNAs are cis-acting siRNAs (casiRNAs; 24 nt). They are transcribed from transposons, repetitive elements and tandem repeats (Chan *et al.* 2004; El-Shami *et al.* 2007; Herr *et al.* 2005; Kanoh *et al.* 2005; Onodera *et al.* 2005; Pontier *et al.* 2005; Xie *et al.* 2004; Zilberman *et al.* 2003). They function in directing DNA methylation and histone modification to silence the loci they originate from (Chan 2008; Chan *et al.* 2004; Llave *et al.* 2002; Mette *et al.* 2000; Tran *et al.* 2005; Zilberman *et al.* 2003).

Trans-acting siRNAs (tasiRNAs) are also an example of endo-siRNAs in plants, and of how the mi- and siRNA pathway can overlap: After miRNAs have directed the cleavage of certain transcripts, the cleaved single stranded 3’ RNA fragment (and a lesser extent the 5’ fragment) can be used to produce siRNAs (Allen *et al.* 2005; Peragine *et al.* 2004; Vazquez *et al.* 2004b; Williams *et al.* 2005; Yoshikawa *et al.* 2005).

Endo-siRNAs are also produced in response to stress, for example from salt stress (Borsani *et al.* 2005) and in response to bacterial pathogen effectors (Katiyar-Agarwal *et al.* 2006). Natural antisense transcript-derived siRNAs (natsiRNA) are produced from pairs of RNAs, were one strand is expressed constitutively and the antisense transcript in stress conditions only (Borsani *et al.* 2005; Katiyar-Agarwal *et al.* 2006).

While endo-siRNAs are well described in plants (Vazquez 2006) and fungi (Catalanotto *et al.* 2002; Reinhart & Bartel 2002), they were only recently discovered in animals. While plant and worm endo-siRNAs are typically produced depending on the RNA-

dependent RNA polymerases action, *D. melanogaster*, mouse and human lack this kind of RdRP (Ghildiyal & Zamore 2009). In cultured human cells, it was discovered that the full-length LINE1 (long interspersed nuclear element1) contain a sense and antisense promoter in its 5` untranslated region (5`UTR). This leads to a bidirectional transcription of an overlapping region, which could be detected by Dicer and processed into siRNAs to silence retrotransposable activity. However, the exact mechanism remains undetermined (Yang & Kazazian 2006). More recently endo-siRNAs in human have been shown to be required for chromatin remodeling in the vicinity of the original siRNA target site, such as H3K9me2 and H3K27me3 (Malecova & Morris 2010).

In flies endogenous siRNAs have been detected in somatic and germ cells. They are 21 nt in length, present in sense and antisense orientations and have 2` O methylated 3`ends. One group of endogenous siRNAs derive from transposons and another from heterochromatic and intergenic sequences, building long extensive structured transcripts like hairpins and mRNA (Chung *et al.* 2008; Czech *et al.* 2008; Ghildiyal *et al.* 2008; Hartig *et al.* 2009; Kawamura *et al.* 2008; Okamura *et al.* 2008a; Okamura *et al.* 2008b).

In mice, endo-siRNAs have been identified in oocytes. As for flies, they are 21 nt in length and originate from a variety of genomic clusters, forming dsRNAs by spliced transcripts of protein-coding genes to homologous antisense transcripts of pseudogenes or an inverted repeat of pseudogenes (Sasidharan & Gerstein 2008; Tam *et al.* 2008; Watanabe *et al.* 2008). Additionally, in embryonic stem cells endo-siRNAs are produced via Dicer-independent cleavage of long hairpins, which mainly derive from genomic loci with tandem inverted short intersperse nuclear elements (SINEs) (Babiarz *et al.* 2008).

Whether the siRNAs originate from exogenous substrates or endogenous transcripts, the siRNA biogenesis begins with a double stranded long RNA. This is cleaved by the dsRNA-specific Ribonuclease III (RNaseIII) Dicer (Bernstein *et al.* 2001; Elbashir *et al.* 2001b; Hammond *et al.* 2000) into a 21 to 24 nt long RNA duplex (Elbashir *et al.* 2001b; Zamore *et al.* 2000), with a two nt overhang at the 3` ends (Elbashir *et al.* 2001b; Elbashir *et al.* 2001c) (see Figure 3). Each strand has a 5` phosphate and a 3` hydroxyl group (Elbashir *et al.* 2001b). The siRNA duplex is then incorporated into the RNA-induced silencing complex (RISC). This occurs via a RISC loading complex (RLC) (see Figure 3), where the relative thermodynamic stability of the 5` ends of each strand is sensed to select the guide strand (Aza-Blanc *et al.* 2003; Khvorova *et al.* 2003; Schwarz *et al.* 2003). In *Drosophila*, the

dsRNA-binding protein R2D2 (two dsRNA-binding domains, associated with Dicer 2), a component of the RLC and a partner of Dicer-2 (DCR-2), is responsible for the recognition of the thermodynamically more stable 5` end of the passenger strand (Liu *et al.* 2003; Tomari *et al.* 2004). The RLC then recruits an AGO protein, for example Agonaute2 (AGO2) in *Drosophila*, which cleaves the passenger strand and releases it (Kim *et al.* 2007; Leuschner *et al.* 2006; Matranga *et al.* 2005; Miyoshi *et al.* 2005; Rand *et al.* 2005). This siRNA mediated cleavage (like the target cleavage) always happens across the phosphodiester bond between the 10^{th} and 11^{th} nucleotide of the guide strand (Elbashir *et al.* 2001c). The cleavage is Mg^{2+} dependent and leads to a 3` hydroxyl and a 5` phosphate group (Martinez & Tuschl 2004; Schwarz *et al.* 2004; Wang *et al.* 2009b). The RNA is now mature in higher animals, but not in flies and plants, where the methyltransferase HEN1 methylates the 3` end of the siRNAs to stabilize the siRNAs and protect them from degradation (Boutet *et al.* 2003; Chen 2007; Ebhardt *et al.* 2005; Li *et al.* 2005; Yang *et al.* 2006b). Following this, the siRNA can guide the RISC to target mRNAs with perfect complementarity, cleaving the mRNA and leading to their degradation because they lack either a cap or poly(A) tail (Elbashir *et al.* 2001a; Elbashir *et al.* 2001b; Elbashir *et al.* 2001c; Hammond *et al.* 2000; Zamore *et al.* 2000).

In plants and flies, the siRNA pathway is partly a defense mechanism against viral infections, but as well a gene regulatory element (Ghildiyal & Zamore 2009; Obbard *et al.* 2006). However, there is yet no evidence that mammals use the RNAi pathway for viral defense, instead they have developed an interferon induced protein based immune-system (Vilcek 2006; Williams 1999).

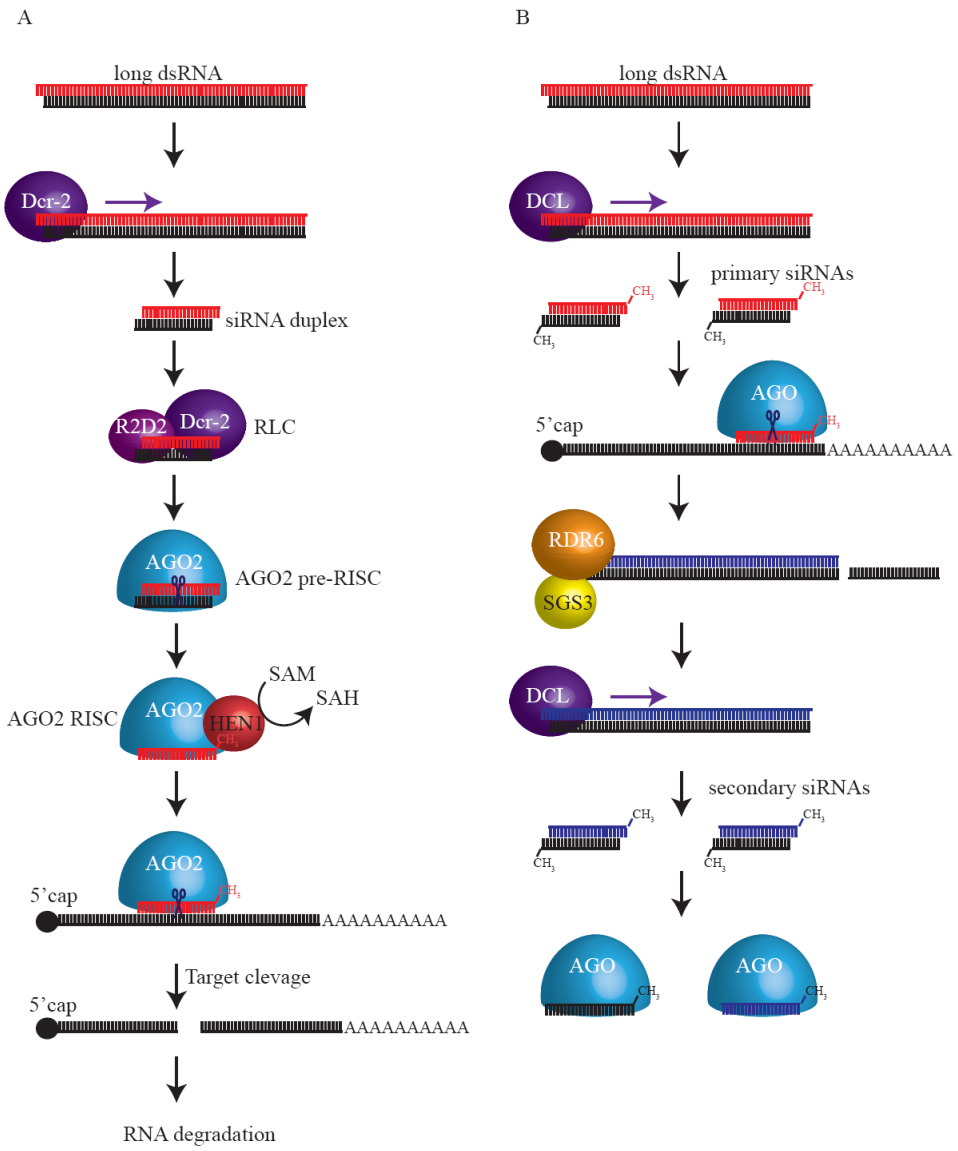


Figure 3: siRNA pathway

(A) siRNAs in insects (the *Drosophila melanogaster* pathway is shown here as an example) starts with a long double stranded RNA, which is cleaved by a Dicer (purple) into ~21 nt siRNA duplexes. The R2D2 protein (lilac) binds to the thermodynamically stable 5` end and together with Dicer the RNA is guided to AGO2 (blue) to form the RISC. The passenger strand (black) is cleaved by the Argonaute enzyme and the guide strand (red) is 2`-O-methylated at the 3` end. The RISC complex cleaves its target mRNA, guided by the siRNA. The mRNA fragments are degraded afterwards.

(B) Amplification of siRNAs in plants: A plant Dicer cuts the primary siRNAs out of a long dsRNA. This siRNAs bind to an Argonaute protein, which cleaves the target mRNA. This can recruit RDR6 with the RNA-binding protein SGS3 (yellow), which leads to a synthesis of a dsRNA. This dsRNA is then cleaved by DCL4 to create secondary siRNAs. These are again bound by AGO. All siRNAs in plants are 2'O methylated. Adapted from (Ghildiyal & Zamore 2009)

1.3.5 Piwi-interacting RNAs, a germline-specific small RNA class

Piwi-interacting RNAs (piRNAs) are the largest class of small RNA molecules expressed in the animal cell (Seto *et al.* 2007), but they were not widely recognized as such until 2006, when five independent research groups simultaneously re-discovered these germline-specific ncRNAs in mice and *Drosophila* (Aravin *et al.* 2006; Girard *et al.* 2006; Grivna *et al.* 2006a; Lau *et al.* 2006; Watanabe *et al.* 2006). Actually, similar ncRNAs were described in *Drosophila* as repeat-associated RNAs (rasiRNAs) (Aravin *et al.* 2003) already in 2003. RasiRNAs have the same features as piRNAs (Saito *et al.* 2006; Vagin *et al.* 2006) and will therefore be referred to as piRNAs hereafter.

piRNAs derive from discrete gene-poor genomic loci (Aravin *et al.* 2006; Girard *et al.* 2006; Grivna *et al.* 2006a; Lau *et al.* 2006; Watanabe *et al.* 2006). Different from miRNAs, whose sequences are conserved across species, no sequence conservation was found for piRNAs (Aravin *et al.* 2006; Girard *et al.* 2006; Lau *et al.* 2006). Many of the larger piRNA clusters are syntenic and range in size from 0.9 to 127 kb (Aravin *et al.* 2006), as well as encoding many homologous piRNAs with overlapping sequences (Ro *et al.* 2007). Most of the piRNAs in mammals derive from uni-directional clusters, but a few bidirectional clusters have been identified (Aravin *et al.* 2006; Girard *et al.* 2006; Lau *et al.* 2006; Ro *et al.* 2007; Watanabe *et al.* 2006).

In *Drosophila* these piRNA clusters derive mainly from transposons, transposon remnants and repeat sequences (Brennecke *et al.* 2007; Gunawardane *et al.* 2007; Saito *et al.* 2006; Yin & Lin 2007). In mammals piRNAs contain exonic, intronic, intergenic and repeat sequences, but the majority are derived from intronic regions (Grivna *et al.* 2006a; Ro *et al.* 2007) and only 17 % of the adult mouse piRNAs map to repeated elements (Aravin *et al.* 2006; Girard *et al.* 2006; Ro *et al.* 2007). Mice piRNAs clusters expressed before birth can generate piRNAs from both strands, as in *Drosophila*, while postnatal piRNAs do not show this feature (single-strand clusters) (Aravin *et al.* 2008).

Interestingly, Piwi proteins show a clear preference for different originating piRNAs during germline development. *Drosophila* contains three Piwi proteins; AGO3, PIWI and AUBERGINE (AUB). AGO3 show a bias for piRNAs encoded on the sense strand of

transposons, while PIWI and AUB preferentially bind piRNAs from the antisense strand (Brennecke *et al.* 2007; Gunawardane *et al.* 2007).

In mice, three Piwi proteins -MIWI, MILI and MIWI2- are present exclusively in testes. MILI is expressed from 12.5 days post coitum (dpc) until round spermatids (20 days post partum, dpp). MIWI2 is found from 15.5 dpc until 3 days after birth and MIWI is expressed from 14 dpp until the round spermatids stage (see Figure 4). MILI binds ~26 nt piRNAs (Aravin *et al.* 2006; Aravin *et al.* 2007b), MIWI2 ~28 nt piRNAs (Aravin *et al.* 2008) and MIWI associates with ~30 nt piRNAs (Girard *et al.* 2006; Grivna *et al.* 2006a; Grivna *et al.* 2006b). At 16.5 dpc most (70%) of the piRNAs are derived from transposon sequences. Between 16.5 dpc and 10 dpp the amount of transposon derived piRNAs is ~35% in MILI. The piRNA profile changes during development, with those derived from long interspersed nuclear elements (LINE) and long terminal repeats (LTR) decreasing as the small interspersed nuclear elements (SINE) and exon originating ones increase. MILI and MIWI2 associate with repeat derived piRNAs during the embryonic development, whereas the piRNA profile of MILI changes towards post-natal stages (Aravin *et al.* 2008; Kuramochi-Miyagawa *et al.* 2008). Prenatal piRNAs show complementarities between MILI- and MIWI2-associated piRNAs. MILI piRNAs have a bias for an U at the first nt (Aravin *et al.* 2006; Aravin *et al.* 2008; Kuramochi-Miyagawa *et al.* 2008; Watanabe *et al.* 2006) and derive from the sense strand of clusters and/or transposons. This suggests they arise from cleavage of a genomically encoded piRNA cluster. MIWI2 has the signature of a preference for an A at the 10^{th} nt and a strong bias for antisense strand transposon sequences. A small proportion of the MILI- and MIWI2-associated repeat sequences also show a 10-nt overlap (Aravin *et al.* 2008; Kuramochi-Miyagawa *et al.* 2008).

From the pachytene stage onwards, MIWI and MILI bind to non-repetitive piRNAs of intergenic origin and of unknown function (Aravin *et al.* 2006; Girard *et al.* 2006). The MIWI-associated piRNAs show a bias for Uracil at the 5` end like MILI piRNAs (Girard *et al.* 2006; Watanabe *et al.* 2006).

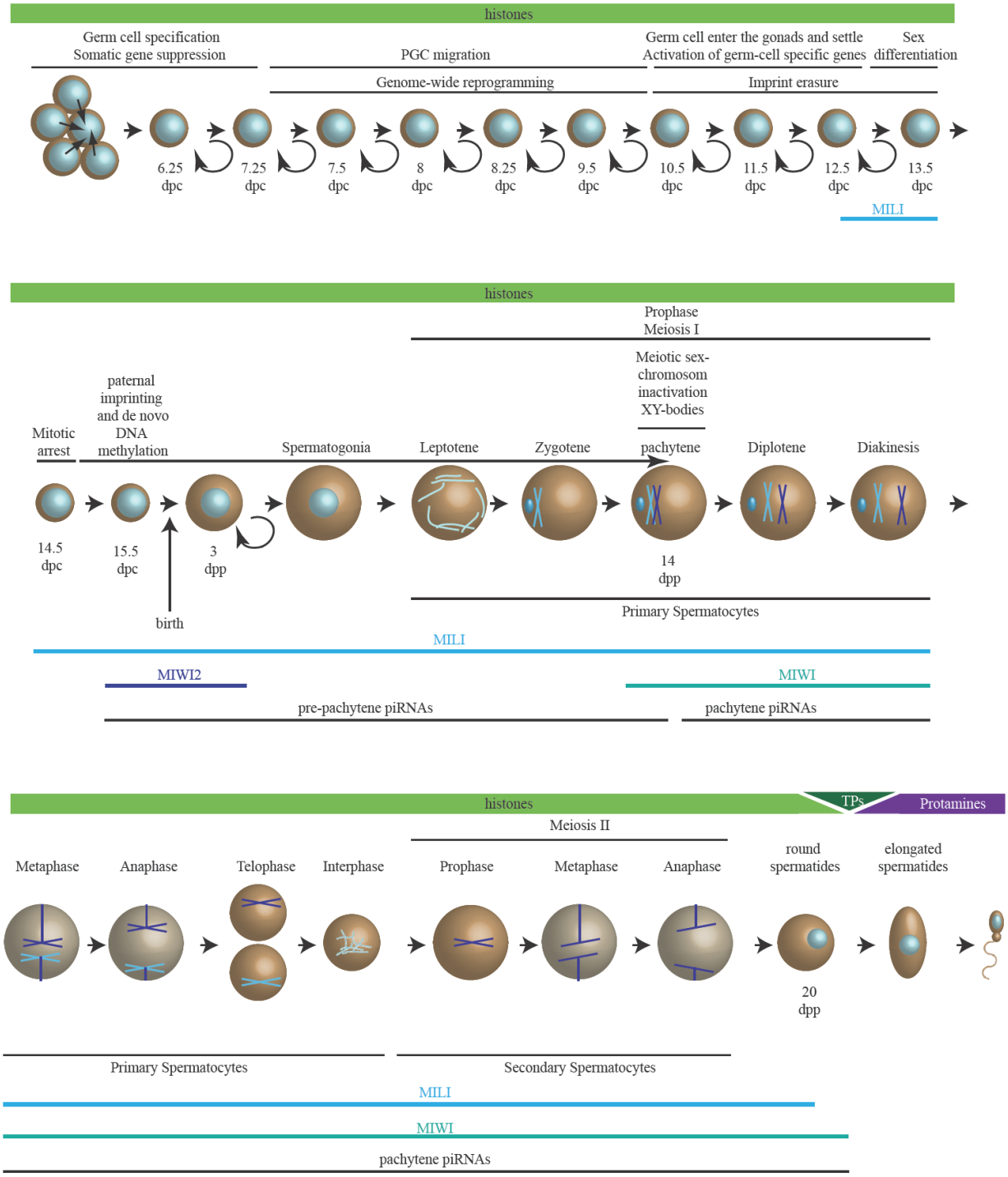


Figure 4: Germ cell development and spermatogenesis in mouse
7.25 dpc the primordial germ cells appear. Between 7.5 and 13.5 dpc genome wide reprogramming takes place before the cells enter mitotic arrest. The expression of piRNAs and Piwi proteins are indicated in blue lines. MILI is expressed from 12.5 dpc until round spermatids, MIWI2 from 15.5 dpc until 3 days after birth and MIWI from 14 dpp until the round spermatids stage.

In the zebrafish (*Danio rerio*) ZIWI and ZILI-associated piRNAs (26-27 nt) are found in the male and female germline (Houwing *et al.* 2008; Houwing *et al.* 2007). The fish piRNAs originate mainly from transposons (mostly LTR elements) and repetitive sequences, the rest are from non-repetitive sequences that lie in intergenic regions. Unlike mammalian, zebrafish piRNAs are derived from discrete genomic loci where they display a strong strand asymmetry. The piRNA density is much higher in ovaries than in testis (Houwing *et al.* 2008; Houwing *et al.* 2007). Another difference is that piRNAs in fish can be maternally passed to the embryo (Houwing *et al.* 2007). Similar to mice, the piRNAs of ZIWI have the 5` U bias while those in ZILI has predominantly A at the 10 nt (Houwing *et al.* 2008)

So far, all the piRNAs analyzed from different species have a 5` monophosphate and a 3` hydroxyl group (Houwing *et al.* 2007; Saito *et al.* 2006; Watanabe *et al.* 2006) that is modified (Gunawardane *et al.* 2007; Houwing *et al.* 2007; Kawaoka *et al.* 2009; Vagin *et al.* 2006). This modification was shown to be a 2`-O-methylation (Horwich *et al.* 2007; Kirino & Mourelatos 2007c; Ohara *et al.* 2007a; Saito *et al.* 2007a; Zhou *et al.* 2010). The piRNAs share this 3' end modification with plant mi- and siRNAs (Akbergenov *et al.* 2006; Katiyar-Agarwal *et al.* 2007; Li *et al.* 2005; Yang *et al.* 2006a; Yang *et al.* 2006b; Yu *et al.* 2005) as well as the siRNAs and hairpin RNAs (hpRNAs) from *Drosophila* (Horwich *et al.* 2007; Okamura *et al.* 2008b; Pelisson *et al.* 2007; Saito *et al.* 2007a; Vagin *et al.* 2006). Similar modification of animal miRNAs or siRNAs in other organisms is not yet described..

To summarize, piRNAs (24 to 31 nt in length) are not conserved between species and are modified at their 3` end with a methyl group. Many piRNAs have a U-bias at the 5` end or an A at the 10^{th} position and an overlap between these piRNAs is suggested. These features of piRNAs may be due to their biogenesis and are discussed in the next part.

1.3.6 piRNA biogenesis

The piRNA pathway is evolutionary ancient (Grimson *et al.* 2008) and there is a great interest in knowing the exact molecular pathway that leads to piRNA biogenesis. piRNAs are not equally distributed over the genome, but rather clustered at defined loci (Aravin *et al.* 2006; Grivna *et al.* 2006a), where 96 % of the piRNA sequences are present in a few hundred genomic sites (Aravin *et al.* 2006; Girard *et al.* 2006). It is suggested that a long single

stranded RNA or a pair of divergently transcribed large single stranded RNAs, from one cluster, serve as a precursor for piRNAs (Aravin *et al.* 2006; Lau *et al.* 2006; Watanabe *et al.* 2006). Given that many piRNAs are generated from a single cluster simultaneously, an individual transcription for each piRNA is unlikely. This idea is supported by the observation that P-element insertions into putative promoter of piRNA loci disrupt the production of all piRNAs from a 180 kb window (Brennecke *et al.* 2007; Kawaoka *et al.* 2008; Malone *et al.* 2009).

The primary processing of piRNAs is not well understood. Factors of this process are only just starting to be discovered. This is mainly from research in flies due to the ease of genetic manipulation. Biochemical studies in mouse characterized the orthologues of the fly proteins and identified new factors.

One factor in *Drosophila* is the putative nuclease Zucchini, which mainly localizes in a perinuclear region of the cytoplasm, called the nuage, in gonad-associated somatic and germline cells (Pane *et al.* 2007; Saito *et al.* 2009). Zucchini was found to interact physically with AUB (Pane *et al.* 2007).The loss of *zucchini* function leads to a reduction of processed piRNAs in somatic as well as germline cells, but does not affect the signatures of the secondary piRNA pathway (Malone *et al.* 2009; Olivieri *et al.* 2010; Pane *et al.* 2007; Saito *et al.* 2009).

Among the factors of the primary pathway identified so far are several putative helicases, Vasa, Spindle-E and Armitage in *Drosophila* and their mouse homologs Vasa (MVH) and MOV10l (Frost *et al.* 2010; Kuramochi-Miyagawa *et al.* 2004; Kuramochi-Miyagawa *et al.* 2010; Malone *et al.* 2009; Olivieri *et al.* 2010; Wang *et al.* 2010; Zheng *et al.* 2010)

Another source of primary piRNAs could be those that are maternally deposited in the *Drosophila* egg. These are mainly bound by AUB and PIWI, and to a lesser extent by AGO3 to prevent transposon activity in the offspring (Brennecke *et al.* 2008). Once the primary piRNAs are present, they are suggested to be used in an amplification loop. This so called “ping-pong” model for piRNA biogenesis is based on sequence analyses of fly piRNA (Brennecke *et al.* 2007; Gunawardane *et al.* 2007) (see Figure 5). The Piwi protein Aubergine (AUB)-containing complex is guided by the primary piRNA to actively transcribed transposons. The transposon sequence is cleaved by AUB, defining the 5` end of a new sense

strand piRNA. This model points out how the 5` ends of new piRNAs are generated from long single-stranded precursors by the slicer action of Piwi proteins, but the mechanism of 3` end formation is less clear (Brennecke *et al.* 2007; Gunawardane *et al.* 2007). It is assumed that an exonuclease or an undefined endonuclease acts on the RNA. Potential candidates were identified in a *Drosophila* genetic screen (Pane *et al.* 2007). The new sense piRNA, a so termed secondary piRNA, because it is generated by Piwi-mediated cleavage, is incorporated into AGO3 RNPs, which target the antisense precursors and lead to the production of more antisense piRNAs to associate with AUB. This mechanism results in an amplification loop of secondary piRNAs (Brennecke *et al.* 2007). PIWI-associated piRNAs show no pairing with AGO3 bound piRNAs (Gunawardane *et al.* 2007), maybe because it is mainly nuclear, while AUB and AGO3 are cytoplasmic localized in the nuage (Brennecke *et al.* 2007; Gunawardane *et al.* 2007; Harris & Macdonald 2001). Therefore it could be speculated that the nuage is the localization center of the ping-pong cycle (Klattenhoff & Theurkauf 2008). Additionally, loss of PIWI has no impact on ping-pong signals, indicating that PIWI does not participate in the amplification cycle, while AUB and AGO3 mutants lack this pathway (Malone *et al.* 2009).

Mutations in *spindle-E*, *vasa*, the Tudor domain-containing protein *krimper* and *armitage* strongly impact on the ping-pong cycle and delocalize AUB and AGO3 from the nuage. Additionally, the piRNA size shifts towards the piRNAs from PIWI. Thus, the Piwi functions in the germline depends on the RNA helicase Armitage, Spindle-E and Vasa, and the Tudor domain protein Krimper. Comparison between Spindle-E and Armitage reveal that Spindle-E is crucially required for the ping-pong cycle. Armitage is only needed for a minority of elements to enter the cycle, so it may not be required for the ping-pong cycle (Malone *et al.* 2009; Olivieri *et al.* 2010). The Heterochromatin protein 1 (HP1) homolog Rhino is not needed for primary piRNA processing but is required for the ping-pong cycle (Olivieri *et al.* 2010). Rhino associates with the 42AB cluster and promotes the production of the RNAs leading to piRNAs from dual-strand heterochromatic clusters, indicating that it could be involved in the processing of primary piRNAs. In *rhino* mutants the ping-pong cycle is impaired and transposon silencing is inefficient. Rhino localizes in the nucleus of germ cells and is responsible for the proper localization of AUB, AGO3 and Vasa to the nuage (Klattenhoff *et al.* 2009). The nuclease Squash also seems to be required for piRNA production, but its phenotype, in terms of the piRNA biogenesis is rather weak, although the

piRNA level decreases overall (Malone *et al.* 2009; Pane *et al.* 2007). Also, there is no identifiable Squash homolog in other species.

A similar post-transcriptional amplification loop has been suggested for vertebrates as well. In mice, MILI mainly binds sense while MIWI2 shows an antisense bias for piRNAs. Prenatal piRNAs show a strong enrichment of secondary piRNA signatures (A at the tenth nucleotide). MIWI2 is enriched two fold with secondary piRNAs compared to MILI. Thus, MILI is primarily biased towards piRNAs with a U as the 1st nt, while MIWI2 binds secondary piRNAs with an A at the 10th position (Aravin *et al.* 2008; Kuramochi-Miyagawa *et al.* 2008). Additionally, in *mili* mutants MIWI2 is not loaded with piRNAs (Aravin *et al.* 2008). Recently it was discovered that the ping-pong cycle in mouse is dependent on MHV, which is probably needed to load piRNAs in MIWI2 (Kuramochi-Miyagawa *et al.* 2010). The Ping-pong signatures have been detected in zebrafish as well. Additionally, piRNAs binding to ZIWI are preferentially antisense while those binding to ZILI have a sense orientation (Houwing *et al.* 2008). At the moment the ping-pong model is an attractive hypothesis for secondary piRNA production based on bioinformatic analysis of deep sequencing, but it is still lacking more experimental support.

Unlike miRNAs in animals, the piRNAs in various organisms have a 2`-O-methyl modification as a universal feature on their 3` terminal nucleotide (Horwich *et al.* 2007; Kirino & Mourelatos 2007c; Ohara *et al.* 2007a; Saito *et al.* 2007a). This modification is performed by the RNA methyltransferase HEN1 proteins (Horwich *et al.* 2007; Kirino & Mourelatos 2007a, b; Okamura *et al.* 2008c; Pelisson *et al.* 2007; Saito *et al.* 2007a; Vagin *et al.* 2006). It is not yet known at which point of the piRNA biogenesis the methylation occurs or what impact the modification has on the binding of piRNAs to Piwi proteins.

In conclusion, piRNAs are produced by at least by two different pathways. Primary piRNAs are generated from long single stranded transcripts of piRNA clusters that are often highly enriched for fragmented, diverged transposon remnants. The primary piRNAs in complex with Piwi proteins can target complementary transcripts, cleaving these and producing a new 5` end of a secondary piRNA. This means that piRNAs can amplify the silencing signal and at the same time silence the transposon transcript at the same time.

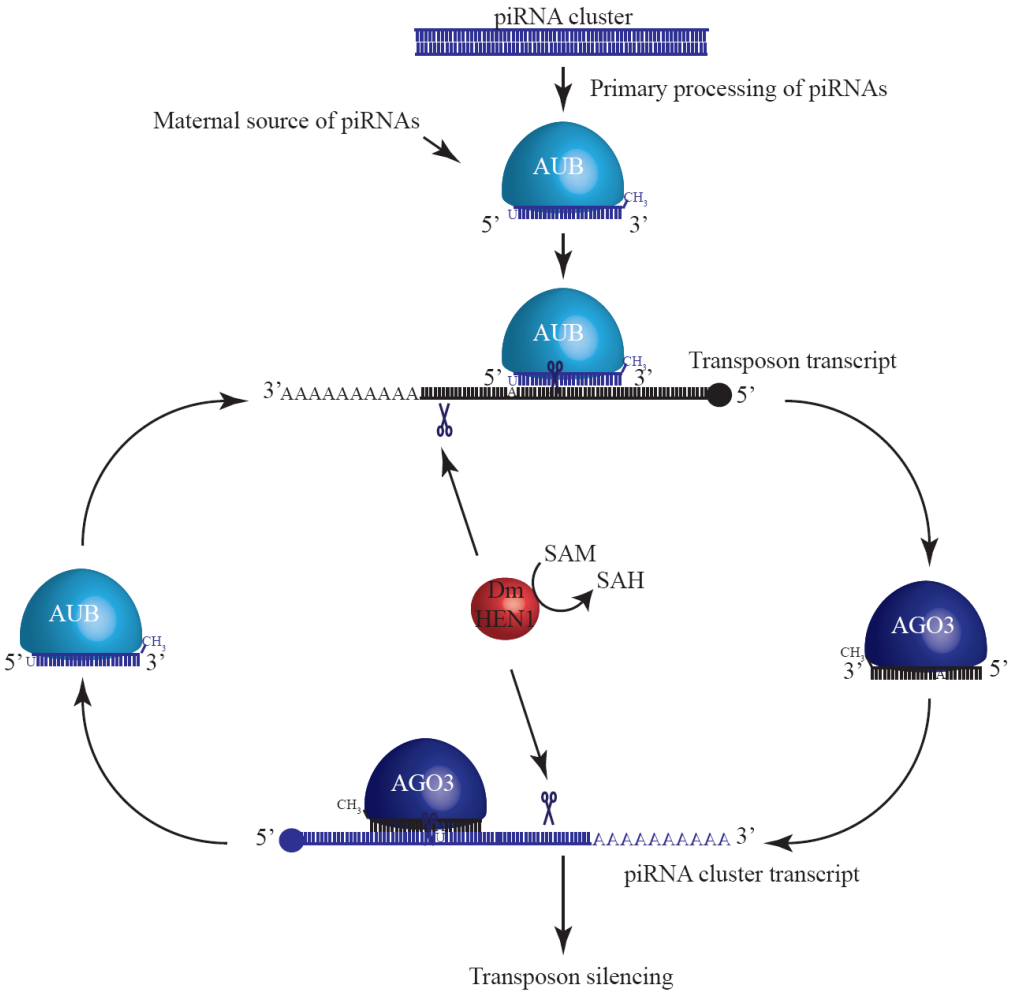


Figure 5: The Ping-pong cycle in *Drosophila*
Primary processed and maternal contributed piRNAs antisense to transposon transcripts are loaded into AUB and lead the complex to its target transposon transcript, whereby the slicer activity of AUB defines the 5`end of a secondary sense piRNA. The 3` end processing is unknown, but this end gets 2`-O-methylated by *Dm*HEN1. The new piRNA is loaded into an AGO3-containing complex to target a piRNA cluster transcript and generate a piRNA, which is antisense to the transposon transcript and incorporated into AUB again. This amplification cycle leads to transposon silencing. (Adapted from (Brennecke *et al.* 2007; Gunawardane *et al.* 2007))

1.3.7 piRNA function

The protection of the genome from transposable elements is crucial to its integrity. Thus, mobile elements have to be distinguished and silenced (Girard & Hannon 2008; Saito & Siomi 2010). Transposable elements can be divided in class I and class II. Class I transpose via an RNA element while class II use DNA elements and are characterized by a terminal-inverted repeat (TIR). The class I can again be subdivided into long terminal repeats (LTR) retrotransposons and non-LTR transposons, which are also called long-interspersed nuclear elements (LINE) or poly(A)-type retrotransposons. An ancient family of transposable elements is built from sequences deriving from small interspersed nuclear elements (SINEs) (Biemont & Vieira 2006).

Studies in mice, *Drosophila* and zebrafish show that piRNAs and Piwi proteins are essential for germline development and the control of transposable elements (Saito & Siomi 2010). An example for gene regulation via Piwi proteins is the AUB regulation of the *Stellate* locus by piRNAs, which are encoded by the *Suppressor of Stellate* (*Su(ste)*) (Aravin *et al.* 2001; Schmidt *et al.* 1999; Vagin *et al.* 2006). *Drosophila* males lacking the Y-chromosome are sterile, form cytoplasmic needles and/or nuclear star-shaped crystalline structures in spermatocytes (Bozzetti *et al.* 1995; Hardy *et al.* 1984) and show defects in meiosis, specifically in chromosome pairing (Palumbo *et al.* 1994). The crystal formation can be suppressed if the flies lack the *Stellate* (*Ste*) locus on the X-chromosome (Livak 1984). On the Y-chromosome the deletion of the *Suppressor of Stellate su(ste)* also leads to male sterility (Hardy *et al.* 1984) because *Ste* is overexpressed (Livak 1990). The euchromatic *Ste* locus and the heterochromatic *Su(ste)* region are homologous (Kalmykova *et al.* 1998). Both strands of the *Su(ste)* region are transcribed and repress the *Ste* locus (Aravin *et al.* 2001). If the *Drosophila spindle-E*, or *aub* is mutated, the *Ste* locus is not suppressed (Aravin *et al.* 2004; Aravin *et al.* 2001; Schmidt *et al.* 1999). piRNAs from this locus are only found to be antisense to *Su(ste)* (Aravin *et al.* 2004; Vagin *et al.* 2006). In *piwi* mutants, piRNAs from the *Su(Ste)* locus hyper-accumulate but *Ste* is silenced. In *aub* mutants the sense transcript accumulates but not the antisense piRNAs, suggesting that the antisense piRNA silence not only *Ste* mRNA but also target the sense piRNA (Vagin *et al.* 2006). The *Su(ste)* piRNAs are

strongly reduced in *ago3* mutant males as well, suggesting that AGO3 is needed to produce the piRNAs bound to AUB (Li *et al.* 2009).

Another example of regulation via piRNAs and Piwi proteins is the flamenco locus. The flamenco locus is a heterochromatic region on the X-chromosome, containing numerous fragments of transposable elements, which encode a large piRNA cluster and is required to repress the transposon activity of Gypsy, Idefix and ZAM (Brennecke *et al.* 2007; Sarot *et al.* 2004). If this region is disrupted, the piRNAs from here are reduced and the transposon activity increases (Brennecke *et al.* 2007). These results show that piRNAs act in the transposable element regulation. The disruption of *aub* has the same effect (Pelisson *et al.* 2007; Savitsky *et al.* 2006; Shpiz *et al.* 2007; Shpiz & Kalmykova 2007; Vagin *et al.* 2004).

In somatic cells, the first insight of the epigenetic role of Piwi proteins was found in *Drosophila*, where AUB and PIWI were shown to regulate position-effect variegation (PEV)(Pal-Bhadra *et al.* 2002). This is a mechanism in which euchromatin genes are silenced if they are in close proximity to a heterochromatic region, for example in chromosomal rearrangement (Dorer & Henikoff 1994, 1997; Fanti *et al.* 1998; Martin-Morris *et al.* 1997). PIWI is present in the nucleus and can be found directly associated with chromosomes and interacting with the Heterochromatin protein 1a (HP1a), a marker for heterochromatin, indicating that the piRNA pathway is directly connected to the heterochromatin-forming mechanism (Brower-Toland *et al.* 2007).

piRNAs are also able to activate transcription. Specific sub-telomeric heterochromatin (telomere-associated sequence, or TAS) piRNAs bound to PIWI promote euchromatic histone modifications and piRNA transcription in TAS on the right arm of chromosome 3 (3R-TAS). PIWI associates to this region and promotes transcription activity. In *piwi* mutants, this 3R-TAS accumulates heterochromatin histone modifications and heterochromatin protein 1 a (HP1a), which silence this locus and lead to a loss of germline stem-cell maintenance (Yin & Lin 2007).

In zebrafish, many piRNAs derive from transposable elements, preferentially from LTR and mainly in the antisense orientation to transposon transcripts. Examples are Gypsy*Dr*1, Gypsy*Dr*2 and Ngaro. In *zili* knockout mutants, the activity of several LTR, non-LTR and DNA elements is increased (Houwing *et al.* 2007). piRNAs are differently expressed in testes and ovaries and play also a role in zebrafish sex differentiation (Houwing

et al. 2007; Zhou *et al.* 2010). These sex-specific piRNAs derive from separate loci. Ovarian piRNAs are more often associated in groups (33%), when compared to testes piRNAs (20%).

Mammalian piRNAs can be divided in two different classes: The pre-pachytene and the pachytene piRNAs, according to the stage of spermatogenesis in which they appear (see Figure 4). Like fly piRNAs, pre-pachytene piRNAs which consists of embryonic and post-natal populations before P10, derive primarily from repetitive sources (35% in MILI or 60% in MIWI2), mainly SINE, LINE and LTR transposons, and are implicated in retrotransposon silencing (Aravin *et al.* 2007a; Aravin *et al.* 2007b; Carmell *et al.* 2007; Kuramochi-Miyagawa *et al.* 2008).

In mammals, piRNAs are implicated in establishment of *de novo* methylation during gametogenesis (Aravin & Bourc'his 2008; Aravin *et al.* 2008). During development, the global methylation patterns are extensively remodeled in waves of demethylation and *de novo* methylation (Hajkova *et al.* 2002). At 8 dpc the DNA methylation decreases together with a transient loss of the DNA *de novo* methyltransferases Dnmt1, Dmnt3a and Dmnt3b. To prevent deregulated transcription, POL II is suppressed as well (Seki *et al.* 2007). During the migration of the primordial germ cells (PGCs), the transcriptional repressor B-lymphocyte maturation-induced protein 1 (BLIMP1, also known as Prdm1) is associated with the histone arginine methyltransferase PRMT5 in the nucleus and mediates dimethylation of H2AR3and H4R3 to suppress their target genes (Ancelin *et al.* 2006). At 11.5 dpc BLIMP1/PRMT5 translocate from the nucleus into the cytoplasm in post-migrated PGCs and their targets are activated (Ancelin *et al.* 2006). Interestingly, H4R3me is also associated with activation of other genes by facilitating the acetylation of H4 tails (Strahl *et al.* 2001; Wang *et al.* 2001a) while H3R8me, which is also mediated by PRMT5, is a transcription repression mark (Pal *et al.* 2004).

It would appear that during 12.5 and 14.5 dpc the genome is not protected, but that is not the case. Post migratory-germ cell-specific genes such as the mouse Vasa homolog (MVH or also known as DEAD box polypeptide 4, Ddx4), Sycp3 (synaptonemal complex protein 3) and Dazl (deleted in azoospermia-like) carry methylation marks in their flanking regions and become hypomethylated when they are expressed at 13.5 dpc (Maatouk *et al.* 2006). Some transposable elements like LINE1 and some intracisternal A particles (IAPs) remain methylated even at 13.5 dpc (Hajkova *et al.* 2002). The male germ cells go into

mitotic arrest between 13.5 and 14.5 dpc. The imprinting marks are erased by demethylation during the PGC migration and new marks are established according to the gender of the animal. The parental methylation imprints (H19, Dlk1/Gtl2 and Rasgrf1) are established between 14.5 dpc and birth (20 dpc), with the paternal are imprinted before the maternal ones in the male germline (Davis *et al.* 1999; Davis *et al.* 2000; Kato *et al.* 2007; Li *et al.* 2004; Ueda *et al.* 2000). Dnmt3a plays a central role in this mechanism for all loci, but Rasgrf1 (RAS protein-specific guanine nucleotide-releasing factor1) additionally requires the *de novo* methyltransferase Dnmt3b (Kaneda *et al.* 2004; Kato *et al.* 2007), most likely due to the fact that several retrotransposons are encoded in this region. Both methyltransferases build a complex with Dnmt3l, which is highly expressed in prospermatogonia but has no methyltransferase activity (Bourc'his & Bestor 2004; Kaneda *et al.* 2004; Kato *et al.* 2007; Webster *et al.* 2005). The imprinting marks stay throughout the rest of the germ cell development.

Together with the imprinting marks retrotransposable elements undergo *de novo* methylation. While SINEB1 is mainly methylated by Dnmt3a, LINE1 and IAP are methylated by Dnmt3a and Dnmt3b, the methylation of all these sequences requires Dnmt3l (Kato *et al.* 2007). In cases of failure of the *de novo* methylation, retroelements are highly transcribed in spermatogonia and spermatocytes, the meiosis fails and germ cells are progressively lost by the mid-pachytene stage (Aravin *et al.* 2007b; Bourc'his & Bestor 2004; Hata *et al.* 2006; Webster *et al.* 2005). Current evidence would indicate a role for piRNAs bound to Piwi proteins in ensuring the genomic stability by silencing the transposable and repetitive elements (Aravin *et al.* 2003; Brennecke *et al.* 2007; Gunawardane *et al.* 2007; Houwing *et al.* 2007; Saito *et al.* 2006; Vagin *et al.* 2006). In *mili* or *miwi2* deficient males, LTR and non-LTR transposons are overexpressed due to a loss of DNA-methylation marks (Aravin *et al.* 2007b; Carmell *et al.* 2007; Kuramochi-Miyagawa *et al.* 2008). Apart from *piwi* mutants, the deletion of small piRNA clusters, such as *Nct1* and *2* in mice, lead to an increased activity of transposable elements, like LINE1 and open reading frame 1 (ORF1). This demonstrates that piRNAs themselves have roles in the transposon control. The mutation of the *Nct1* and *2* does not affect gene expression, fertility, or spermatogenesis (Xu *et al.* 2008).

How piRNAs perform silencing of the transposable elements and which factors are involved is not known. However, it is possible that silencing by piRNA is similar to the

mechanism of heterochromatin formation at peri-centromers induced by siRNAs in *Schizosaccaromyces pombe*. Here, the siRNA recruits the RNA-induced transcriptional silencing (RITS) complex to nascent transcripts from peri-centromers, which in turn recruits the HP1a homolog SWI6, triggering heterochromatin formation (Buhler *et al.* 2006; Verdel *et al.* 2004; Volpe *et al.* 2002).

Piwi proteins also play a role in dsDNA break-repair. Isolation of rat PIWI (RIWI) complexes identified the adenosine triphosphate (ATP)–dependent DNA helicase rRECQ1 (Lau *et al.* 2006), the homolog of the *Neurospora* QDE-3, which together with an AGO protein has a role in gene silencing (Catalanotto *et al.* 2002).The presence of RecQ in the RIWI complexes (Lau *et al.* 2006) is a indication for the function of the piRNA pathway in dsDNA-break repair (Hunter 2008). On the other hand, in absence of Piwi proteins, dsDNA breaks are caused by transposon accumulation (Klattenhoff *et al.* 2007).

Pachytene piRNAs are bound to MILI and MIWI, but their function is still not clear. Isolation of RNAs from polysome and RNP fractions containing MIWI shows that MIWI and piRNAs interact in polysome and are associated with mRNAs (Grivna *et al.* 2006a; Grivna *et al.* 2006b). MIWI is also associated with eIF4E, an mRNA cap binding protein that is responsible for translational control, and Dicer, the endonuclease needed for si- and miRNA production. The quantity of certain miRNAs is significantly reduced in MIWI mutants, meaning MIWI could be involved in miRNA-mediated translational control. The target mRNAs are also down-regulated, too, supporting a MIWI involvement in mRNA stability (Deng & Lin 2002). MILI seems to be involved in translational control as well, where it forms a stable complex with eIF3a and associates with eIF4E and eIF4G containing the m7G cap-binding complex (Unhavaithaya *et al.* 2009). However, the data supporting relationship between Piwi proteins and translation control needs to be obtained from other model systems to become accepted as one of the modes by which piRNA pathway regulates gene expression.

In mice, as in other mammals, piRNAs are only found in significant amounts in the male germline. While the male germline contains lifelong regenerating stem cells, the female gender relies on meiotically arrested oocytes stored from prenatal development. Nevertheless, the female germ cells have to be protected from transposable elements too. Perhaps this function in oocytes is carried out by endogenous small interfering RNAs deriving from

retroelements including LINE, SINE and LTR (Tam *et al.* 2008; Watanabe *et al.* 2006; Watanabe *et al.* 2008), as Piwi mutants display normal fertility in females.

Although a lot of knowledge was acquired since piRNAs were discovered in 2006, many questions still remain unanswered. While the ping-pong model explains secondary piRNAs, the biogenesis of primary piRNA remains unclear. Factors involved in the biogenesis and function of piRNAs are only just starting to be discovered. Although the pre-pachytene piRNAs are associated with transposable element silencing, many of these piRNAs do not fit this scheme. Also, there is still no function for pachytene piRNAs.

1.4 Factors involved in the piRNA pathway

1.4.1 The methyltransferase HEN1 adds methylation marks on the 3` end of small RNAs

Hen1 was identified in *Arabidopsis thaliana* as a gene involved in the specification of the stamen and carpel identities during the flower development, while searching for factors which enhances HUA1 and HUA2, leading to the name HUA ENHANCER 1 (HEN1) (Chen *et al.* 2002). Plants carrying *hen1* mutations are smaller, their fertility is reduced and their leaf shape alters (Chen *et al.* 2002). HEN1 is indirectly involved in stress response (Zhang *et al.* 2008), acquisition and resolution of copper (Abdel-Ghany & Pilon 2008), and the long distance transport of information via small RNAs (Buhtz *et al.* 2010). It S-adenosyl-L-methionine dependent methylates miRNA and siRNA duplexes with a preferential length of 21-24 nt after Dicer cleavage (Akbergenov *et al.* 2006; Katiyar-Agarwal *et al.* 2007; Li *et al.* 2005; Yang *et al.* 2006a; Yang *et al.* 2006b; Yu *et al.* 2005). The methylation of the small RNAs by HEN1 protects them from the addition of extra nucleotides, mostly U-residues, by an unknown terminal transferase or polymerase which leads to their destabilization and degradation (Chen 2005; Ebhardt *et al.* 2005; Li *et al.* 2005; Yang *et al.* 2006b). The miRNAs and the siRNA levels are greatly reduced in *hen1* mutants (Akbergenov *et al.* 2006; Boutet *et al.* 2003; Katiyar-Agarwal *et al.* 2006; Park *et al.* 2002; Vazquez *et al.* 2004a; Vazquez *et al.* 2004b).

The *Arabidopsis thaliana* HEN1 (942 aa) is a multi-domain protein. It contains two dsRNA-binding domains (dsRBD 1 and 2) separated by a LA-motif (another RNA binding site) containing domain at the N-terminus (Tkaczuk *et al.* 2006). This is followed by a PPIase (peptidyl-prolyl cis-trans isomerase)-like domain (PLD), lacking a typical PPIase active site (Tkaczuk *et al.* 2006). The C-terminus contains the catalytic domain, a typical Rossmann-fold MTase fold with an S-adenosyl methionine (SAM)-binding motif (Asp-Phe-Gly-Cys-Gly) and a nuclear localization motif (Anantharaman *et al.* 2002; Park *et al.* 2002; Vazquez *et al.* 2004a; Yu *et al.* 2005). Four of the five domains directly interact with the RNA substrate in an A-form confirmation (the dsRBD1 and 2, the MTase domain and the La domain). HEN1 binds to the small RNAs as a monomer and specifically recognizes duplexes with 2 nt overhangs at the 3` end (Huang *et al.* 2009) (see Figure 6A).

Based on homology and structure prediction, members of the HEN1-family were identified in metazoan, fungi, *Viridiplantae* and bacteria. Evolutionary analyses of this protein family suggest that the protein appeared first in the common ancestor of eukaryotes, before they separated into Viridiplantae and Metazoan/Fungi and was subsequently horizontal transferred to the bacteria.

The bacterial HEN1 is more similar to animal than plant, the only common feature of all being the MTase domain. Only the plant HEN1 contains the double stranded RNA binding domains and PPIase domain (Jain & Shuman 2010). The crystal structure of the MTase domain of bacteria HEN1 revealed a core seven-stranded β-sheet (β1-7), which is flanked on two sides by α-helices (αZ, αA-E), a common fold of Ado Met-dependent MTases. Additionally the MTase of HEN1 contains a short helix between β4 and αD and support domain between β5 and αE (see Figure 6 B). This is a conserved structure and unique to HEN1 proteins (Mui Chan *et al.* 2009) (see Figure 6 C). Recently it was shown, that different to former suggestions, the methyltransferase activity of bacteria HEN1 is manganese, and not magnesium dependent (Jain & Shuman 2011).

In bacteria HEN1 is essential for RNA repair. For example, in the bacterium *Anabaena variabilis*, HEN1 (ca. 460 aa) interacts with the polynucleotide Kinase-phosphatase PNKP, a protein involved in RNA repair, and forms a stabile heterotetramer *in vitro*. PNKP alone cannot repair RNAs, which were cleaved by ribotoxins. Together with HEN1 the RNAs are not only repaired but a methyl group at the cleavage site is added during

the progress. This protects the RNA from being cut again (Chan *et al.* 2009; Jain & Shuman 2010).

In *Tetrahymena*, Piwi-associated scanRNAs (26-31 nt) are 2`-O-methylated (Kurth & Mochizuki 2009). scanRNA are expressed only during the sexual reproduction (Mochizuki *et al.* 2002). This modification is catalyzed by HEN1p and protects the small RNAs from degradation. HEN1p and the Piwi proteinTwi1p interact where the single-stranded scanRNA are methylated after loading and exclusion of one RNA strand. This modification ensures the function of scanRNAs in DNA-elimination, influencing the efficient of the production of sexually functional progeny (Kurth & Mochizuki 2009).

Studies of the HEN1 *Drosophila* homolog (*Dm*HEN1), a 391 aa protein, shows that it is essential for piRNA, siRNA and hpRNAs methylation (Horwich *et al.* 2007; Okamura *et al.* 2008b; Pelisson *et al.* 2007; Saito *et al.* 2007a; Vagin *et al.* 2006). *Dm*HEN1 methylates single-stranded RNA (ssRNA) substrates (Horwich *et al.* 2007; Saito *et al.* 2007a) with a size range between 22 and 38 nt (Saito *et al.* 2007a). Loss of activity results in reduced length and abundance of piRNAs and siRNAs (Horwich *et al.* 2007; Saito *et al.* 2007a). Analysis of *Dm*HEN1 and the zebrafish *Dr*HEN1 revealed that the methylation of the small RNAs is needed for their stability. This occurs by protecting them from transferases preferentially adding Us at their 3` end, which subsequently attract 3`-5` exonucleases that trim the small RNAs (Ameres *et al.* 2010; Kamminga *et al.* 2010).

The mouse mHEN1 (395 aa) is expressed exclusively in the male germline and the recombinant protein expressed in *E. coli* is reported to have methylation activity, suggesting that this could potentially be the piRNA methyltransferase. This protein was able to methylate 20-40 nt long ssRNAs on their 3` ends (Kirino & Mourelatos 2007a, b). Until now it is not clear how this modification affects the piRNA pathway in mice and how the methyltransferase is recruited to this pathway. Further, it is not known how the methylation mark on small RNAs in animals is recognized by other components of the piRNA pathway, for example the PAZ domain of the murine Piwi proteins.

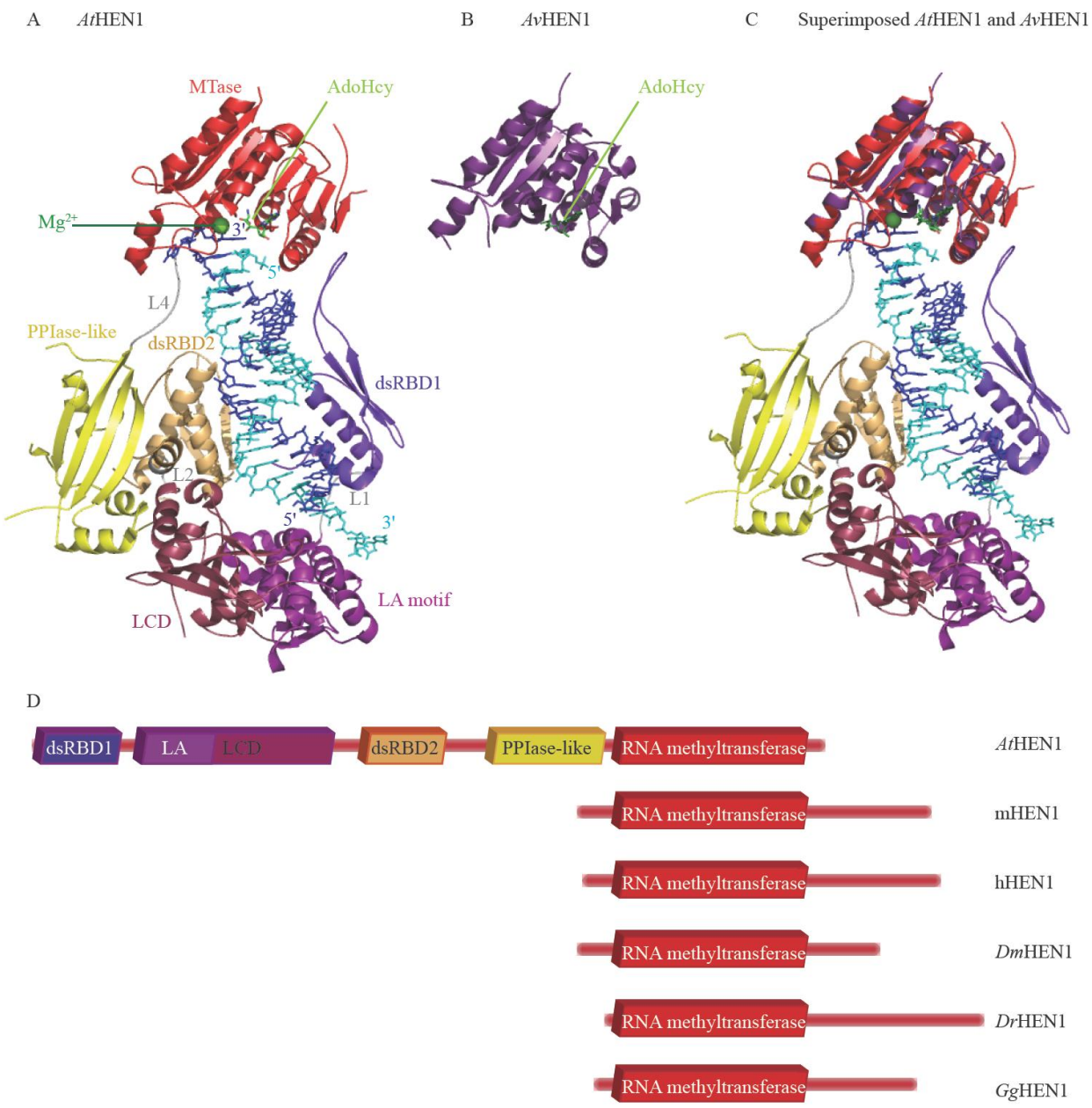


Figure 6: The methyltransferase of HEN1 is highly conserved.

(A) Ribbon diagram of *A. thaliana* HEN1 structure. The dsRBD1 is colored in blue, the La motif in magenta, the LCD in bordeaux, the dsRBD2 in sand, the PPIase-like domain in yellow and the MTase domain in red. The linkers are colored in gray. Indicated as a green sphere is the Mg2+ and the AdoHcy is indicated in green sticks. The RNA strands are indicated in blue and cyan (adopted from (Huang *et al.* 2009) and analyzed by Pymol)

(B) Ribbon diagram of the *Anabaena variabilis* HEN methyltransferase domain (purple). The bound AdoHcy is indicated in green (adopted from (Mui Chan *et al.* 2009) and analyzed by Pymol)

(C) Superimposed *At*HEN1 and *Av*HEN1(analyzed by Pymol), indicating the high similarity of the structures.

(D) Domain architecture of different HEN1. *At*, *Arabidopsis thaliana*; m, mouse; h, human; *Gg*, *Gallus gallus* (chicken); *Dr*, *Danio rerio* (zebrafish); *Dm*, *Drosophila melanogaster* (fruit fly). The methyltransferase domain is indicated in red and the LAM domain in green. The figure is based on (Chen 2007).

1.4.2 Piwi proteins carry methylated Arginines at their N-terminal region

Recently it was discovered that Piwi proteins carry post-translational modifications in form of symmetrically di-methylated arginines (sDMA) at their N-termini (Kirino *et al.* 2010a; Kirino *et al.* 2010b; Nishida *et al.* 2009; Reuter *et al.* 2009; Vagin *et al.* 2009). The first identified sDMA carrying residue was R74 in MILI (Reuter *et al.* 2009) (see Figure 7). R100, R146, R163 and R549 in MILI were identified as dimethylated and R95 as monomethylated by Vagin and colleagues (Vagin *et al.* 2009), while R74 was found to carry both types of methylation in the latter study. In contrast to these findings, Chen *et al.* published that R74, R83, R95 and R100 can be mono- and dimethylated, while R45, R146, R156 and R163 are only dimethylated (Chen *et al.* 2009). MIWI was first described to carry R49 and R371 monomethylated and R14 mono-or dimethylated (Vagin *et al.* 2009). Later Chen and colleagues published that R53 in MIWI can be both mono and dimethylated (Chen *et al.* 2009). In *Drosophila* the AUB protein showed sDMAs at R11, R13 and R15, while AGO3 contains sDMAs at R4, R68 and R70 (Nishida *et al.* 2009). These different findings for the methylation status maybe due to the source of material used for analysis, suggesting that Piwi methylation is a dynamic process.

sDMAs are known to modify the ability of proteins to carry out their biological functions. For example, Sm proteins, which are splicing machinery factors, contain sDMAs that are recognized by the single Tudor domain of the SMN (survival motor neuron) proteins to recruit U snRNAs (Buhler *et al.* 1999; Meister *et al.* 2002; Selenko *et al.* 2001). The methylation of the Sm proteins and Piwi proteins is catalyzed by protein methyltransferase 5 (PRMT5) in *Drosophila* as well in mouse (Kirino *et al.* 2009; Nishida *et al.* 2009; Vagin *et al.* 2009). In *Drosophila dprmt5* null-mutants, AGO3 and AUB are shown to be significantly reduced and PIWI only mildly, the mRNA level of these proteins remains unaffected (Kirino *et al.* 2009). In contrast, Nishida *et al.* could not detect the reduction of AUB (Nishida *et al.* 2009). piRNAs are also reduced, while the retrotransposon activity increases in *dprmt5* null-mutants, while the binding capacity of Piwi proteins to piRNAs is unaltered (Kirino *et al.* 2009). Interestingly, the AUB localization to the nuage is disturbed in *dprmt5* mutants (Nishida *et al.* 2009). A mouse mutant for *Prmt5* is not yet available (likely to be embryonic lethal given its role in transcription regulation during development, as described above), but it

was biochemically shown that the co-factor of PRMT5, WDR77, associates with MIWI, MILI and MIWI2 (Vagin *et al.* 2009). The *Drosophila* homolog to WDR77 is Valois, a nuage component whose genetic disruption leads to the grandchild-less phenotype (Anne & Mechler 2005; Cavey *et al.* 2005; Harris & Macdonald 2001; Schupbach & Wieschaus 1991)

In *Drosophila* and mice the methylation of the arginines at the N-termini of the Piwi proteins is essential for protein interactions with Tudor domain-containing proteins (Chen *et al.* 2009; Kirino *et al.* 2010a; Kirino *et al.* 2010b; Nishida *et al.* 2009; Reuter *et al.* 2009; Vagin *et al.* 2009; Vasileva *et al.* 2009; Wang *et al.* 2009a). Without these interactions the piRNA pathway is affected, display transposon activation, and the animals are male-sterile (Kirino *et al.* 2010a; Kirino *et al.* 2010b; Reuter *et al.* 2009; Vagin *et al.* 2009).

In zebrafish, two Piwi proteins are described; zebrafish PIWI (ZIWI) and zebrafish PIWI like (ZILI). Considering sequence homology, ZIWI seems to be the ortholog of MIWI and ZILI of MILI (Houwing *et al.* 2008; Houwing *et al.* 2007). ZILI and ZIWI display RG repeats in their N-termini similar to *Drosophila* and mouse Piwi proteins, but their methylation status has not been examined yet.

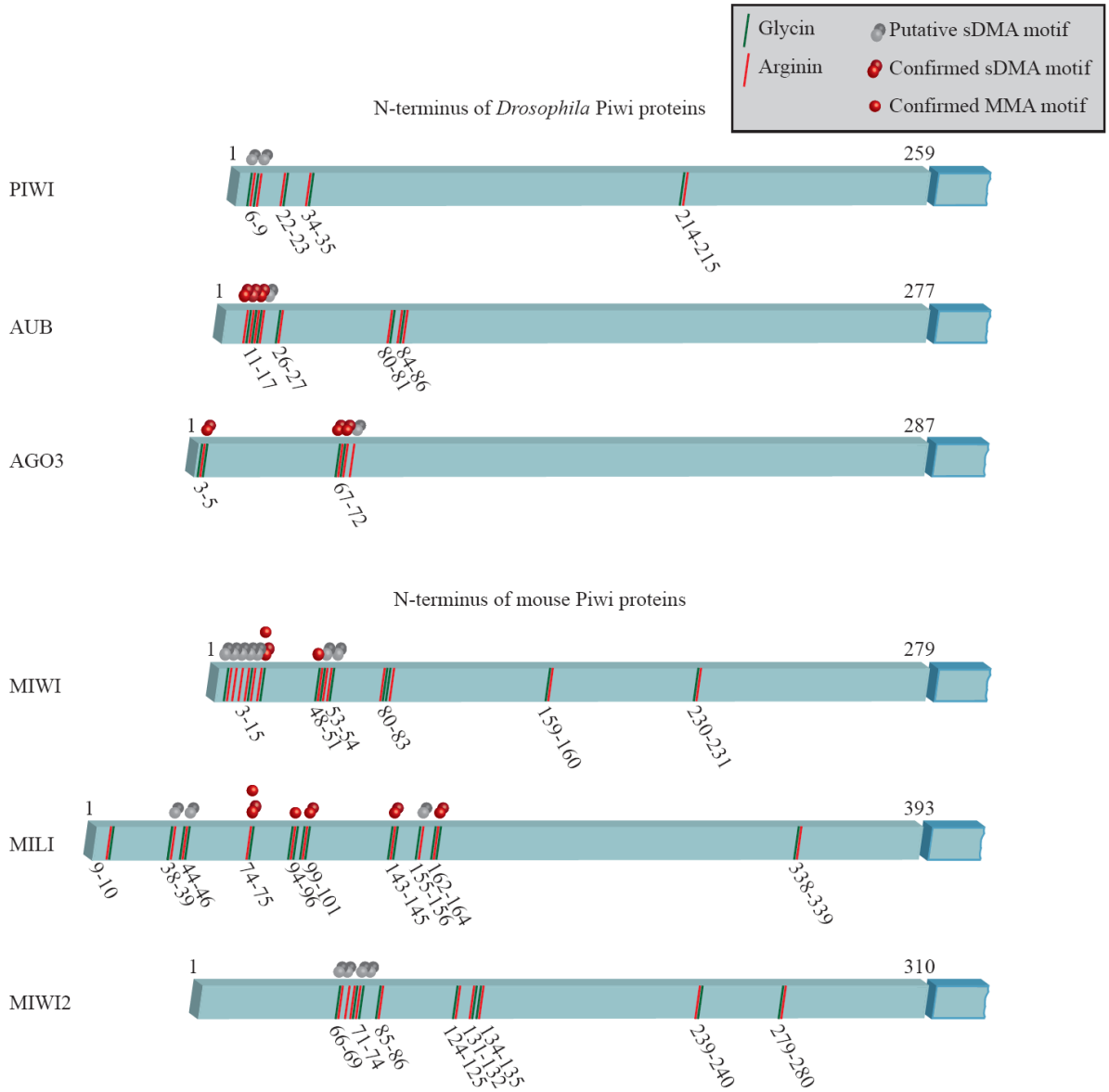


Figure 7: Symmetrical dimethylation at the N-termini of mouse and *Drosophila* Piwi proteins
The N-terminal region of *Drosophila* and mouse Piwi proteins (blue line) carry RG/GR repeats (R in red, G in green). Putative sequences for symmetrical dimethylation are RG (Kirino *et al.* 2009). Putative methylation sites are marked with grey balls and confirmed arginine methylations with red balls (Symmetrically dimethylation sites (two balls) and monometylation sites (one ball). Figure adapted from (Siomi *et al.* 2010)).

1.4.3 Tudor domain-containing proteins recognize post-transcriptional modifications on Piwi proteins

Tudor domains are found in many proteins across species and are named after the *Drosophila* TUDOR protein, containing 11 Tudor domains (Boswell & Mahowald 1985; Ponting 1997; Talbot *et al.* 1998). The PIWI protein and piRNAs levels in *tudor* mutants are not affected in the female germline (Kirino *et al.* 2010b), but the transposon derived level of piRNAs is altered (Nishida *et al.* 2009). But most importantly, the localization of AUB in the pole plasm is strongly reduced. Thus, the arginine methylation of AUB by dPRMT5 is required for an interaction with TUDOR, which in return is needed for the correct localization of AUB (Kirino *et al.* 2010b; Nishida *et al.* 2009). Transposon silencing is only faintly affected in *tudor* mutants, suggesting that the piRNA pathway in PGC specification is independent from transposon control (Kirino *et al.* 2010b). TUDOR can also associate with AGO3 through R68 and R72 methylation (Nishida *et al.* 2009). TUDOR associates with AGO3 and/or AUB is found in heteromeric complexes in different combinations, which contain pre-piRNA-like molecules of a few hundred nts in length. It was speculated that TUDOR provides a kind of quality control of piRNAs, which associate with AUB or AGO3, and is involved in loading (Nishida *et al.* 2009). Recently, the structure of an extended Tudor domain (Tudor 11) of TUDOR with methylated peptides of AUB was solved (Liu *et al.* 2010). The structure contains a typical Tudor domain, which is comprised of four β-strands, and an oligonucleotide binding (OB) fold. The Tudor domain is an insertion between β-strands 2 and 7 of the OB domain and is connected to the OB domain via linker helices. The structure also revealed that the interaction with methylated peptides is dependent on a conserved aromatic cage (Liu *et al.* 2010) (see Figure 8).

A total of 26 Tudor domain-containing proteins are expressed in mice, but only 10 are considered to be members of the Tudor domain-containing (TDRD) family (Siomi *et al.* 2010). TDRD1, containing a MYND domain and four Tudor domains, interacts with MILI and MIWI, where it was shown that the Tudor domains recognize the sDMAs R74 at the N-termini of MILI (Chen *et al.* 2009; Kojima *et al.* 2009; Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a). It may also interact with MIWI2, but the data is controversial (Kojima *et al.* 2009; Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a). Not only TDRD1

interacts with Piwi proteins in mouse testes, seven TDRDs (TDRD1-TDRD9 except for TDRD 3 and 5) are associated with these proteins. For example, TDRD6, the mouse homolog of *Drosophila* TUDOR (Hosokawa *et al.* 2007), interacts with MILI and MIWI (Chen *et al.* 2009; Vagin *et al.* 2009; Vasileva *et al.* 2009). TDRD9, the homolog of *Drosophila* Spindle-E, interacts with MIWI2 and MIWI (Aravin *et al.* 2009; Shoji *et al.* 2009; Vagin *et al.* 2009). However, it is not clear if TDRD interactions with MIWI2 require sDMAs, as Spindle-E was not found to associate with AUB peptides (Nishida *et al.* 2009). MIWI2 and MIWI can also interact with TDRD2 (Chen *et al.* 2009; Vagin *et al.* 2009), as well as MILI and MIWI with TDRD7 (Kirino *et al.* 2010b; Vagin *et al.* 2009; Vasileva *et al.* 2009). In conclusion, the methylation of the RG repeats is needed for the interaction of Piwi proteins with the Tudor domain-containing protein family, but most probably is not sufficient for the binding (Siomi *et al.* 2010). Taking into account that many TDRDs interact with Piwi proteins, it might be that the piRNA pathway is a central function of TDRDs.

Functional evidence of the involvement of the TDRDs in the piRNA pathway was provided by studies of TDRD1 and TDRD9. Mutations in *tdrd1* and *tdrd9* lead to male sterility (Chuma *et al.* 2006; Shoji *et al.* 2009).

The loss of these TDRDs activate transposable elements (Reuter *et al.* 2009; Shoji *et al.* 2009; Vagin *et al.* 2009), and, while the amount of piRNAs bound to MILI is not affected in *tdrd1* mutants, the small RNA profile changes (Reuter *et al.* 2009; Vagin *et al.* 2009). Interestingly, TDRD1 and TDRD9 are not involved in the loading of piRNAs into the Piwi proteins (Reuter *et al.* 2009; Shoji *et al.* 2009). Additionally, the nuclear localization of MIWI2 is disturbed in *tdrd1* mutants as in a *mili*-mutant (Aravin *et al.* 2008; Reuter *et al.* 2009; Vagin *et al.* 2009). All these studies suggest that the TDRDs are involved in the piRNA pathway, but that the mechanism of their association with transposable elements silencing is not known.

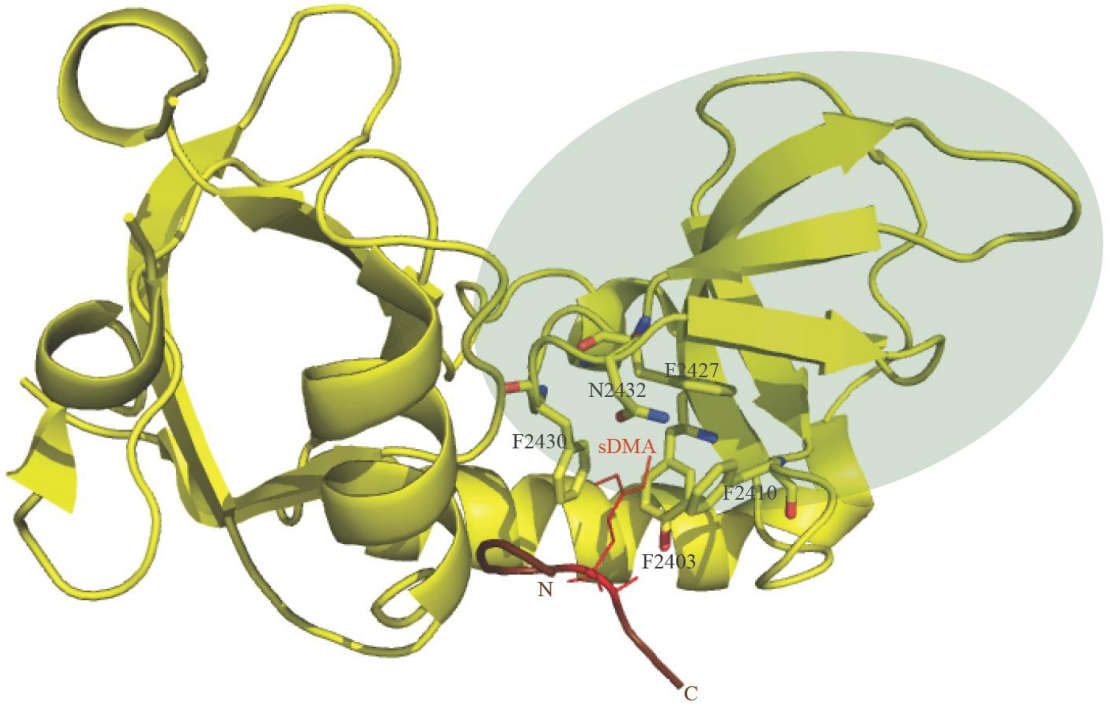


Figure 8: Structure of the Tudor domain 11 of *Drosophila* TUDOR
Ribbon display of the structure of the extended Tudor 11 domain (yellow) of *Drosophila* TUDOR with a Aubergine peptide (red) containing a sDMA as a in stick model. The canonical Tudor domain is presented in the green oval. The sDMA binding pocket residues are also shown as stick (adapted from (Liu *et al.* 2010) using PyMol)

1.4.4 Helicases in the piRNA pathway

Many Piwi-associated factors and genes genetically linked to the piRNA pathway have been described, but their direct involvement has not been shown yet. Among those factors are several putative helicases such as Vasa, Spindle-E and Armitage in *Drosophila* and the mouse homologs of Vasa (MVH) and Armitage (MOV10L) (Frost *et al.* 2010; Kuramochi-Miyagawa *et al.* 2004; Kuramochi-Miyagawa *et al.* 2010; Malone *et al.* 2009; Olivieri *et al.* 2010; Wang *et al.* 2010; Zheng *et al.* 2010).

The DEAD-box RNA helicase Vasa is an evolutionarily conserved RNA helicase and essential for germ cell development (Noce *et al.* 2001). In *Drosophila* Vasa is responsible to regulate target mRNAs such as nanos (Becalska & Gavis 2009; Noce *et al.* 2001), a component of the germ cell nuage (Hay *et al.* 1990; Lasko & Ashburner 1990). In MVH-deficient mice there is a block of the spermatogenesis at the first meiotic division (Tanaka *et al.* 2000). In both organisms, Vasa, and accordingly MVH are localized in the nuage in *Drosophila* (Liang *et al.* 1994) and chromatoid body in round spermatids in mice (Fujiwara *et al.* 1994; Meikar *et al.* 2010; Raz 2000).

MVH is expressed in the PGCs as they arrive at the gonad from 10.5 dpc onwards, and in the germ cells undergoing gametogenic processes until the post-meiotic stage in both sexes (Toyooka *et al.* 2000). It is shown that MVH associates with MILI and MIWI (Kuramochi-Miyagawa *et al.* 2004), connecting the protein to the piRNA pathway. In *vasa* mutants the spermatogenesis is disrupted between the leptotene to zygotene stage due to apoptosis of the premeiotic germ cells and decreased differentiation and proliferation activity, resulting in male sterility, while the female fertility is unaffected (Tanaka *et al.* 2000). In male germ cells of the *mvh*-deficit mice, the DNA methylation is significantly impaired at the postnatal stage and consequently the activity of retrotransposable elements is up regulated (Kuramochi-Miyagawa *et al.* 2010). While piRNAs are severely reduced, some do remain. This phenotype is similar to those of *mili*- and *miwi2*-deficient mice (Kuramochi-Miyagawa *et al.* 2010; Kuramochi-Miyagawa *et al.* 2008). Interestingly, while MILI is still loaded with piRNAs in *mhv*-deficit mice, MIWI2 is not associated with piRNAs (Kuramochi-Miyagawa *et al.* 2010), indicating a defective ping-pong cycle. Additionally, the localization of MIWI2 is shifted from the nucleus to the cytoplasm and out of the P-bodies, and MILI is not

localized in the intermitochondrial cement, a form of germ granule in fetal post-spermatogonia (Kuramochi-Miyagawa *et al.* 2010). These results indicate an involvement of MVH in the piRNA biogenesis, presumably during the early ping-pong cycle and the creation of the intermitochondrial cement (Kuramochi-Miyagawa *et al.* 2010), but it is not clear what the molecular function of this protein is.

The putative helicase Armitage in *Drosophila*, which is mainly expressed in germline cells (Cook *et al.* 2004), but also detected in somatic follicle cells (Olivieri *et al.* 2010), seems to be involved in the primary piRNA processing (Olivieri *et al.* 2010). Although that it was reported to be only required for the secondary pathway (Malone *et al.* 2009). The differing observations are probably due to the fact that the Armitage locus contains two promoters, transcribing a longer and shorter mRNA, and leading to two different isoforms of the protein. The longer isoform is found in somatic and germline cells, while the shorter form is only present in the somatic cells. *armitage* knockdown flies lose primary/somatic piRNAs and show increased activity of some retro-transposable elements (Olivieri *et al.* 2010). Similar, in *mov10l*-deficient mice neither MILI nor MIWI2 are loaded with piRNAs, as piRNAs are not produced in these knockout mice. Due to this loss, retrotransposons are not DNA-methylated and are thus activated (Zheng *et al.* 2010).

In the follicle cells Armitage is found in 1-3 perinuclear foci and co-localizes with the Tudor domain-containing putative helicase Yb (Zheng *et al.* 2010), with which it interacts (Olivieri *et al.* 2010). Yb accumulates in somatic support cells of male and female gonads in the cytoplasm, the so called Yb bodies, which are in close proximity to mitochondria. It is required for regulation of PIWI expression and for female fertility (Szakmary *et al.* 2009). In *yb* knockdowns the Armitage localization is lost, while in *armitage* mutants the Yb bodies are not even detectable (Olivieri *et al.* 2010). As with *armitage*, *yb* knockdowns increase the transposon activity (Olivieri *et al.* 2010).

Loss of *armitage* and *yb* lead to a loss of the nuclear localization of PIWI. Interestingly, in knockdowns of the putative nuclease *zucchini* PIWI localizes in the perinuclear granule with and dependent on Armitage. This finding suggests that PIWI also transiently localizes to the cytoplasm and may explain why PIWI and Armitage are found to interact. Similar results were obtained in the germ cells, where Armitage localizes to the nuage (Olivieri *et al.* 2010). The direct interaction of Armitage and PIWI is supported by the

result of Zheng *et al*. that shows that the mouse Armitage homolog MOV10L interacts with all three mouse Piwi proteins (Zheng *et al.* 2010). The level of PIWI is reduced in *armitage*, and less affected in the y*b* mutants, indicating that Yb is only required in the somatic piRNA pathway. The level of Armitage is neither affected in *piwi*, nor *yb* mutants. These results suggest that a failure in the loading of PIWI with piRNAs destabilizes the protein resulting in mis-localization and that the PIWI loading occurs in the cytoplasm (Olivieri *et al.* 2010). Similar results were obtained in *mov10l* knockout mice, where the predominantly nuclear localization of MIWI2 is disturbed and observed in the cytoplasm (Zheng *et al.* 2010). Unfortunately, the mechanism by which Armitage or MOV10L regulate the loading or the stability of piRNAs during the primary piRNA pathway, is not known yet.

1.5 Project description

This thesis aims to study piRNA pathway factors and to understand their contributions to this pathway. The project can be divided broadly into two parts: the studies on the piRNA methyltransferase mHEN1 and the characterization of *Dr*TDRD1 and *Dr*MOV10L in zebrafish embryos.

1.5.1 The role of the piRNA methylation and the methyltransferase mHEN1

piRNAs carry a 2`-O-methyl group on their 3` terminal nucleotide (Horwich *et al.* 2007; Kirino & Mourelatos 2007c; Ohara *et al.* 2007a; Saito *et al.* 2007a). When the project was started the methyltransferase of piRNAs was not known, but soon, homologues of the plant HEN1 RNA methyltransferase were identified in *Drosophila*, mice and other organisms (Horwich *et al.* 2007; Kamminga *et al.* 2010; Kirino & Mourelatos 2007b; Saito *et al.* 2009). While it was directly shown in *Drosophila* that *Dm*HEN1 acts on si-and piRNAs (Horwich *et al.* 2007; Saito *et al.* 2007a), the mouse homolog of HEN1 was suggested to be a piRNA methyltransferase due to its expression in the germline and because the recombinant protein is able to methylate 20-40 nt long ssRNAs on their 3` ends (Kirino & Mourelatos 2007a, b).

The aim of this project was to show that endogenous mHEN1 is the actual piRNA methyltransferase and to answer the following questions: How is mHEN1 recruited to the piRNAs? Is there an association between the Piwi proteins and mHEN1 and which domains are involved in this interaction? Do post-translational marks on the Piwi proteins play a potential interaction role with mHEN1, as they do with Tudor domain-containing proteins? What is the sub-cellular localization of mHEN1 in the context of Piwi proteins? And finally, when does mHEN1 catalyzes the methylation of piRNAs?

Further, it is not clear how the methylation marks on small RNAs in animals is recognized by other components of the piRNA pathway, for example the PAZ domain of the murine Piwi proteins. Another aim was therefore to characterize the potential 3` RNA

binding of PAZ domain and to perform structural and biophysical studies of the association with methylated and non-methylated substrates.

To answer these questions, immunoprecipitations of the endogenous mHEN1 protein were needed and analysis of the methyltransferase activity of the protein. Interaction studies of mHEN1 with different piRNA pathway components and subsequent deletion constructs were therefore required to link the methyltransferase directly to the pathway. In order to gain more insights into mHEN1 biological functionality, localizations of the different components in purified germ cell populations were considered necessary to place the protein and complex in time and space. Furthermore, the structure of a Piwi PAZ domain could reveal the recognition mechanism of methylated piRNAs.

Chapter 2 addresses the direct recruitment of the piRNA methyltransferase mHEN1 via interaction with the Piwi protein MILI.

1.5.2 TDRD1 and MOV10L, two proteins involved in the piRNA pathway

piRNAs are responsible for silencing transposable elements in the germline (Saito & Siomi 2010), but while this function is known, their biogenesis factors are poorly understood. Two piRNA pathway components, TDRD1 and MOV10L from mouse, were studied in our laboratory (Reuter *et al.* 2009; Zheng *et al.* 2010). TDRD1 is a component of the transposon silencing machinery, interacting with the mouse Piwi proteins (Kojima *et al.* 2009; Lau *et al.* 2006; Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a). MOV10L is a putative helicase which is thought to be involved in primary piRNA biogenesis given that it associates with all three murine Piwi proteins and that in its absence, neither MILI nor MIWI2 are loaded with piRNAs (Frost *et al.* 2010; Zheng *et al.* 2010). The third chapter aims to identify the homologs of the mouse TDRD1 (*Dr*TDRD1) and MOV10L (*Dr*MOV10L) in zebrafish and to characterize the function of these proteins using *Danio rerio* embryos as a model system. Other questions include; how does *Dr*TDRD1 perform its function in transposon silencing? What is the function of *Dr*MOV10L in the primary piRNA pathway? When and where are these proteins expressed and localized in the zebrafish? Are they associated with the zebrafish Piwi proteins, as shown for mouse homologs, or is there interplay between

*Dr*TDRD1 and *Dr*MOV10L? Are *Dr*MOV10L and *Dr*TDRD1 involved in transposon silencing?

Besides biochemical studies using RT-PCRs, immunoprecipitations, binding assays and localization experiments, another aim in this project was to develop a new method to gain rapid functional data of piRNA-associated components. This was carried out by developing a fluorescence reporter assay to detect changes in the piRNA content in the zebrafish embryo system. A zebrafish facility at the EMBL Grenoble had to be established for this purpose.

1.6 Description du project

Cette thèse a pour objectif l'étude des facteurs de la voie des piRNA et la compréhension de leur contribution à cette voie. Le projet peut être divisé en deux grandes parties: les études sur la piRNA-méthyltransférase mHEN1, et la caractérisation de *Dr*TDRD1 et *Dr*MOV10L dans les embryons de poissons zèbres.

1.6.1 Le rôle de la méthylation des piRNA et de la méthyltransférase mHEN1

Les piARN portent un groupement 2'-O-méthyle sur leur nucléotide 3'terminal (Horwich *et al.* 2007; Kirino & Mourelatos 2007c; Ohara *et al.* 2007a; Saito *et al.* 2007a). Lorsque le projet a débuté, la méthyltransférase des piARN n'était pas connue, mais rapidement les homologues de HEN1, ARN-méthyltransférase de plante, ont été identifiés chez la drosophile, la souris et d'autres organismes (Horwich *et al.* 2007; Kamminga *et al.* 2010; Kirino & Mourelatos 2007b; Saito *et al.* 2009). Bien qu'il ait été directement démontré chez la drosophile que *Dm*HEN1 agit sur les si- et piARNs (Horwich *et al.* 2007; Saito *et al.* 2007a), il a été suggéré que l'homologue de HEN1 chez la souris soit une piARN-méthyltransférase du fait de son expression dans la lignée germinale et aussi parce que la protéine recombinante est capable de méthyler les ARN simple brin de 20-40 nt de long au niveau de leur extrémité 3'(Kirino & Mourelatos 2007a, b).

L'objectif de ce projet était de montrer que la protéine mHEN1 endogène est la véritable piARN-méthyltransférase et de répondre aux questions suivantes: comment mHEN1 est recrutée par les piARN ? Y a-t-il une association entre les protéines Piwi et mHEN1 et quels domaines sont impliqués dans cette interaction ? Est-ce que les marques post-traductionnelles des protéines Piwi jouent un rôle potentiel d'interaction avec mHEN1, comme c'est le cas avec les protéines contenant des domaines Tudor ? Quelle est la localisation sub-cellulaire de mHEN1 dans l'environnement des protéines Piwi ? Et enfin, quand la protéine mHEN1 catalyse-t-elle la méthylation des piARN ?

En outre, la façon dont les marques de méthylation des petits ARN chez les animaux sont reconnues par d'autres composants de la voie de piARN n'est pas claire, c'est le cas par

exemple du domaine PAZ des protéines Piwi murin. Un autre objectif était donc de caractériser le domaine de liaison potentiel de PAZ avec l'extrémité 3' de l'ARN et d'effectuer des études structurales et biophysiques de l'association avec des substrats méthylés et non méthylés.

Pour répondre à ces questions, des immunoprécipitations de la protéine endogène mHEN1 ainsi que l'analyse de son activité méthyltransférase étaient nécessaires. Les études d'interaction de mHEN1 avec différentes composantes de la voie des piARN et l'utilisation de constructions permettant leur délétion ont donc été nécessaires pour relier la méthyltransférase directement à la voie des piARN. Afin d'en savoir plus sur la fonction biologique de mHEN1, des expériences de localisation des différents composantes dans des populations de cellules germinales purifiées ont été nécessaires pour replacer la protéine et le complexe dans le temps et l'espace. De plus, la structure d'un domaine PAZ Piwi a pu révéler le mécanisme de reconnaissance des piARN méthylés.

Le chapitre 2 présente le recrutement direct de la piARN-méthyltransférase mHEN1 via l'interaction avec la protéine Piwi MILI.

1.6.2 TDRD1 et MOV10L, deux protéines impliquées dans la voie des piARN

Les piARNs sont responsables du mécanisme de répression des elements transposables dans les lignées germinales (Saito & Siomi 2010). Alors que leur fonction est connue, ce n'est pas le cas de leur biogénèse. Deux éléments de la biogénèse des piARNs, TDRD1 et MOV10L, ont été étudiés dans notre laboratoire (Reuter *et al.* 2009; Zheng *et al.* 2010). TDRD1 est une composante du transposon de la machinerie de répression, et intéragit avec les les protéines Piwi chez la souris (Kojima *et al.* 2009; Lau *et al.* 2006; Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a). MOV10L est quant a elle une hélicase putative qui est supposée être impliqué dans les premières étapes de la biogénèse des piARNs puisqu'elle s'associe avec les trois protéines Piwi murines et, lorsque celle-ci est absente, MILI et MIWI2 sont dépourvues de piARNs (Frost *et al.* 2010; Zheng *et al.* 2010). Le troisième chapitre à pour but de d'identifier les homologues des versions murines de TDRD1 (*Dr*TDRD1) and

MOV10L (*Dr*MOV10L) chez le poisson-zèbre et de caractériser la fonction de ces protéines en utilisant *Danio rerio* comme modèle. Les autres questions que nous nous sommes posées sont : par quel mécanisme est-ce que *Dr*TDRD1 inactive le transposon ? Quel est la fonction de *Dr*MOV10L dans la biogénèse des piARNs? Où et quand ces protéines sont-elles exprimées chez le poisson-zèbre ? Sont-elles associées aux protéines Piwi du poisson-zèbre, comme il l'a été montré chez la souris, ou est-ce qu'elles interagissent entre elles ? *Dr*MOV10L et *Dr*TDRD1 sont-elles impliquées dans le mécanisme de répression du transposon?

En marge des études biochimiques utilisant les techniques de RT-PCR, immunoprécipitation, de tests d'intéraction et d'expériences de localisation, nous avions aussi pour but de développer une nouvelle méthode nous permettant d'obtenir rapidement des données fonctionnelles sur les protéines associées aux piARNs. Dans cette optique, nous avons développé un test de fluorescence pouvant détecter les changements de quantités de piARN chez l'embryon du poisson-zèbre. Une plateforme à été crée à l'EMBL Grenoble dans ce but.

Chapter 2: Direct recruitment of the piRNA methyltransferase mHEN1 via interaction with the Piwi protein MILI

2.1 Contribution remark

This chapter describes studies aimed at understanding the role of the mouse piRNA methyltransferase, mHEN1, in piRNA biogenesis. In writing this chapter as a journal manuscript I have assembled data generated mainly by myself and some data of a postdoc (Dr. Michael Reuter) in the laboratory. Michael's contribution to this manuscript is the piRNA methylation assays with endogenous mHEN1, described in Figure 9, and the immunoprecipitation to detect the endogenous mHEN1 associated with MILI or co-purification of piRNAs, shown in Figure S 2. Both experiments relied on an anti-mHEN1 antibody that I prepared and purified myself. All the rest of the data is from my own work.

2.2 Remarques concernant ma contribution

Ce chapitre décrit des études visant à comprendre le rôle de la piARN-méthyltransférase mHEN1 chez la Souris, dans la biogenèse des piARN. En écrivant ce chapitre comme un manuscrit de journal, j'ai rassemblé les données générées principalement par moi-même ainsi que celles d'un post-doctorant (Dr Michael Reuter) du laboratoire. La contribution de Michael à ce manuscrit est un essai de méthylation des piARN utilisant la protein mHEN1 endogène (décrit dans la figure 9), l'immunoprécipitation permettant de détecter la proteine mHEN1 endogène associée à MILI et la co-purification de piARNs (représentée sur la figure S 2). Les deux expériences sont basées sur un anticorps de mHEN1 que j'ai préparé et purifié. Le reste des données provient uniquement de mon travail.

2.3 Direct recruitment of the piRNA methyltransferase mHEN1 via interaction with the Piwi protein MILI

Stephanie Eckhardt[1], Michael Reuter[1], Emerald Perlas[2], Sophie Pison-Rousseaux[3], Saadi Khochbin[3] and Ramesh S. Pillai[1]

[1]European Molecular Biology Laboratory, 6 Rue Jules Horowitz, BP 181, 38042 Grenoble, France.

[2]Mouse Biology Unit, EMBL Monterotondo Outstation, via Ramarini 32, 00016, Monterotondo, Rome, Italy.

[3]Institut National de la Santé et de la Recherche Médicale (INSERM), U823, Institut Albert Bonniot, Grenoble F-38706, France.

Correspondence to:

Ramesh Pillai
Phone: xx-33-4-76 20 7446
Fax: xx-33-4-76 20 7199
E-mail: pillai@embl.fr

2.3.1 Abstract

Piwi-interacting RNAs (piRNAs) are gonad-specific small RNAs that associate with the Piwi subfamily of Argonaute proteins. Together they are implicated in silencing of transposons and other selfish genetic elements within the animal germline. One defining feature of all characterized piRNAs to date is the presence of a 2`-O-methyl group at their 3` terminal nucleotide. This modification is recognized by the PAZ domain of Piwi proteins. The methyl modification is catalyzed by the RNA methyltransferase HEN1. Herein we present the methyltransferase activity of the endogenous mouse homolog of HEN1 (mHEN1) from testes extracts. Although all three murine Piwi proteins, MIWI, MILI and MIWI2, in testes incorporate methylated piRNAs, only MILI shows detectable interaction in both testes extracts and HEK293 cell cultures. The interaction site was mapped to the N-terminal part of MILI, to a region proximal to the PAZ domain, but distinct from the region mediating methylation-dependent interaction with Piwi-associated factors like Tudor proteins. mHEN1 recognizes this site with a C-terminal region close to the methyltransferase domain, bringing its active site in close proximity to the 3`end of the piRNA. Piwi proteins and other piRNA processing machinery components are all localized in cytoplasmic granules called nuage in germinal cells. Strikingly, mHEN1 is localized to a singular spot, distinct from the nuage, in haploid round spermatids. This distinct localization suggests a temporal assembly of the mHEN1/MILI complex. Considered together, this study uncovers the physical basis for recruitment of the piRNA methyltransferase to the piRNA pathway.

2.3.2 Resume

Les piARNs sont des petits ARNs spécifiques des gonades qui s'associent avec la sous-famille Piwi, appartenant à la famille des protéines Argonaute. Ensemble, ils sont impliqués dans la répression des transposons et autres éléments génétiques égoïstes dans la lignée germinale animale. Une propriété fondamentale de tous les piARNs caractérisés à ce jour est la présence d'un groupement 2'-O-methyl sur le nucléotide terminal de l'extrémité 3'. Cette modification est reconnue par le domaine PAZ des protéines Piwi. La modification du groupement méthyl est catalysée par l'ARN méthyltransferase HEN1. Ici, nous présentons l'activité méthyltransferase de l'homologue de souris endogène (mHEN1) obtenu à partir d'extraits testiculaires. Bien que les trois protéines Piwi murines, MIWI, MILI et MIWI2, incorporent les piARNs méthylés dans les testicules, seul MILI montre une interaction détectable à la fois dans les extraits testiculaires et les cellules HEK293. Le site d'interaction a été identifié sur la partie N-terminale de MILI, dans une région proche du domaine PAZ, mais distincte de la région médiant l'interaction méthylation-dépendante avec les facteurs associés à Piwi, tels que les protéines Tudor. mHEN1 reconnait ce site avec une région C-terminale proche du domaine méthyltransferase, de sorte que le site actif vienne à proximité de l'extrémité 3' du piARN. Les protéines Piwi, ainsi que les autres composants de la machinerie traitant les piARNs, sont tous localisés dans des granules cytoplasmiques appelés « nuage » dans les cellules germinales. De facon surprenante, mHEN1 se concentre au niveau d'un seul point distinct du nuage, dans les spermatides haploïdes ronds. Cette localisation distincte suggère donc un assemblage temporel du complexe mHEN1 / MILI. Mis en commun, ces résultats suggèrent donc une base physique pour le recrutement des piARN methyltransferases dans la voie de signalisation des piARNs.

2.3.3 Introduction

A wide variety of non-coding RNAs are involved in gene regulation in both prokaryotes and eukaryotes (Birney *et al.* 2007; Mattick 2009). A class of non-coding RNAs of the size range 19-31 nucleotides (nt) form a distinct group and are generally called the small RNAs (Ghildiyal & Zamore 2009). One characteristic feature of small RNAs is their association with RNA binding proteins called Argonautes. The Argonaute family is highly conserved in both prokaryotes and eukaryotes (Cox *et al.* 1998; Hock & Meister 2008), however, small RNAs are currently described only in eukaryotes. Argonaute proteins are defined by the presence of two conserved domains - PAZ and PIWI (Hock & Meister 2008). Structural studies of prokaryotic full-length Argonautes and eukaryotic Argonaute domains have revealed the PAZ domain to resemble an oligonucleotide-binding (OB) fold with a hydrophobic cavity that binds the 3` end of a small RNA. The PIWI domain comprises two distinct regions: the MID domain which has a 5` phosphorylated nucleotide binding pocket and an RNase-H-like fold domain, which contains an endonuclease activity site present in some Argonautes (slicers) (Parker *et al.* 2004; Song *et al.* 2004; Wang *et al.* 2008a; Wang *et al.* 2008b; Yuan *et al.* 2005).

Argonaute proteins can be broadly classified into two subfamilies: Ago and Piwi (Sasaki *et al.* 2003; Tolia & Joshua-Tor 2007). The ubiquitously expressed Ago subfamily members bind small RNAs which are ~21-24 nt in length. These small RNAs are called microRNAs (miRNAs) and small interfering RNAs (siRNAs). These RNAs guide the Ago proteins to participate in gene silencing events at both transcriptional and post-transcriptional levels (Buhler & Moazed 2007; Filipowicz 2005; Tomari & Zamore 2005). Piwi proteins, on the other hand, are found only in animals and exclusively expressed in the gonads. They bind longer (~26-31 nt) small RNAs called Piwi-interacting RNAs (piRNAs). The mouse genome expresses three Piwi proteins named MIWI (mPIWIL1), MILI (mPIWIL2) and MIWI2 (mPIWIL4) (Aravin *et al.* 2007a; Hartig *et al.* 2007; Kim 2006; Peters & Meister 2007; Seto *et al.* 2007). MILI and MIWI2 are implicated in silencing transposons and repetitive elements in the germline (Aravin *et al.* 2003; Brennecke *et al.* 2007; Gunawardane *et al.* 2007; Houwing *et al.* 2007; Saito *et al.* 2006; Vagin *et al.* 2006). The function of MIWI is presently unknown.

The biogenesis of piRNAs is different from that of mi- and siRNAs. While the latter originate from double stranded RNAs, which are processed by the enzyme Dicer (Ghildiyal & Zamore 2009), piRNAs are believed to be transcribed from discrete gene-poor clusters in intergenic repetitive element or transposon-rich regions (Aravin *et al.* 2006; Girard *et al.* 2006; Grivna *et al.* 2006a; Watanabe *et al.* 2006) in the genome as a single-stranded long precursor RNA (Brennecke *et al.* 2007) in a Dicer-independent manner (Houwing *et al.* 2007; Vagin *et al.* 2006). Activity of unknown nucleases is implicated in cleaving this precursor into mature piRNAs. Most factors involved in the piRNA biogenesis are currently unknown, but recent research has begun to identify Tudor domain-containing proteins and RNA helicases as key factors in the piRNA pathway (Frost *et al.* 2010; Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a; Zheng *et al.* 2010)

While, the exact mechanism by which piRNAs are generated is unknown, the resulting piRNAs carry a 5` monophosphate and a 3` hydroxyl group similar to all known small RNAs (Houwing *et al.* 2007; Saito *et al.* 2006; Watanabe *et al.* 2006). Additionally, piRNAs are 2`-O-methylated at their 3`ends (Horwich *et al.* 2007; Kirino & Mourelatos 2007c; Ohara *et al.* 2007a; Saito *et al.* 2007a; Zhou *et al.* 2010). This modification is reported for example, for fly small RNAs entering AGO2 (mainly siRNAs), as well as for both miRNAs and siRNAs in plants (Boutet *et al.* 2003; Chen 2007; Ebhardt *et al.* 2005; Horwich *et al.* 2007; Li *et al.* 2005; Yang *et al.* 2006b). The 2`-O-methyl group is now regarded as a defining feature of piRNAs in all systems analyzed (Horwich *et al.* 2007; Kirino & Mourelatos 2007c; Ohara *et al.* 2007a; Saito *et al.* 2007a; Zhou *et al.* 2010).

In plants, the 2`-O-methyl group at the 3` end of miRNA and siRNA duplexes is added by the RNA methyltransferase HUA ENHANCER 1 (HEN1). Apart from the Rosmann fold methyltransferase domain, the plant HEN1 has two double-stranded RNA binding domains (dsRBD) and an additional RNA recognition motif. Based on predicted sequence and structural homology, members of the HEN1-family were identified in metazoan, fungi, Viridiplantae and in bacteria. In contrast to the plant HEN1, the animal homologs contain similarity only to the methyltransferase domain (see Figure S 1) (Chen 2007; Tkaczuk *et al.* 2006). Based on genetic studies, the *Drosophila* and fish homologs of HEN1 have been shown to be essential for piRNA methylation (Horwich *et al.* 2007; Kamminga *et al.* 2010; Saito *et al.* 2007a). While the plant HEN1 acts on dsRNAs (Akbergenov *et al.* 2006; Li *et al.* 2005; Yang *et al.* 2006a; Yang *et al.* 2006b; Yu *et al.*

2005), recombinant *Drosophila melanogaster* (*Dm*)HEN1 methylates only single-stranded RNA (ssRNA) substrates. This is consistent with the observed absence of dsRNA precursors and Dicer-independency in the piRNA pathway (Horwich *et al.* 2007; Saito *et al.* 2007a). The *in vitro* RNA methylation activity of recombinant mouse homolog of HEN1 (mHEN1) protein was investigated, and is shown to be capable of methylating single-stranded -but not double-stranded- small RNAs (Kirino & Mourelatos 2007b).

The 3` terminal modification of small RNAs is shown to enhance their stability and is not necessary for their recognition. In *hen1* mutants of *Arabidopsis thaliana* the miRNA and siRNA levels are greatly reduced (Akbergenov *et al.* 2006; Boutet *et al.* 2003; Katiyar-Agarwal *et al.* 2006; Park *et al.* 2002; Vazquez *et al.* 2004a; Vazquez *et al.* 2004b; Xie *et al.* 2003). In fact, the modification protects the 3` termini from an uridylation activity, which usually acts as a signal for destabilization (Chen 2005; Ebhardt *et al.* 2005; Li *et al.* 2005; Yang *et al.* 2006b). Similarly, the loss of *Dmhen1* activity results in reduced length and abundance of piRNAs and siRNAs while miRNAs are not affected (Horwich *et al.* 2007; Saito *et al.* 2007a). Methylation by *Dm*HEN1 also protects fly small RNAs from uridylation and 3` trimming activities that are recruited to small RNAs with extensive complementary targets (Ameres *et al.* 2010). Loss of HEN1 in zebrafish and *Tetrahymena* is shown to destabilize piRNAs and scanRNAs, respectively (Kamminga *et al.* 2010; Kurth & Mochizuki 2009). In fly and fish *hen1* mutants, transposons are slightly activated due to defective piRNA pathway. Importantly, methylation appears not to play any role in loading of small RNAs into an RNP, as in all the mutants described above, small RNAs correctly sort into their respective Argonautes (Horwich *et al.* 2007; Kamminga *et al.* 2010; Saito *et al.* 2007a).

One important question that remains to be addressed is how HEN1 is recruited to specific Argonaute proteins for catalyzing methylation of their associated small RNAs. Fly HEN1 is shown to interact with Piwi proteins but not with *Dm*AGO1 that carries microRNAs, which are not modified by methylation (Saito *et al.* 2007a). However, the nature of the interface mediating this interaction with Piwi proteins is not known. In addition, little is known about the biology of the endogenous HEN1 protein in mouse. Here, the characterization of the RNA methylation activity of endogenous mHEN1 is described. Of the three Piwi proteins, only MILI interacts with mHEN1 in both testes extracts and in HEK293 cells. This interaction surface was mapped to the N-terminus of MILI, to a region proximal to the PAZ domain. Finally, immunofluorescence analysis of mHEN1 localizes the protein to a

cytoplasmic granule distinct from that accumulating other piRNA pathway components. Taken together, these studies identify a physical interaction that recruits mHEN1 to the mouse piRNA pathway and identify compartmentalization of the protein in a distinct cytoplasmic domain.

2.3.4 Results

2.3.4.1 The endogenous mHEN1 is an RNA methyltransferase in mouse testes

Endogenous murine piRNAs (28 nt and 30 nt) carry a 2`-O-methyl group on their 3` terminal nucleotide (Kirino & Mourelatos 2007c; Ohara *et al.* 2007a). A similar 3` terminal modification of plant small RNAs is catalyzed by the RNA methyltransferase HEN1 (Yu *et al.* 2005). Its homologs in fly and fish have been shown to be responsible for modifying piRNAs (Horwich *et al.* 2007; Kamminga *et al.* 2010; Saito *et al.* 2007a). The recombinant version of the mouse homolog of HEN1 (mHEN1) (see Figure S 1) is demonstrated to have methylation activity for small single-stranded RNAs (Kirino & Mourelatos 2007b), but the activity of the endogenous protein is not reported. To characterize the piRNA methylation role of the endogenous mHEN1, antisera was raised against the C-terminal part of the protein in rabbits (GQG and GOY). Western blotting analysis with the affinity-purified antibodies detects a strong band at ~45 kDa corresponding to the expected size of the protein in mouse testes extracts, and recombinant 6XHis-mHEN1 (see Figure 9 A and B Western blot). No signal was detected with pre-immune sera, purified in a similar manner (see Figure 9 A).

Next, the endogenous mHEN1 complexes was immunopurified from adult mouse testes extract and *in vitro* methylation assays were performed, as previously described (Yu *et al.* 2005). Briefly, a mixture of 21 and 30 nt RNAs were incubated with immunopurified HEN1 and ^{14}C-SAM (methyl donor). Subsequently, an aliquot of the reaction mixture was removed at the end of the incubation for 5`-end-labelling to confirm the integrity of RNAs during the experiment. The remaining reaction mixture was processed for detection of $^{14}$C signal due to any RNA methylation activity. The 5` end-labeling experiment reveals that both RNA substrates were intact at the end of the reaction (see Figure 9 B, middle panel). A signal indicating RNA methylation is observed only in the presence of immunoprecipitated mHEN1, whereas the control samples without any added protein, or purifications with an

unrelated antibody remained negative (Figure 9 B, top panel; ^{14}C). Interestingly, only the 30 nt RNA and not the 21 nt RNA substrate is methylated by endogenous mHEN1 (see Figure 9 B, ^{14}C). Similar assays performed with other 21 nt RNA oligonucleotides also showed no 3` terminal methylation in the RNA methylation assay (unpublished data). These data are in contrast to activity assays on the recombinant protein, which was able to methylate RNAs in the range of 20 nt to 40 nt (Kirino & Mourelatos 2007b). It is likely that specific associated factors confer specificity to endogenous mHEN1 complex isolated from testes extracts for 30 nt piRNA-sized RNA substrates. Taken together, these results provide direct evidence that mHEN1 is the piRNA methyltransferase in mouse testis.

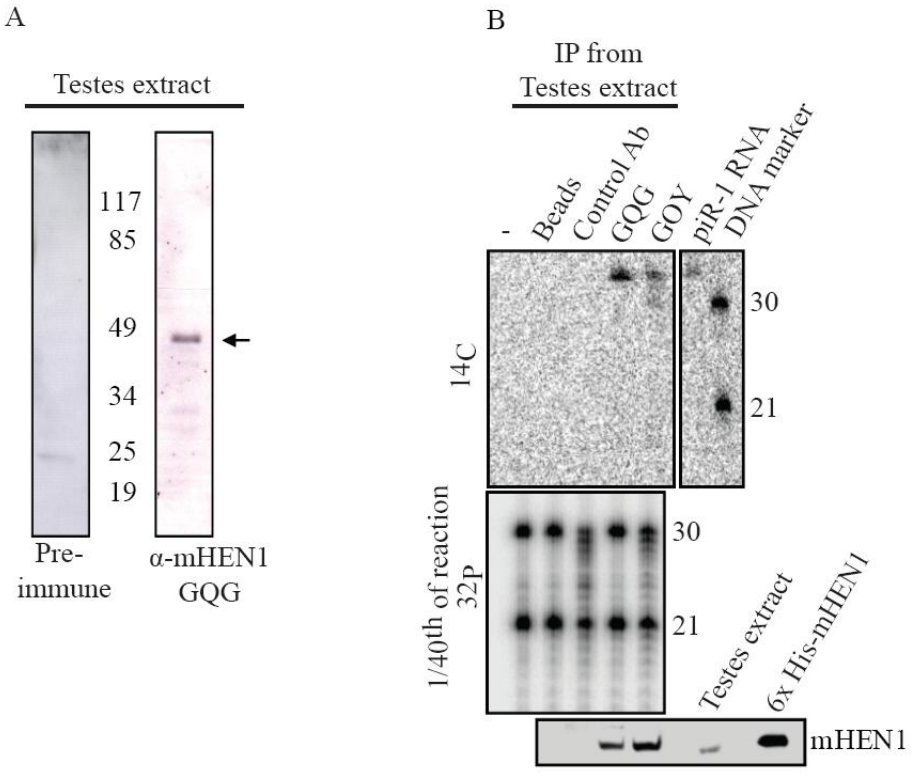


Figure 9:Endogenous mHEN1 in testes extracts is an RNA methyltransferase

(A) Affinity purified mHEN1 antibodies and the pre-immune serum was tested with adult testes extract. The protein markers are in kDa. The antibody specifically detects a band of ~45 kDa (indicated with an arrow).

(B) *In vitro* methylation assays with mHEN-1 immunopurified (IP) (using two independent antibodies GQG, GOY) from adult testes extracts. All reactions contain 21 nt and 30 nt RNAs as substrates and [^{14}C] SAM as methyl donor. In the presence of mHEN1, only the 30 nt RNA is labeled (top panel, ^{14}C) in these reactions. RNA substrates in the reactions remain intact as shown by 5'-end labeling (middle panel). Bottom panel shows Western analysis for mHEN1 (using GQG antibody) with 20 % of the IP reaction. Control Ab, IP using an unrelated antibody; —, reaction with no protein added. Size (in nucleotides) of ^{32}P-endlabelled DNA markers and 30 nt RNA (piR-1 RNA) are indicated.

2.3.4.2 mHEN1 interacts with MILI in mouse testes extract and in the HEK 293 cell line

In *Drosophila* and zebrafish *hen1* mutants, piRNAs are loaded into Piwi proteins, but remain unmethylated (Kamminga *et al.* 2010; Saito *et al.* 2007a). This suggests that recognition of the substrate occurs in the context of the Argonaute proteins. Recombinant *Dm*HEN1 associates with three Piwi proteins (AUB, PIWI and AGO3) in *hen1* mutant extract, and methylates the bound unmodified piRNAs (Saito *et al.* 2007a). In contrast the miRNA-binding AGO1 does not interact with *Dm*HEN1. To directly test interactions of the endogenous mHEN1 to murine Piwi proteins, MILI and MIWI, two of the mouse Piwi proteins present in adult mouse testis, were immunoprecipitated and presence of mHEN1 probed by Western blotting. As shown in Figure 10 A, mHEN1 is robustly detected only when MILI in present. A possible reason for this could be that the interaction with MIWI, which also carries methylated piRNAs, is rather weak and undetectable in the assay conditions. Alternatively, since most of the piRNAs in the endogenous Piwis are already methylated, there might be structural changes in MIWI that hinders re-binding of the methyltransferase.

To further confirm the interaction with MILI in a heterologous system, HA-tagged Piwi proteins (MILI, MIWI, and MIWI2) and hAGO2, a member of the Ago subfamily of Argonaute proteins, that does not bind to methylated small RNAs and should therefore not interact with mHEN1 (negative control), were co-transfected with Myc-tagged mHEN1 in HEK293 cell cultures. The tagged Piwi proteins/hAGO2 were purified with α-HA- beads from lysates prepared 48 hours post-transfection and subjected to Western analyses. Expressions of all the tagged proteins were detected in the input fraction of the lysate. No interaction was detected between mHEN1 and hAGO2, MIWI or MIWI2 (see Figure 10 C). Confirming the detected association between endogenous proteins, Myc-mHEN1 interacts specifically with HA-MILI (see Figure 10 C). The interaction was further confirmed with reciprocal tagging and immunoprecipitations (see Figure 10 D).

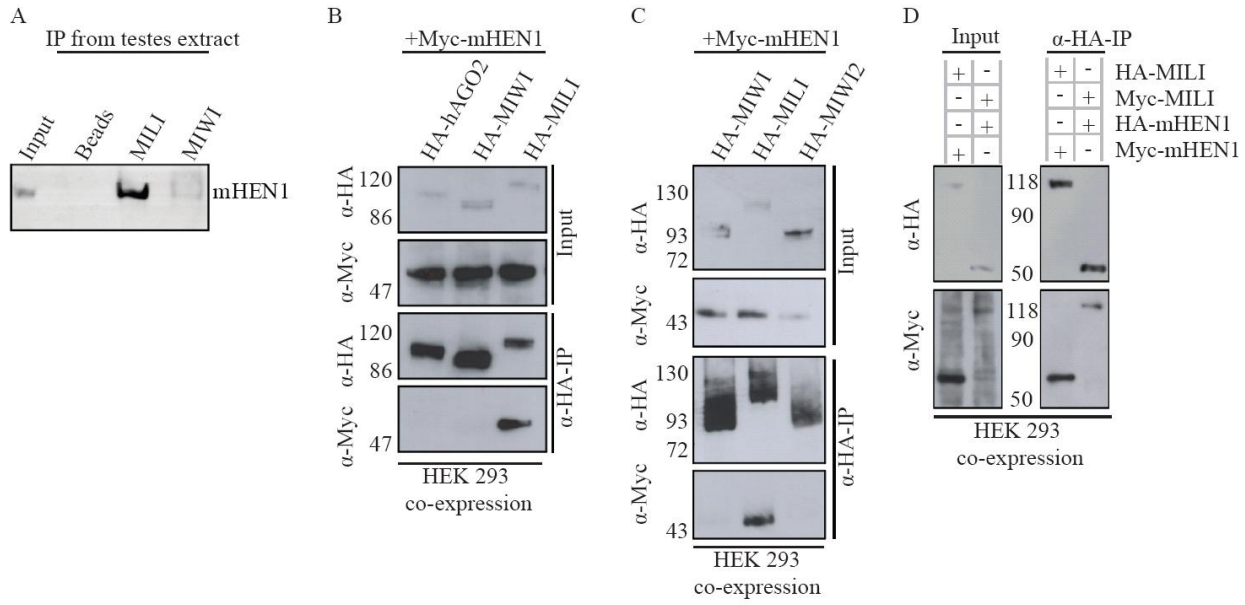


Figure 10: mHEN1 interacts with MILI

(A) Immunoprecipitations with testes extract using α-MILI and α-MIWI antibodies reveals that endogenous mHEN1 interacts only with MILI. Western blot analysis for mHEN1 is with GQG antibody.

(B) Myc-tagged mHEN1 interacts with HA-tagged MILI in HEK293 cell cultures, but not with HA-hAGO2 or HA-MIWI. Indicated HA-tagged fusions of different Argonaute proteins were co-expressed with a Myc-tagged fusion of mHEN1 in HEK293 cells. Western analysis with indicated antibodies of α-HA purifications and input extracts is shown. Input is 5% of the extract used in IPs.

(C) Myc-tagged mHEN1 interacts only with HA-tagged MILI in HEK293 cell cultures and not with HA-MIWI or HA-MIWI2.

(D) Reciprocal interaction of Myc-tagged MILI with HA-mHEN1.

2.3.4.3 Interaction of mHEN1 with MILI is independent of bound RNA and Piwi methylation status

To rule out the possibility that the MILI/mHEN1 interaction is mediated by RNA, the purified HA-MILI/Myc-mHEN1 complex was subjected to treatment with RNase A. This treatment did not affect retention of Myc-mHEN1 on the HA-beads when HA-MILI was present. Again, MIWI failed to interact with mHEN1. This result suggests that complex formation between MILI and mHEN1 is direct and independent of bound RNA. In fact, in immunoprecipitations of mHEN1 from mouse testes extracts, piRNAs were never detected, suggesting an interaction with 'empty' unloaded MILI (see Figure S 2).

Tudor domain-containing proteins interact with Piwi proteins via recognition of symmetrical dimethylated arginines mediated by the Tudor domains. Tudor domain–containing protein 1 (TDRD1) is one such factor which has been shown to interact with modified RG/GR dipeptides present on the N-terminus of MILI (Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a). Since in *tdrd1* mutants, MILI is still loaded with piRNAs, TDRD1 might act subsequent to/or during piRNA loading, which is the same step when mHEN1 interacts with piRNA-loaded Piwi proteins. To determine if TDRD1 interacts as well with mHEN1, HEK293 were co-transfected cells with Myc-TDRD1 and HA-mHEN1, and performed immunoprecipitation with anti-HA-beads. As shown in Figure 11 B, the two proteins do not interact.

Next, we investigated, whether the post-transcriptional modification of MILI could be important for the interaction with mHEN1. HA-MILI/Myc-mHEN1 co-transfections were performed in the presence of the methylation-inhibitor 5`Deoxy-5` (methylthio) adenosine (MTA) or the solvent, DMSO. The HA-MILI/Myc-TDRD1 pair was used as a positive control for methylation-dependent interaction in these experiments. While MILI/TDRD1 interaction is affected by MTA treatment, MILI/mHEN1 interaction was not (see Figure 11 C). Thus, we demonstrated that mHEN1 interacts with unloaded MILI and this interaction is independent of Piwi methylation status.

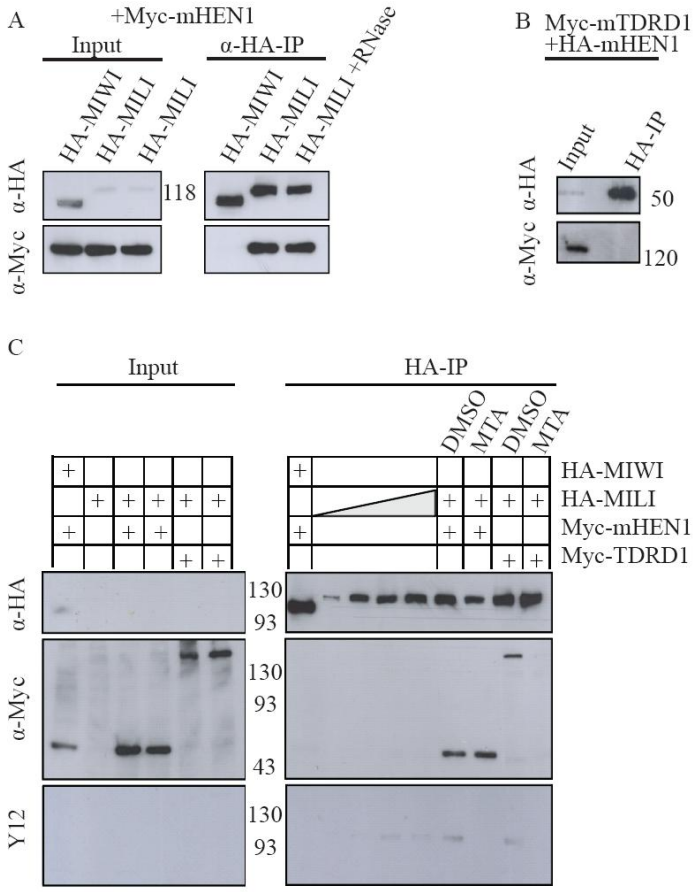


Figure 11: mHEN1 interacts with MILI independent of RNAs and Piwi methylation status

(A) mHEN1/MILI interaction is RNA-independent. HEK293 cells were co-transfected as indicated and subject to α-HA immunoprecipitation, followed by RNaseA treatment where mentioned. Immunoprecipitates were then analyzed by anti-Myc Western blotting**.**

(B) mHEN1 andTDRD1 do not interact with each other

(C) HA-MIWI and -MILI are co-expressed with Myc-mHEN-1 and treated with 5`Deoxy-5` (methylthio) adenosine (MTA) or DMSO when indicated. After α-HA-purification HA-MIWI and Myc-mHEN1 do not show any interaction. For controlling the detection sensitivity a HA-MILI gradient is marked. MTA treated samples show still the interaction of MILI and mHEN1 (middle panel) but no methylation (Y12 panel).

2.3.4.4 The mHEN1/MILI interaction surface brings the methyltransferase domain in close proximity to the 3` end of the piRNA

The N-terminus of MILI is shown to be an interaction site for other proteins, for example TDRD1 (Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a). To gain insight into whether the mHEN1 interaction with MILI is mutually exclusive with TDRD1, full-length HA-MILI or an N-terminal deletion construct of MILI were co-expressed with Myc-tagged mHEN1 in HEK 293 cells. The HA-tagged proteins were purified with α-HA-beads from lysates prepared 48 hours post-transfection and subjected to Western analyses. While the full-length HA-MILI co-immunoprecipitates Myc-mHEN1, the N-terminus deletion did not show any interaction with the methyltransferase nor did the HA-tagged MIWI pull down mHEN1 (see Figure 12 A). From this experiment, we can conclude that mHEN1 interacts with the N-terminal region of MILI.

To precisely define the binding surface of mHEN1 and MILI, deletion constructs of HA-tagged MILI carrying increasingly shortened N-termini (see Figure 12 B) were co-expressed with full-length Myc-mHEN1 in HEK 293 cells. Western blot analysis of the HA-beads purified cell extract indicated that deletions of the first 221 aa of HA-MILI does not affect the binding of Myc-tagged mHEN1, while deletions starting with amino acid position 236 or more abolish the interaction (see Figure 12 C). Thus, the interaction surface for mHEN1 is in a region proximal to the PAZ domain of MILI. The binding site could be a composite surface generated by the PAZ domain and N-terminal region proximal to the PAZ domain.

Then the minimal N-terminal deletion fragment that retains interaction with mHEN1 was examined by comparing known HEN1 binders (PIWI, *Dm*AGO3, AUBERGINE,) and hypothetical binding proteins (DmAGO2, the zebrafish ZILI and ZIWI, the *Bombyx mori* *Bm*AGO3 and SIWI) to non-binders (*Dm*AGO1) in a sequence alignment to investigate possible motif conservation (see Figure 12 D). The minimal N-terminal region of MILI reveals 7 amino acids that are strictly conserved in all Argonaute proteins that are known to bind methylated small RNAs. Argonautes that bind unmethylated small RNAs lack this consensus sequence. The significance, if any, of this sequence motif is presently unknown as they are widely separated, but perhaps in the folded protein these amino acids might form a defined interaction surface. Structural studies of the MILI/mHEN1 complex should shed light on this question.

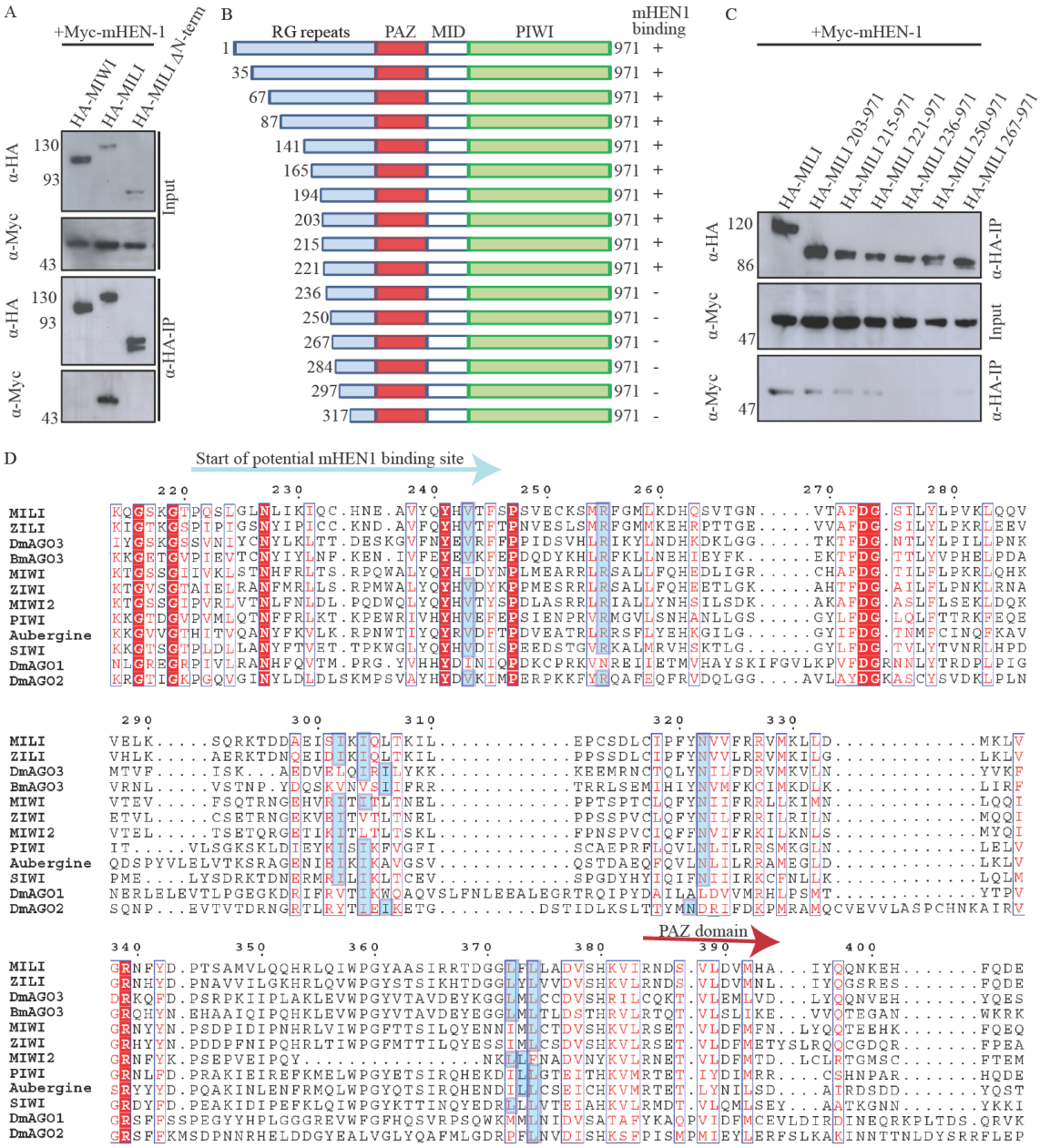


Figure 12: The mHEN1 interaction site on MILI is close to the PAZ domain

(A) mHEN1 interacts with the N-terminus of MILI. HEK293 cells were co-transfected as indicated and subject to anti-HA immunoprecipitation, followed by analysis by anti-Myc Western blotting**.**

(B) Schematic view of the N-terminal deletion constructs of MILI. On the right side interactions with mHEN1 are indicated (+ as positive, - as negative)

(C) Indicated HA-tagged deletion constructs of MILI were co-transfected with Myc-mHEN1 and HA-purified. The interaction between mHEN1 and the MILI deletions were analyzed by anti-Myc Western blot staining. MILI deletions of more than the first 221 aa do not interact with mHEN1 any longer.

(D) Alignment of the potential mHEN1 binding site in the N-terminus of MILI with the same region of HEN1 interacting Piwi proteins (PIWI, *Dm*AGO3, AUBERGINE), potential interacting Argonaute proteins (DmAGO2, ZILI and ZIWI, *Bm*AGO3 and SIWI) and the N-terminus of DmAGO1 as an example for an Argonaute protein which does not interact with HEN1 aligned using ClustalW. The Alignment was analyzed using ESPript. In this region seven amino acids (in blue) are conserved in all Argonaute proteins that are known to bind methylated small RNAs (indicated in red boxes; for MILI: V243, R255, I302, I304, N322, L372 and L374). Argonautes that bind unmethylated small RNAs lack them.

The region of mHEN1 important for the interaction with MILI was also investigated. It was shown that the C-terminal domain of the zebrafish *Dr*HEN1 is involved in the correct localization of the protein to cytoplasmic RNA processing centers, called germ granules. This suggests that this region is bound by other factors, probably a Piwi protein for proper localization (Kamminga *et al.* 2010). It was speculated that the recruitment of the *D*rHEN1 could be due to a Piwi protein, although it was not shown. To examine this hypothesis, several C-terminal deletion constructs of Myc-tagged mHEN1 were constructed (see Figure 13 A) and co-transfected with HA-tagged MILI in HEK293 cell. After HA-beads purification and Western blot analysis with α-Myc antibodies, we demonstrated that constructs with C-terminal deletions longer than 303 aa did not affect the interaction of mHEN1 to MILI (see Figure 13 B). However, mHEN1 constructs with shorter C-termini (1-234 aa) did not show interactions with the co-expressed HA-MILI. In conclusion, the far end of the C-terminus of mHEN1 (303- 395 aa) is not involved in the interaction to MILI. In assumption that the C-terminus of *Dr*HEN1 is needed to bind to the fish Piwi proteins for correct localization and that the methyltransferase domain is not needed for this interaction, the region of the *Dr*HEN1 minimal localization construct (Kamminga *et al.* 2010) was aligned with the minimal C-terminal mouse mHEN1, which still was able to bind to MILI. These data suggest that the MILI binding region on mHEN1 is between 234-303 aa. The close proximity to the methyltransferase domain might facilitate the methylation of the 3` end of the piRNAs bound in the PAZ domain of MILI.

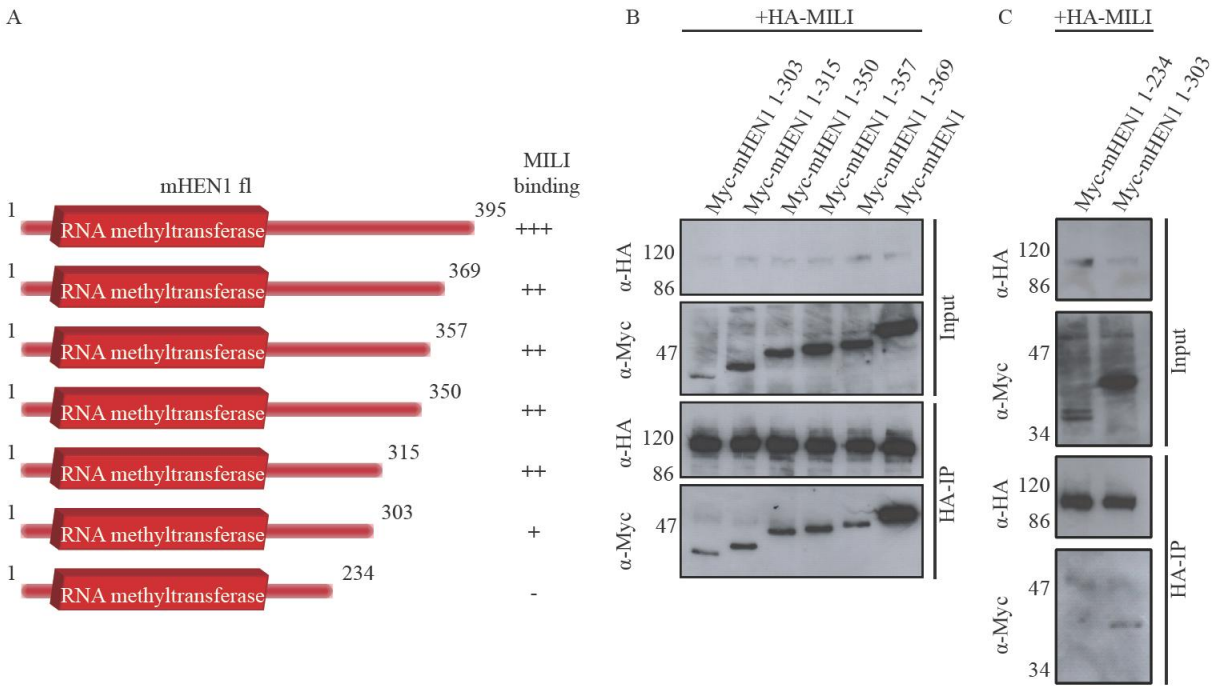


Figure 13: MILI interacts with the C-terminus of mHEN1 close to methyltransferase domain

(A) Domain alignment of the full-length and deletion constructs of mHEN1 used in the following experiments. The intensity of the interaction is indicated by (strong; +++,++,+ and -, none)

(B) Myc-tagged mHEN1 and the indicated deletions were co-expressed with HA-MILI in HEK 293 cells. After purification via HA-beads, the samples were analyzed by Western blot analysis using α-Myc antibody. All C-terminal deletions until 303 aa are interacting with MILI.

(C) Myc-tagged mHEN1 1-234 and Myc-tagged mHEN1 1-303 were co-expressed with HA-MILI in HEK293 cells. The samples were purified using HA-beads and analyzed via Western blot using α-Myc antibodies. While the Myc-tagged mHEN1 1-303 construct is still co-purified by HA-MILI, Myc-tagged mHEN1 1-234 does not bind to HA-MILI.

2.3.4.5 mHEN1 is a cytosolic protein with a localization distinct from other piRNA pathway components

Since mHEN1 is involved in 3` end maturation of piRNAs and directly interacts with MILI, their expression pattern was examined in mouse testes sections by indirect immunofluorescence. In seminiferous tubules of the male gonads, spermatogenesis proceeds through mitotic and meiotic divisions of the germ cells to give rise to the haploid round spermatids (RS), which undergo further development to form mature sperm. mHEN1 is detected in both meiotic pachytene spermatocytes (PS) and RS. In both cell types, the enzyme is cytosolic, with RS showing a single punctuate spot (see Figure 14 A). To investigate this pattern in more detail, the localization of mHEN1 and different components of the piRNA machinery were performed using specific antibodies in purified germ cells (Pivot-Pajot *et al.* 2003). The murine Piwi proteins MILI and MIWI are reported to be cytoplasmic with maximal expression seen in PS and RS respectively (Deng & Lin 2002; Kuramochi-Miyagawa *et al.* 2008). In isolated meiotic PS mHEN1 was detected together with the pachytene cell marker SCP3, a component of the axial element of the synaptonemal complex which becomes associated with the two sister chromatides of each homolog (Lammers *et al.* 1994; Moens *et al.* 1987) (see Figure 14 B). mHEN1 is diffusely distributed in the cytoplasm, with occasional granules being present. In RS, the mouse VASA homolog (MVH) (Kotaja *et al.* 2006; Tanaka *et al.* 2000) is reported to be present in a single large perinuclear structure called the chromatoid body (CB) (see Figure 14 C). MIWI and MILI are also reported CB components (Aravin & Bourc'his 2008; Aravin *et al.* 2008; Wang *et al.* 2009a), which was confirmed by co-staining experiments. Detection of mHEN1 in RS indicates its localization in a single cytoplasmic focus, which is slightly smaller than that observed for the Piwi proteins (see Figure 14 C). Surprisingly, co-staining with MILI reveals that the focus stained by anti-mHEN1 antibody is distinct from that stained by MILI (see Figure 14 D third row). This separate sub-structure was called the “HEN body”. Examination of multiple cells and in different experiments confirms this distinct localization pattern for mHEN1, but also indicates a general tendency of the HEN body to remain in close proximity to the CB. The functional relevance of this sub-cellular compartmentalization is presently unknown, but serves probably temporal regulation of the interaction. Immunoprecipitations of MILI revealed in Western blot analysis with α-mHEN1 antibody that the interaction of MILI and

HEN1 is present in PS and in RS, although the localizations of the two proteins are distinct (see Figure 14 F).

Figure 14: Cytoplasmic localization of mHEN1
mHEN1 localizes in a singular granule distinct from the chromatoid body in round spermatids

(A) Adult mouse testes section stained with α-mHEN1 (red) and DAPI (blue). White arrow indicates round spermatids (RS) and red arrow marks a pachytene spermatocyte (PS)

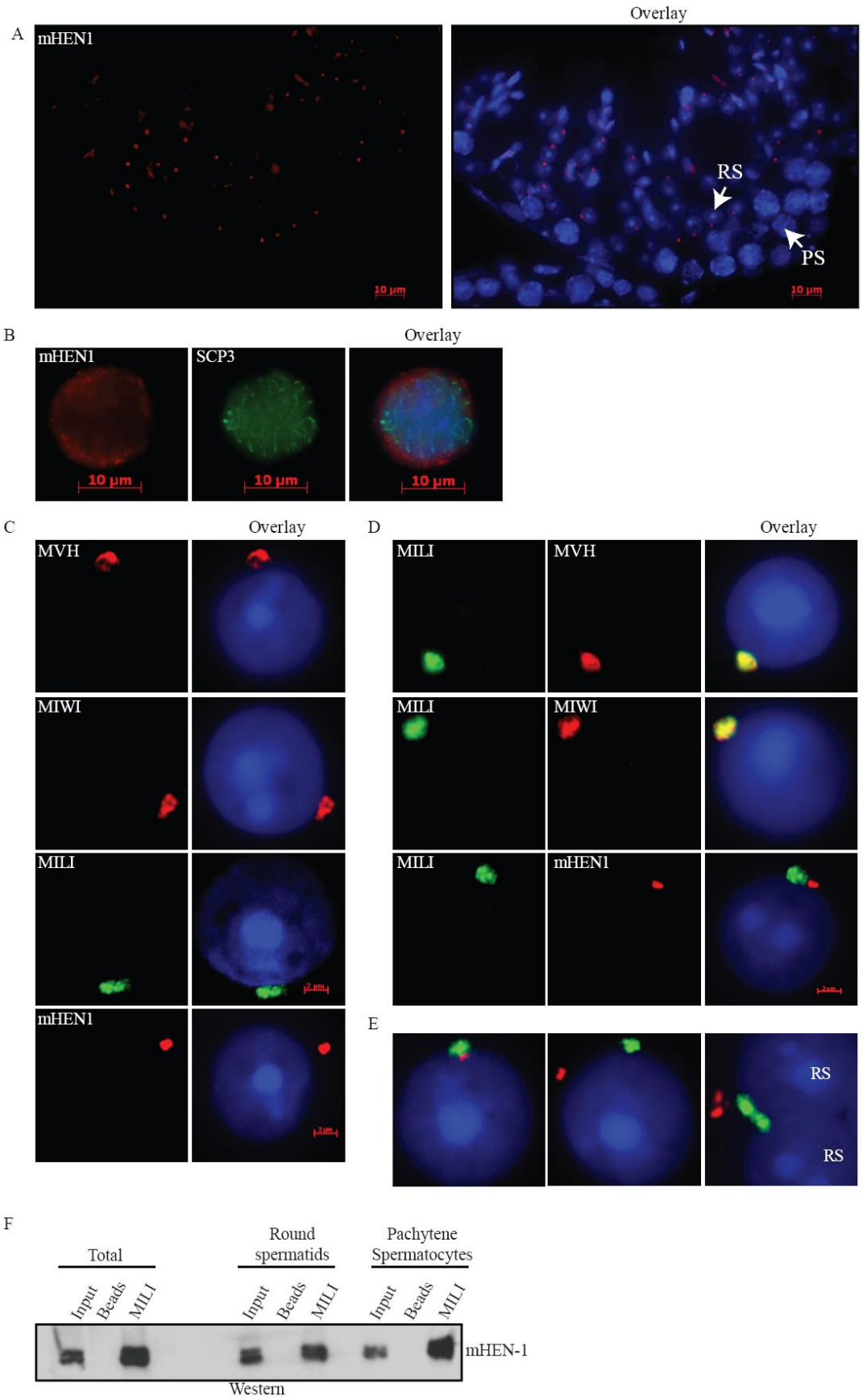
(B) Pachytene spermatocytes stained with mHEN-1 (red) and SCP3 (green)

(C) Purified RS stained with indicated antibodies. Overlay shows merged channels with DAPI staining (blue) for DNA.

(D) MIWI and MILI are in the chromatoid body (CB) while mHEN1 forms a distinct granule. Double staining with anti-MILI mouse monoclonal antibody (green) and rabbit polyclonal antibodies for indicated proteins (red). Overlay shows merged channels with DAPI staining (blue) for DNA.

(E) A few more representative examples of MILI/mHEN1 overlay images are shown. One of the panels shows two adjacent RS, each with a CB and a HEN body.

(F) Immunoprecipitations using α-MILI antibodies and as a control empty beads revealed that in RS as well as in PS MILI co-precipitates mHEN1.


A
mHEN1
Overlay
RS
PS
10 µm
B
mHEN1
SCP3
Overlay
10 µm
C
MVH
MIWI
MILI
mHEN1
Overlay
D
MILI
MVH
MIWI
mHEN1
Overlay
E
RS
RS
F
Total
Round spermatids
Pachytene Spermatocytes
Input
Beads
MILI
mHEN-1
Western

2.3.5 Discussion

HEN1 homologs catalyze addition of a methyl group to the 2`OH at the 3` end of small RNAs. Here, evidence was presented that the endogenous mouse homolog of HEN1 (mHEN1) from testis is an RNA methyl transferase. Its interaction with one of the Piwi proteins, MILI, was shown both in testis extracts and in a heterologous cell culture system. This interaction surface was mapped on the N-terminus of MILI to a region proximal to the PAZ domain and to the C-terminus of mHEN1 in vicinity of the methyltransferase domain. Finally, the endogenous protein was localized in seminiferous tubules and isolated germ cells to show that mHEN1 is a component of a nuage, distinct to the chromatoid body where other piRNA biogenesis factors accumulate. Together, these results provided evidence implicating the endogenous mouse homolog of HEN1 in the murine piRNA pathway.

In *Arabidopsis*, both ~21 nt siRNAs and miRNAs carry a 3` terminal methyl group. Since these RNAs use double-stranded RNA (dsRNA) precursors, the plant HEN1 has at its N-terminus dsRNA-binding domains for recognizing small RNA duplexes for methylation. In agreement with this, the enzyme is inactive on single-stranded RNA substrates (Yu *et al.* 2005). Animal homologs are much smaller than the plant protein, and contain only the conserved SAM-domain required for RNA methylation (Chen 2007; Tkaczuk *et al.* 2006). Consistent with the lack of dsRNA precursors in the piRNA pathway, animal HEN1 proteins do not act on dsRNA substrates (Horwich *et al.* 2007; Kirino & Mourelatos 2007a, b; Saito *et al.* 2007a). The question is how the animal protein specifically picks its substrates among other single-stranded RNAs in the cell. One possibility is the variable sizes of piRNA substrates HEN1 encounters in animal systems.

In mouse for example, all Piwi proteins recognize piRNAs of a specific length (MILI binds to ~26 nt, MIWI2 to ~28 nt and for MIWI associates with ~29 nt piRNAs) (Aravin *et al.* 2006; Aravin & Bourc'his 2008; Girard *et al.* 2006; Grivna *et al.* 2006a; Grivna *et al.* 2006b). Since the Piwi proteins are the only factors capable of discriminating these minor size differences in piRNAs, it is reasonable to expect that methylation of piRNAs might occur in the context of the RNA being bound to the Piwi protein. Supporting this idea is the finding that in fly *Dmhen1* mutants, piRNAs remain unmethylated, but are loaded correctly into Piwi proteins. In fact, supplementing purified Piwi complexes from such mutants with

recombinant fly *Dm*HEN1 allowed piRNA methylation to proceed. Modification of siRNAs also can be rescued in the *Dmhen1* mutant ovary extracts by addition of recombinant *Dm*HEN1, but only after unwinding of the siRNA duplex intermediate and loading of the single-stranded siRNA into fly *Dm*Ago2 (Horwich *et al.* 2007). Additionally, *Dm*HEN1 did not modify *Drosophila* Ago1 bound miRNAs, suggesting specific requirements for protein-mediated recruitment of HEN1 or accessibility of the enzyme to the small RNA 3`end.

In this study, the interaction of mHEN1 and MILI was demonstrated in both testes extracts and also in a heterologous cell culture system. Similarly, recombinant fly *Dm*HEN1 interacts with endogenous Piwi proteins in *Dmhen1* mutant extracts, indicating that it is likely that the Piwi proteins recruit the enzyme to the associated piRNAs (Saito *et al.* 2007b). Surprisingly, our results did not show an interaction between mHEN1 and MIWI or MIWI2. Very likely such an interaction might occur only with a piRNA loaded Piwi protein, containing an unmethylated RNA, similar to that used in the fly *Dmhen1* mutant study. Similar experiments in mouse will have to wait for the availability of a mouse *mhen1* mutant.

Accessibility of the small RNA 3` end for methylation can be influenced by at least three factors: first, the affinity of the PAZ domain for the 3` end in its modified or unmodified form. Recently it was shown that the MIWI PAZ domain has a slightly increased (2-fold) affinity for a methylated 3` end. Further it was revealed by structural studies that the methyl-group takes the position of the last nucleotide in the MIWI PAZ domain compared to a hAGO1 PAZ domain (Simon *et al.* 2011), suggesting that the methyl group could bind the piRNA deeper into the binding pocket. Because an unmethylated RNA binds more weakly than a methylated piRNA, the unmodified piRNA might have its 3` end exposed and accessible to HEN1 for methylation. A second factor contributing to the accessibility of the 3`end of the piRNA could be the extent of base-pairing of the small RNA uses to recognize its targets. Crystal structures of Argonaute-guide RNA-target RNA ternary complex reveal that extensive complementarity forces the PAZ domain to let go of the small RNA 3` end to facilitate the formation of the required guide-target A-helix (Filipowicz 2005). This possibility is supported by the intricate link between piRNA function in transposon mRNA destruction and (secondary) piRNA biogenesis. In this context, the mapping of the interaction surface on MILI to a region close to the PAZ domain is interesting and might offer yet a third possibility. This interaction might force a conformational change near the PAZ domain, enabling the release of the unmethylated and weakly-bound small RNA 3` end. After

methylation, a stronger interaction with the PAZ domain might result in the release of the Piwi protein. All these possibilities will need to be tested by future structural and biophysical analyses.

The localization studies of mHEN1 indicate a cytoplasmic localization of the piRNA during the maturation process. All currently known processing factors and most Piwis are components of the cytoplasmic perinuclear granules (called nuage in worms, flies and mice) (Kotaja & Sassone-Corsi 2007). In mouse, only MIWI2 is mostly nuclear, but it is also observed in nuage structures, called pi-P-bodies in the cytoplasm of gonocytes (Aravin & Bourc'his 2008; Aravin *et al.* 2009). In fact, interfering with MIWI2 loading results in a cytoplasmic localization (Aravin *et al.* 2009; Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a), reinforcing the hypothesis that piRNA biogenesis is a cytoplasmic process and only when MIWI2 is properly loaded can it enter the nucleus. In flies too, where only PIWI is nuclear, depletion of piRNA biogenesis factors like Armitage, Yb or Zucchini lead to a cytoplasmic localization (Olivieri *et al.* 2010; Saito *et al.* 2009). Taken together, germline nuage appears to be the site of actual piRNA production.

It was observed that mHEN1 shares the same localization pattern as MILI in pachytene stage cells. This result is similar to the localization pattern of ZIWI and *Dr*HEN1 in zebrafish, which are both localized in the nuage (Kamminga *et al.* 2010). However, here it is shown that in later stages of the spermatogenesis in mouse, in round spermatids, mHEN1 is in a separate body in the vicinity of the chromatoid body. The novel granule structure was called a HEN-body. The spatial separation from the CB raises the question whether or not there is a functional role for this body in piRNA biogenesis.

Regardless of the different localizations of MILI and mHEN1 in haploid round spermatids, the interaction of the two proteins that was reported here remains unchanged both in pachytene spermatocytes and round spermatids. Modification of piRNAs is also not different in these cells as all piRNAs in mouse total testis RNA are methylated (Kirino & Mourelatos 2007c; Ohara *et al.* 2007a). Thus, accumulation of mHEN1 in a distinct body is unlikely to have a direct role in piRNA biogenesis. One possible function of the HEN body is to sequester mHEN1 away from the CB which contains miRNAs and other stored developmentally important mRNAs (Kotaja & Sassone-Corsi 2007). Any promiscuous methylation of its RNA components might be disastrous for developmental processes, as it can alter the regulated half-lives of these RNAs. Another possible explanation is that given

that mHEN1 is a strong interactor with MILI, the protein has to be locked in a separate body in order to not interfere with the silencing function of the effector MILI complex. Further characterization of mHEN1-interacting factors will shed light on the contents of the HEN body and the role of the HEN1 localization.

2.3.5.1 Model of mHEN1 activity

We propose a model of mHEN1 activity based on our results and currently available published data (see Figure 15): Our results have shown that post-translational modifications in the form of symmetrically dimethylated arginines in the N-terminus of MILI, catalyzed by the methyltransferase PRMT5 and its associate Piwi binding protein WDR77, are not needed for the interaction with mHEN1. In *Drosophila dprmt5* null mutants, piRNAs are still methylated (Kirino *et al.* 2009), confirming that HEN1 performs its activity independent of the post-translational modifications of the PIWI proteins. In fly and fish *hen1* mutants the piRNA pathway is only slightly impaired in its function to silence transposable elements (Horwich *et al.* 2007; Kamminga *et al.* 2010; Saito *et al.* 2007a). This presumes that the Piwi proteins must be methylated on their N-terminal arginines in the *hen1* mutant, because otherwise a stronger activity of the transposons would be expected as shown for fly *dprmt5* mutants (Kirino *et al.* 2009) or in mutants of proteins which bind dependent on the symmetrical dimethylated arginines of Piwi proteins like for example TDRD1 in mice (Reuter *et al.* 2009; Wang *et al.* 2009a). Immunoprecipitations using an antibody against symmetrical dimethylated arginines or against TDRD1 co-purified matured piRNAs (Reuter *et al.* 2009). However, piRNAs were never observed in immunoprecipitations of mHEN1. Additionally, mHEN1 interacts with MILI in an RNA-independent manner, suggesting that the mHEN1/MILI interaction can occur before the complex contains RNAs and before MILI is post-translationally modified. Evidence shows that HEN1 is not required for piRNA loading onto Piwi proteins (Kamminga *et al.* 2010; Saito *et al.* 2007a). These data suggest a model where the methylation of piRNAs occurs before the post-transcriptional modification of Piwi and the subsequent complex arrangement including TDRD1.

We found that mHEN1 is localized in a distinct granule in close proximity to the chromatoid body where MILI is localized. We speculate that mHEN1 leaves the MILI complex after the methylation of the piRNAs in order to allow formation of MILI-containing

multi-protein complexes. One component of this complex would be TDRD1, with which MILI associates and co-localizes throughout development (Reuter *et al.* 2009). This localization could as well serve to separate mHEN1from other small RNAs, like miRNAs which are present in the chromatoid body (Kotaja *et al.* 2006; Meikar *et al.* 2010).

Thus, the action of mHEN1 can be placed after the loading of the piRNAs onto MILI and before or at the same time as the post-translational modification of MILI. After piRNA methylation, mHEN1 likely leaves the complex and separates into a distinct cytoplasmic granule.

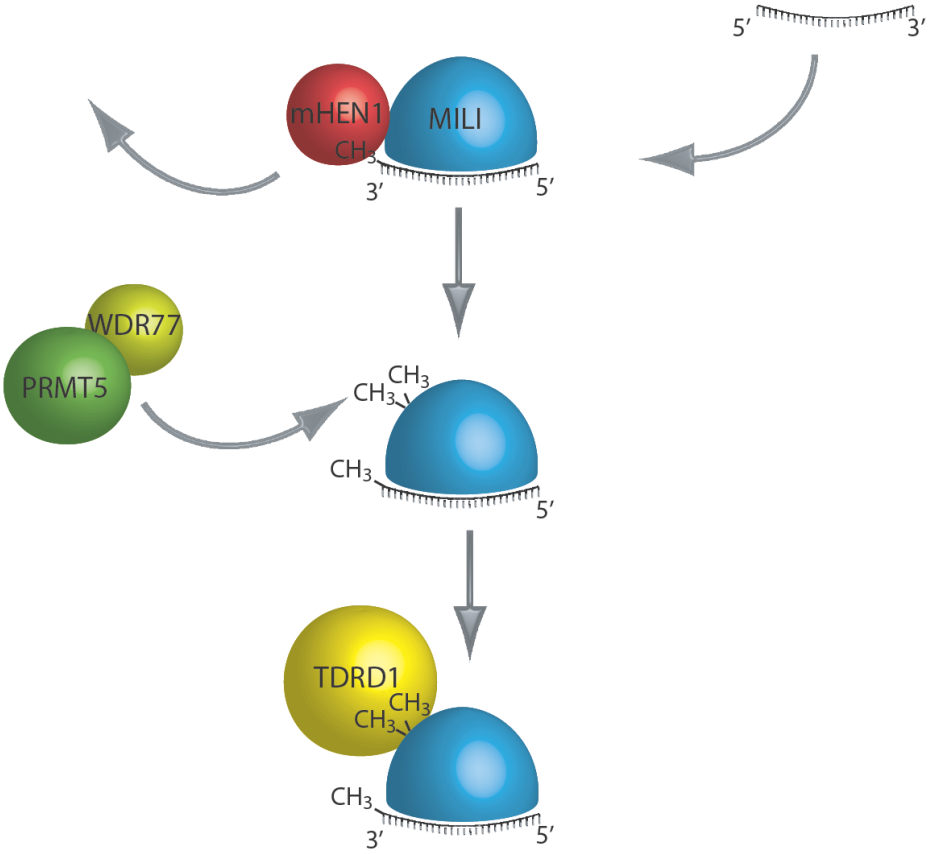


Figure 15: Model
mHEN1 (red) is not needed for the loading of MILI (blue), but can interact with the unloaded protein. After loading of MILI with piRNAs mHEN1 methylates the bound RNA and leaves the complex. Then PRMT5 (green) guided by WDR77 (olive) methylates the arginines in the N-term of MILI. This methylation is essential for binding of TDRD1 (yellow).

2.3.6 Acknowledgements

We thank T. Noce for anti-MVH antibody. We thank EMBL Monoclonal Antibody Core Facility and Protein Expression and Purification Facility for production of antibodies. Work in RP laboratory is supported by the EMBL.

2.3.7 Methods

2.6.1 Antibodies

For the production of antibodies, GST-fusions of specific protein fragments were produced in *E. coli* strain BL21 and injected into rabbits. The fragments used were anti-mHEN1 (GQG and GOY, 185-395 aa) and anti-MIWI (BVI, 1-200 aa; kindly provided by Dr Michael Reuter). The anti-MILI monoclonal antibody (mAb, 13E-3; kindly provided by Dr Michael Reuter) recognizes the peptide corresponding to 107-122 aa of MILI and was raised at the Monoclonal Antibody Facility, EMBL. Control antibodies used in IPs were rabbit polyclonal antibodies directed against the Glutathione-S-Transferase (GST) moiety. All antibodies were affinity-purified.

2.6.2 Mouse testes extracts and RNA analysis

CD1 mouse testes extract preparation, immunoprecipitation and 5`-end-labelling of piRNAs were performed as described previously (Aravin *et al.* 2006; Girard *et al.* 2006). Dried gels were exposed to a storage phosphor screen and scanned with a Typhoon scanner (Amersham).

The *in vitro* methylation assays was performed as described (Yu *et al.* 2005). Endogenous mHEN1 from mouse testes extract was immunopurified (using antibodies chemically-cross linked to Protein G sepharose) and bead-bound protein was used in the assay. After the incubation period, an aliquot (1/40th) of the *in vitro* methylation reaction was 5'-end labeled with ^{32}P to verify the presence of both substrate RNAs.

The following RNAs were used in the *in vitro* methylation assay:

TGACATGAACACAGGTGCTCA (21 nt)

TGACATGAACACAGGTGCTCAGATAGCTTT

(30 nt, the same sequence as piR-1 piRNA). The 21 nt RNA is a 3' end-truncated version of the 30 nt RNA.

2.6.3 Clones and cell culture

For mammalian cell transfections, full-length ORFs of Argonaute proteins (MILI, MIWI and MIWI2) or mHEN1 (Acc. no. AK038994, RIKEN), or deletion constructs were inserted into pCIneo-HA vector (Pillai *et al.* 2004) or into pcDNA3-Myc (Invitrogen), as required.

For HEK293 transfections, 2 µg of each construct of interest (pCIneo-HA-Piwi or mHEN1 plasmids and/or pcDNA3-Myc-mHEN1 or Piwi protein vectors) were used and transfected into 6 cm dishes at 80% confluence using Lipofectamine/Plus (Invitrogen) following the manufacturer's protocol, and harvested 48 hours post-transfection. When mentioned, transfected cells were treated with methyltransferase inhibitor 5`-deoxy-5`-(methylthio) adenosine (MTA; Sigma) at a final concentration of 750 µM or the solvent DMSO alone. Cells from one 6 cm plate were lysed and mixed with 10 µL α-HA-beads (anti-HA affinity matrix beads (Roche) following the purification in the manufacturer's protocol. When mentioned, purified samples were treated with RNase A (20 µg) for 15 min at 37 °C, followed by three washes.

Western analysis was performed with affinity-purified antibodies to specific proteins at a dilution of 1:250. Other antibodies used were: anti-HA antibodies (Santa-Cruz), anti-Myc and Y12 at 1:250 dilutions. An aliquot of 5% input material was probed with anti-HA antibody to control for mHEN1 expression in all reactions.

2.6.4 Immunofluorescence

For tissue sections, testes from 20-day old CD10 mice were fixed in 4% paraformaldehyde. Staining with anti-mHEN1 (used at 1:50) was performed on 7 µm sections using microwave antigen retrieval in 10 mM citrate buffer pH 6.0 and detected by immunofluorescence with anti-rabbit Alexa 488 (Invitrogen). The tissues were mounted with Vectrashield (Reactolab) containing DAPI and visualized using 100x objective of a Zeiss Axiovision Z1 wide-field microscope.

Preparation of purified cell populations by sedimentation (Pivot-Pajot *et al.* 2003) and fixation to glass slides as well as the indirect immunofluorescence detections were performed

as described (Kotaja *et al.* 2006). The following affinity-purified primary antibodies were used at 1:4 dilutions: anti-MIWI (BVI), anti-mHEN1 (GQG), anti-MILI (13E-3). Anti-MVH was used as a dilution of 1:1000 (kindly provided by Toshiaki Noce, Mitsubishi Kagaku Institute of Life Sciences, Tokyo, Japan). The secondary antibodies anti-rabbit IgG Alexa Fluor 594 and anti-mouse-IgG Alexa Fluor 488 (Invitrogen) were both used at 1:10,000 dilutions. Cells were mounted with Vectrashield (Reactolab) containing DAPI and visualized using 100x objective of a Zeiss Axiovision Z1 wide-field microscope. Images were taken with Axiocam camera, processed with manufacturer's software and mounted with Photoshop.

2.3.8 Supplementary data

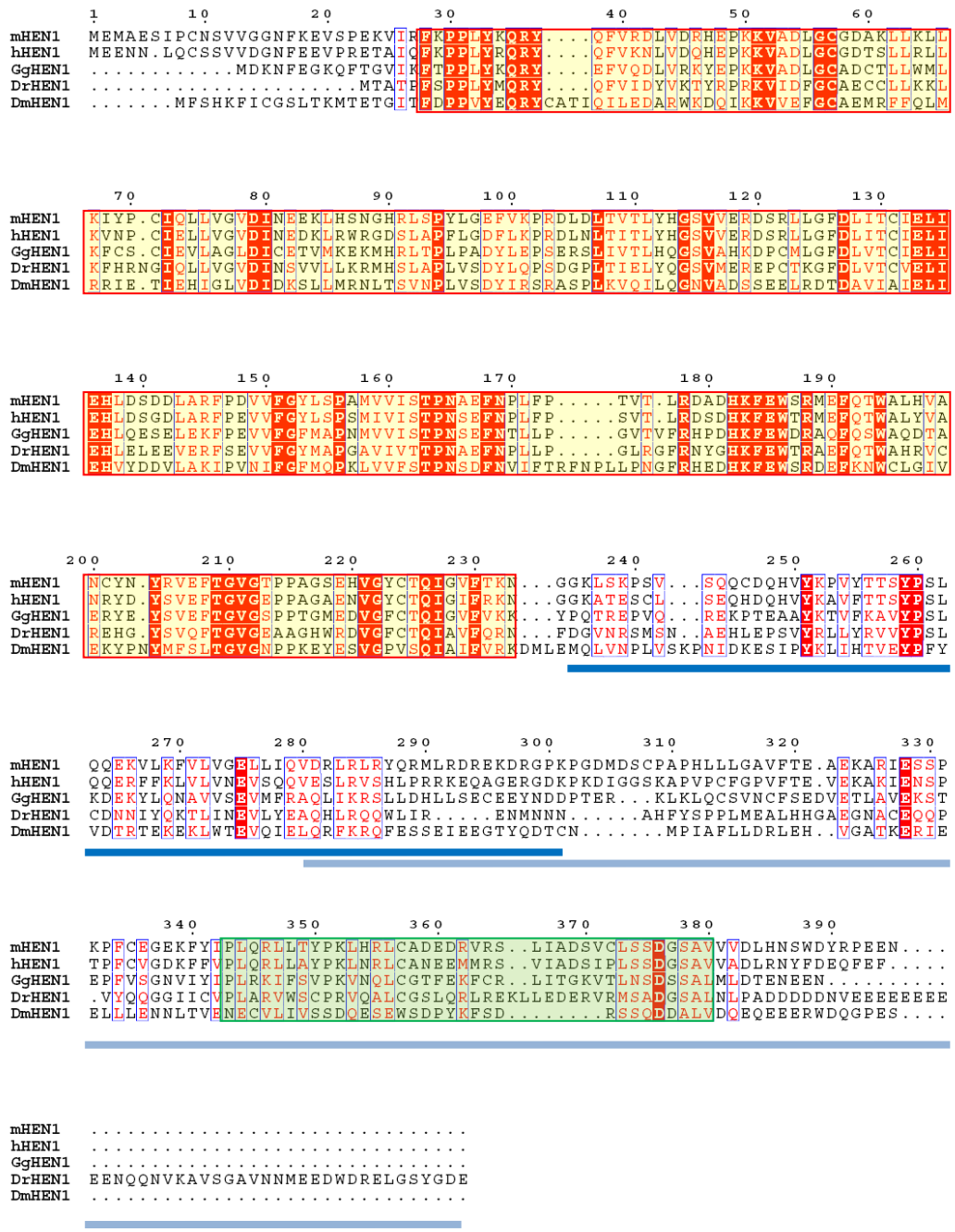


Figure S 1: Alignment of HEN1 from different animals
Sequence alignment of full-length animal HEN1 homologs aligned using ClustalW. The Alignment was analyzed using ESPript. The figure is based on (Huang *et al.* 2009; Kamminga *et al.* 2010)

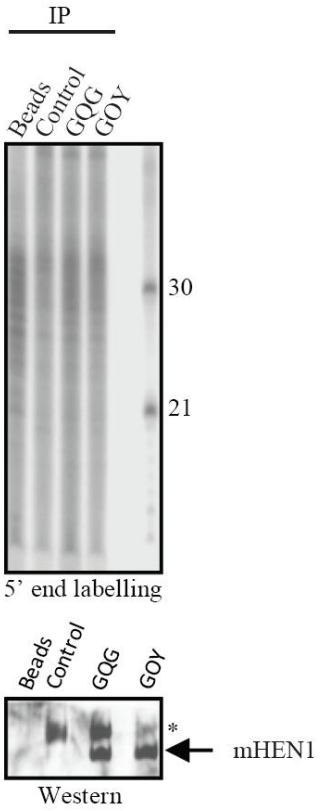


Figure S 2: Endogenous mHEN1 from testes extracts is devoid of any associated small RNAs
mHEN1 was immunoprecipitated (IP) using antibodies (GQG or GOY) and analyzed by 5'-end-labelling and Western (GQG). Control, IP with an unrelated antibody; beads, pull down without any antibodies; asterisk, heavy chain of antibodies. Size (in nucleotides) of RNA markers is indicated.

2.4 Additional experiments for the characterization of mHEN1/MILI complex

2.4.1 Attempt to express the mHEN1/ MILI complex in insect cells

2.4.1.1 Aim

RIWI, a Piwi protein in rat (Lau *et al.* 2006), as well as the *Drosophila* Piwi proteins are described as endonucleases, which are able to slice their target RNAs guided by the loaded piRNAs (Gunawardane *et al.* 2007; Nishida *et al.* 2007; Saito *et al.* 2006). Based on the assumption that MILI and MIWI2 are slicers, the piRNA biogenesis was described as similar to the ping-pong model of *Drosophila* piRNAs (Aravin & Bourc'his 2008; Kuramochi-Miyagawa *et al.* 2010; Kuramochi-Miyagawa *et al.* 2008). In this model, the Piwi proteins create, guided by their bound RNAs, the 5` end of a secondary piRNA via their slicer activity. The 3`end formation remains unknown. The novel piRNA is subsequently incorporated into one of the Piwi orthologs and performs a similar reaction (Brennecke *et al.* 2007; Gunawardane *et al.* 2007). It was shown in the previous section that mHEN1 and MILI interact in both testis extract and in HEK 293 cells. Analysis of the slicer activity of MILI and determination if the slicing event of the 5` end of the secondary piRNAs and the methylation of the 3`ends by mHEN1 are functional connected are still missing in the literature.

In order to study the mechanism of methylation of piRNAs in the context of Piwi proteins, mHEN1 and MILI were required to be produced recombinantly. Bacterial expressions of full-length recombinant proteins was unsuccessful: neither mHEN1 nor MILI are soluble as full-length constructs. Therefore, co-expressions of mHEN1 and MILI as a recombinant complex in insect cells using a dual ORF expression construct were tried. Insect cell expression in *Sf*9 or *Sf*21 cells is also attractive because the reported lack of recombinant Sm and Piwi protein methylation (Brahms *et al.* 2000; Kirino *et al.* 2009), which enables to confirm the previous described results that the interaction of MILI and mHEN1 is independent of the post-translational methylation marks in the N-terminus of MILI.

2.4.1.2 Results

A fusion vector was designed encoding 6XHis-tagged MILI and CBP-tagged mHEN1 with independent promoters, which was inserted into the baculovirus genome by Tn7 transposition. The baculovirus was used to transfect *Sf*21 cells. The expression was monitored via YFP co-expressed by the bacmid (Berger *et al.* 2004; Bieniossek *et al.* 2008; Fitzgerald *et al.* 2006) (see Figure 16 A and B). Insect cells were harvested at the highest possible YFP expression. As indicated in Figure 16 B, the expression of both proteins was analyzed via Western blot of the total cell extract and the soluble fraction after lysis, using specific purified antibodies against MILI and mHEN1. Both proteins were detected in both fractions. Nickel-beads were used to purify the 6XHis-MILI and stained via affinity purified anti-mHEN1 antibody in a Western blot analysis, demonstrating that the CPB-tagged mHEN1 was co-purified (see Figure 16 C), suggesting that the two proteins are expressed and form a complex in the insect cells.

The 6XHis-MILI/CBP-mHEN1 complex was poorly expressed and purifications did not yield homogenous proteins, as judged by SDS-PAGE (see Figure 16 D first gel). Again, Western blot analysis of one of the purified elutions was performed, which showed the most abundant protein concentration (Elution 3). The Western blot indicated that MILI and mHEN1 are present in the nickel column elution (see Figure 16 D second and third column). Six bands in the proximity of the expected molecular weight of mHEN1 and MILI were excised and subjected to mass spectrometry. Indeed, MILI was identified in band #2 and mHEN1 in bands #4 and #5. The other bands are likely to be endogenous proteins from the insect *Spodoptera fugiperda* and unidentifiable as the genome has not been fully sequenced.

To validate the findings that mHEN1 interacts with MILI independent of the post-translational modifications of the arginines in the N-terminus of the Piwi protein, the Western blot membrane was re-stained with Y12 antibody. However, contrary to the published data (Brahms *et al.* 2000; Kirino *et al.* 2009) MILI was at least partially methylated (see Figure 16 C third row). Based on this result it is impossible neither to confirm nor to disprove the independence of the interaction of mHEN1 with MILI on the methylation state of the Piwi protein. Also, data from the laboratory show that low level expression of Bombyx mori Piwi proteins can lead to their methylation in *Sf*9 cells (personal communication from Zhaolin

Yang). Due to these reasons, although the original goal was to use the insect expression system to produce proteins lacking post-translational modifications, it is not possible to conclude that this is the case.

In the laboratory it was shown before (personal communication from Zhaolin Yang), that *Sf*21 cell contain piRNAs. Knowing that the MILI-mHEN1 complex is present, although in small amounts, whether bound *Sf*21 piRNAs are present in the complex should be determined. Therefore, RNA were isolated from the Ni-column elution fractions, ^{32}P end-labelled and seperated on a 15% denaturating polyacrylamid gel. No specific enrichment of a specific RNA species around 30 nt in length were identified (see Figure 16 E), indicating that the complex did not bind *Sf*21 piRNAs, likely due to the failure of MILI-mHEN1 complex in entering the piRNA pathway.

Given the difficulties in expressing the mHEN1/MILI complex and the impurities of the elutions we did not proceed further to functional analysis of the complex.

Figure 16: MILI/mHEN1 complex studies in *Sf*21

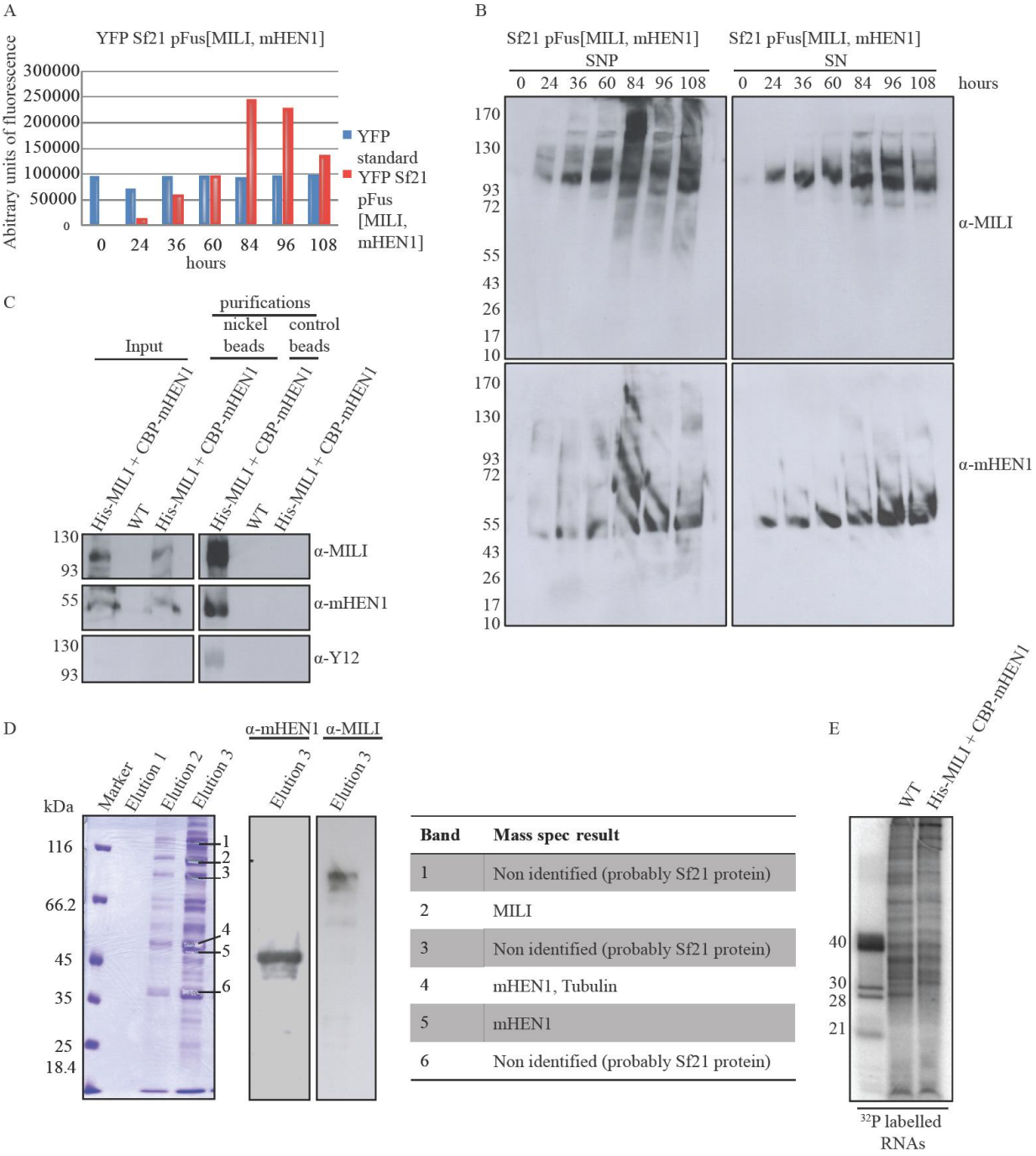
(A) 6XHis-MILI and CBP-mHEN1 were co-expressed in a fusion vector in *Sf*21. The expression was monitored by YFP-co-expression by the bacmid (red bars), analyzed against a laboratory internal standard at different time points (blue bars) after the cells growth arrested. As indicated, the expression was the highest after 84 hours after growth stop of the insect cells.

(B) The aliquots taken at each time point mentioned above were analyzed via Western blot using affinity purified anti-mHEN1 and anti-MILI antibodies, indicating that both proteins are expressed from 24 h after cell growth stagnation.

(C) The *Sf*21 cell extract of the transfected cells as well as non-transfected (WT) cell were purified via nickel-column. As a control Glutathione beads were used which are not supposed to bind the tagged-proteins. The purified samples were analyzed by Western blot using affinity purified anti-mHEN1 and anti-MILI antibodies. Nickel-column purifications of 6XHis-MILI co-purified CBP-mHEN1. The WT sample as well as the beads control remained negative. In difference to published data, MILI can be detected carrying symmetrically dimethylated arginines indicated by Y12 staining of the Western blot.

(D) Attempted nickel-column purification of the complex was analyzed via a coomassie stained gel, indicating that the complex is not significantly overexpressed. Although the 6XHis-MILI/CBP-mHEN1 complex was expressed, the coomassie stained gel revealed multiple other bands. Several bands of the approximate size of 6XHis-MILI and CBP-mHEN1 were excised from the coomassie stained gel and analyzed by mass spectrometry. In the adjusted table, the results are summarized, which indicate that CBP-mHEN1 and 6XHis-MILI are present but mixed with other *Sf*21 proteins like Tubulin.

(E) RNAs of nickel-column purified 6XHis- MILI/CBP-mHEN1 and WT samples exposed to the same experimental set up were extracted and analyzed on a 15% denaturating RNA gel. A significant enrichment of small RNAs in the size of 30 nt (which could indicate bound piRNAs) was not detected.



Band	Mass spec result
1	Non identified (probably Sf21 protein)
2	MILI
3	Non identified (probably Sf21 protein)
4	mHEN1, Tubulin
5	mHEN1
6	Non identified (probably Sf21 protein)

2.4.1.3 Discussion

In an attempt to express the recombinant MILI-mHEN1 complex in the insect cell line *Sf*21, analysis of the expression by Western blot analysis using affinity purified antibodies indicated that the complex can be detected by this sensitive method, but Coomassie stained gels show that the complex is not abundantly expressed in insect cells.

In contrast to the published data indicating that recombinant human and mouse proteins do not carry symmetrically dimethylated arginines in this expression system (Brahms *et al.* 2000; Kirino *et al.* 2009), it was shown that the 6XHis-MILI is at least partial methylated. This result makes it impossible to confirm or disprove the suggested independency of post-translational modification on MILI for the mHEN1 binding.

It was determined if the *Sf*21 expressed 6XHis-MILI/CBP-mHEN1 complex is bound to small RNAs in the size of 30 nt, which could indicate a loading of the complex with *Sf*21 piRNAs. A specific band of piRNA size was not observed. This is probably due to the fact that the recombinant mouse complex could not enter the piRNA pathway of the insect cells.

2.4.1.4 Method *Sf*21 expression of MILI and mHEN1

A fusion vector of pFastBac-6XHis-MILI and pUCS-CPB-mHEN1 was inserted into the baculovirus genome by Tn7 transposition and used to express 6XHis-tagged MILI and CBP-tagged mHEN1 in *Sf*21 cells. The expression was monitored by the co-expression of the bacmid internal YFP (Berger *et al.* 2004; Bieniossek *et al.* 2008; Fitzgerald *et al.* 2006). Aliquots of 1x 106 cells of the transfected Sf21 cells were harvested at each time point. After lysis, 1/10th of the total cell extract (SNP) were removed and the soluble part (SN) and pellet separated. 1/10th of the cell lysis was removed and the rest analyzed at a photo-spectrometer for YFP expression.

The cells were purified via the 6XHis-tag: small scale for immunoprecipitation: Lyses of 1 x 10^7 cells via hypotonic buffer (50 mM Tris pH=7.5, 15 mM KCl, 25 mM $MgCl_2$) using sonication, binding to a Nickel column, 8 times washing (50 mM Tris, 50 mM Imidazole,

150 mM NaCl, 0.02 % NP40). The beads were directly loaded onto a 10 % SDS gel used for western blots.

Large scale: The cell pellets were lysed (50 mM Na2HPO4, 300 mM NaCl, 10 mM Imidazole, 1 tablet of Complete EDTA free protease inhibitor cocktail; pH 7.5). The purification took place on a nickel column (2 CVs of a low salt solution: 50 mM Na2HPO4, 300 mM NaCl; then 10 CVs of a low Imidazole solution: 40 mM Imidazole; 10 CVs of a high salt solution: 50 mM Na2HPO4, 750 Mm NaCl, 750 mM of KCl, 40 mM Imidazole and, finally, 2 CVs of low salt solution again; eluted with 8 times 0.5 mL of elution buffer: 50 mM Na2HPO4, 300 mM NaCl, 250 mM Imidazole and dialyzed in 1x PBS, 0.03 % Tween20, 1 mM DTT). Antibodies used to detect 6XHis-tagged MILI and CBP-tagged mHEN1 via western blot analysis were anti-mHEN1 (GQG; 185-395 aa) and a anti-MILI monoclonal antibody (mAb, 13E-3; kindly provided by Dr Michael Reuter, recognizes the peptide corresponding to 107-122 aa of MILI and was raised at the Monoclonal Antibody Facility, EMBL).

Extracting the RNA and the 5'-end-labelling were performed as described previously (Aravin *et al.* 2006; Girard *et al.* 2006). Dried gels were exposed to a storage phosphor screen and scanned with a Typhoon scanner (Amersham).

To analyze if the 6XHis MILI/CBP-mHEN1 is able to bind RNAs, 2 x 10^7cells of wild type *Sf*21 or 6XHis MILI/CBP-mHEN1 expressing cells were used for purification via nickel beads. The beads were incubated in 100 mM KCl, 2 mM $MgCl_2$, 10 mM Tris/HCl pH 7.5, 40 U RNasin, 1 µg tRNA and 25 fmol 28 nt RNA. After 30 min at 30 °C, the complex was washed 5 times, the RNA extracted, labeled and analyzed on a 15 % Urea gel (Aravin *et al.* 2006; Girard *et al.* 2006).

2.4.2 RNF17 is does not influence the 3'end processing of piRNAs

2.4.2.1 Aim

The piRNA methyltransferase mHEN1 interacts with the murine Piwi protein MILI in testes extract as well as co-expressed in HEK293 cells. Here, it was attempted to identify other binding partners of mHEN1, which might be involved in the 3`end processing and modification of piRNAs. Immunoprecipitations using the affinity purified mHEN1 antibody were performed and analyzed via mass spectrometry, searching for co-purified protein (personal communication from Michael Reuter). Unfortunately, interesting candidates which could be involved in the 3` end formation and modification of piRNAs were not detected.

Based on the findings that mHEN1 localizes in a perinuclear granule distinct from the chromatoid body, proteins with similar localization and expression pattern were searched. One candidate was the Tudor domain and RING finger domain-containing RNF17/TDRD4 (Pan *et al.* 2005). The two isoforms of the protein are exclusively expressed in testis. The short isoform (RNF17S) contains three and the long isoform (RNF17L) five Tudor domains. RNF17S localizes diffusely in the cytoplasm of all germ cells, while RNF17L builds a structure distinct from the chromatoid body in late pachytene spermatocytes, round spermatids and elongating spermatids (Pan *et al.* 2005), similar to the HEN1-body (see Figure 17) . RNF17 is essential for the spermatogenesis. Like *miwi* mutants, *Rnf17* mutant mice are male sterile, due to a block of the spermatogenesis after round spermatids (Pan *et al.* 2005), indicating that the protein could be connected to the piRNA pathway due to its knockout phenotype. It was investigated if RNF17 and mHEN1, due to their similar localization pattern, are functionally connected.

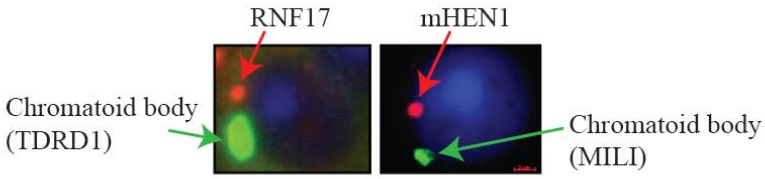


Figure 17: RNF17 and mHEN1 show a similar localization pattern.
Pan *et al.* localized RNF17 (red) and TDRD1 (green) a protein known to be localized in the chromatoid body in round spermatids (left picture). In the right image the mHEN1 granule (red) distinct from the chromatoid body (green) is indicated by co-staining of MILI in round spermatids. Despite the fact that in the testes section image of Pan et al the resolution is not that good as in purified germ cells, the localization pattern of mHEN1 and RNF17 (red arrows) in comparison to the chromatoid body (green arrows) is similar in round spermatids.

2.4.2.2 Results

RNF17 and mHEN1 show similar localization pattern. In round spermatids both proteins are found to form granules distinct from the chromatoid body (Pan *et al.* 2005). The antibodies to both proteins were produced in rabbits, making it rather difficult to perform co-localization experiments using fluorescent secondary anti-rabbit antibodies. To avoid inconclusive localization experiments, first a possible interaction between both proteins was investigated. The anti-mHEN1 antibody was used for immunoprecipitations from adult mouse testes extracts and the proteins were analyzed via western blot. Using an anti-RNF17 antibody a co-purification of RNF17 could not be detected in the anti-mHEN1 immunoprecipitate (see Figure 18 A).

Then the influence of RNF17 in the piRNA pathway, especially in the 3` end formation, was investigated. *rnf17* mutant mouse testis were obtained from Dr. Jeremy Wang, University of Pennsylvania. Immunoprecipitations of MILI and MIWI were performed from adult mouse testes extract of WT, $RNF^{-/+}$ and $RNF^{-/-}$ and the RNAs were isolated. After radioactive labeling of the RNAs, they were analyzed on a denaturating gel. In all samples piRNA size and abundance were not affected (see Figure 18 B). To analyze if RNF17 influences the 3` end modification of piRNAs, the piRNAs of MIWI and MILI immunoprecipitations of WT, $RNF^{-/+}$ and $RNF^{-/-}$ samples were subjected to β-eliminations. The piRNA mobility of all testes extract samples did not change after the treatment, indicating that the RNAs were modified at their 3' end, while artificial unmodified RNAs run faster due to a loss of the last nucleotide (see Figure 18 C).

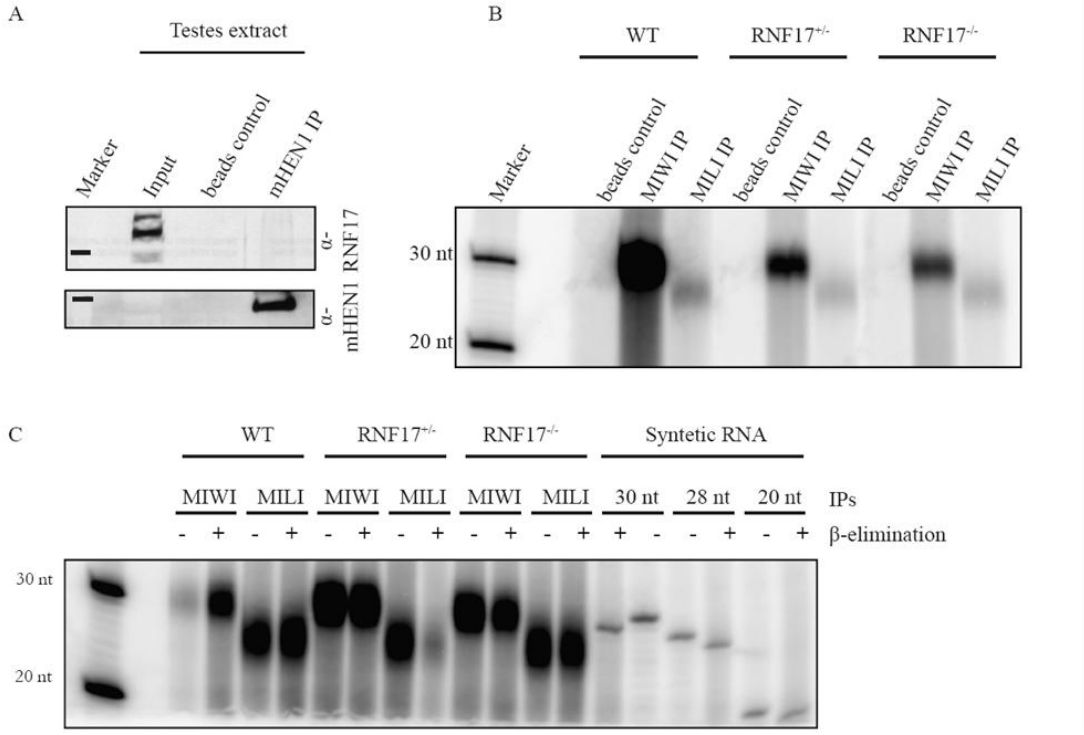


Figure 18: RNF17 does not affect the 2'O methylation of the 3'end of piRNAs

(A) mHEN1 was immunopurified using an affinity purified antibody and analyzed on a Western blot. Staining with anti-RNF17 identified the protein in the input sample but not as a co-purified protein in the mHEN1 IP.

(B) MIWI and MILI bound RNAs were immunoprecipitated from of WT, RNF17$^{+/-}$ and RNF17$^{-/-}$ adult testes extracts. The RNAs were ^{32}P labeled and analyzed on a denaturating gel. No significant difference in the samples could be detected.

(C) The RNA of MIWI and MILI IPs were oxidized and a β-elimination was performed. In contrast to the synthetic control RNA oligos of 30nt, 28nt and 20nt the β-elimination did not affect the mobility of the immunopurified RNA on the gel indicating that they are 3` end modified.

2.4.2.3 Discussion

Many different types of perinuclear electron-dense structures, called nuages, are identified in the germ cells in different species and have been described as important for germ cell specification and differentiation. One of these nuages in mice is the chromatoid body, thought to be an active center for the piRNA pathway (Kotaja & Sassone-Corsi 2007). Another example in mice is the RNF17 granule. RNF17 is a testes-specific protein with a RING domain and three or five Tudor domains, according to the isoform. The long isoform RNF17L forms dimers or polymers, indicating that it may be responsible for assembly of the RNF17 granules (Pan *et al.* 2005). These granules are distinct from the chromatoid body, like the HEN-body. Although RNF17 shows a similar localization pattern to mHEN1, it does not interact with the methyltransferase nor does it seem to be involved into the 3` end processing of piRNAs. These results indicate that RNF17 is probably not needed at the same step of the piRNA biogenesis as mHEN1. It is possible that the protein is involved further downstream in piRNA pathway, potentially in DNA methylation, because it contains Tudor domains, which could bind to histones or Piwi proteins carrying methylated arginines. Until now there is no reason to believe that RNF17 functions in the 3`end processing of piRNAs.

2.4.2.4 Methods

Testes extracts were prepared from adult CD1 WT mice, +/- RNF17 (heterozygous) and -/- RNF17 (knockout) testes kindly provided from the Wang laboratory (Pan *et al.* 2005). Endogenous mHEN1 or Piwi proteins from mouse testes extract were immunopurified using the following antibodies linked to Protein G sepharose: anti-mHEN1 (GQG and GOY, 185-395 aa), anti-MIWI (BVI, 1-200 aa) and anti-MILI monoclonal antibody (mAb, 13E-3). For Western blot analysis purified rabbit anti-RNF17 (Wang laboratory) was used in a dilution of 1:500 and affinity purified anti-mHEN1 in a dilution of 1:250. Mouse testes extract preparation and immunoprecipitation and 5'-end-labelling of piRNAs were performed as described previously (Aravin *et al.* 2006; Girard *et al.* 2006). Dried 15% Urea gels were exposed to a storage phosphor screen and scanned with a Typhoon scanner (Amersham). The β-eliminations were performed as described (Vagin *et al.* 2006).

2.4.3 Structural insight of binding of the methylated piRNAs by the MIWI PAZ domain

2.4.3.1 Contribution remarks

The biophysical characterization of the PAZ domain is based on collaboration between the Carlomagno group from the Structural and Computational Biology unit at the European Molecular Biology Laboratory Heidelberg, Germany, and the Pillai group of the European Molecular Biology Laboratory, Grenoble Outstation Grenoble, France. While the NMR structure of the MIWI PAZ domain was solved in the Carlomagno group, I focused on the expression, optimization of the expression and the purification of the PAZ domain for biochemical and biophysical experiments. In our lab, a trainee student, Elsa Rocha, designed and established the first purification protocols for the soluble MIWI PAZ construct. Two point mutant constructs were prepared by Jerome Bourdin, a research associate in the lab. The biophysical measurements using SPR were performed by Dr. Peter Sehr at the Chemical Biology Facility, EMBL Heidelberg. My contribution consisted in following up the project in the lab by optimization and purification of the PAZ domain and the mutants, as well as the design of new mutant constructs for Biacore experiments in Heidelberg. I also performed several UV-crosslinking experiments that were not included in the final paper.

2.4.3.2 Remarques concernant ma contribution

La caractérisation du domaine PAZ est basée sur une collaboration avec le groupe Carlomagno de l'unité de biologie structurale et bioinformatique d'EMBL Heidelberg et le groupe Pillai de l'antenne EMBL de Grenoble, France. Alors que la structure NMR du domaine PAZ de MIWI a été résolue dans le groupe Carlomagno, je me suis concentrée sur l'expression et la purification du domaine PAZ pour des expériences biochimiques et biophysiques. Dans notre laboratoire, une stagiaire Elsa Rocha a conçu et rédigé les protocoles de purification préliminaires pour la construction Miwi PAZ. Deux mutants ponctuels ont été préparés par Jerome Bourdin, notre assistant de recherche. Les mesures

biophysiques utilisant la SPR (Surface Plasmon Resonance) ont été réalisées par le Dr Peter Sehr (Chemical Biology Facility, EMBL Heidelberg). Mon rôle a été de purifier le domaine PAZ et ses mutants ainsi que de créer de nouvelles constructions pour des analyses complémentaires Biacore à Heidelberg. J'ai également réalisé plusieurs expériences de pontage covalent par UV non inclus dans le papier final.

2.4.3.3 Aim

All piRNAs in total mouse testes RNA are demonstrated to carry the 2`-O-methyl modification (Kirino & Mourelatos 2007c; Ohara *et al.* 2007b). The role of this modification in RNA binding to the PAZ domain has not yet been established. To test the possibility of a methylation-requirement of the piRNAs in the binding to the MIWI PAZ domain, the NMR structure of the MIWI PAZ domain in complex with a 2`O methylated ssRNA (8 nt) was solved in collaboration with the Carlomagno group at the EMBL Heidelberg based on a construct by Elsa Roch. Like the PAZ domain of Dicer and AGO proteins, the MIWI PAZ domain shows an OB-fold. The short β-barrel is capped by 2 α-helices on one side and an α-helix and β-hairpin on the other. The RNA binds in the cavity, built by the central β-barrel and the α-helix/β-hairpin motif. The 2`-O-methyl group of the 3` end is recognized by a hydrophobic cleft consisting of F333 on the β-hairpin and M382 and A381 on strand β7 of the core β-sheet. The carbonyl and amide of M382 of strand β7 forms hydrogen bonding interactions with the 3`-hydroxyl and the 2`methoxy groups of the RNA 3` end. Based on chemical shifts, the protons of the 3`nt ribose are in close proximity to residues of the β-barrel and the α-helix/β-hairpin insert (A381, L383, F333). The second to last nucleotide of the ssRNA does not show interactions with the protein. Several contacts were observed between the ribose of the third to the sixth nucleotides and strand β2 of the central barrel (T317, Y318, R319) and the C-terminus (L393). That leads to a stretched conformation of the phosphate backbone between the 3`end nt and the second to last nt. In the MIWI PAZ domain, it seems that the 2`-O-methyl group of the RNA occupies the space which is taken by the 3` end nucleotide of unmethylated RNAs bound to the hAGO1 PAZ domain. This indicates that the RNA in the MIWI PAZ/RNA complex is shifted out of the binding pocket

by one nucleotide, while the last nucleotide takes the position close to the second to last nucleotide in the hAGO1 complex with unmethylated 3` end RNA (Simon *et al.* 2011).

Based on the NMR structure of our collaborators and alignments of different PAZ domains, the biophysical nature of the interaction of the different amino acids of the PAZ domain to the piRNAs was characterized. The goal was to understand the impact of the methyl group on the 3` end of piRNAs on the recognition and binding by the MIWI PAZ domain. To this end Surface Plasmon Resonance (SPR) experiments were performed in collaboration with Dr Peter Sehr at the EMBL Core facility in Heidelberg, as well as UV cross linking experiments of purified proteins and mutants with synthetic RNAs.

2.4.3.4 Results

2.8.3.4.1 PAZ mutant preparation and protein expression

From the NMR structure, three PAZ domain residues T332, F343 and M382 were determined to be important for 3`end piRNA binding in the RNA binding pocket of the PAZ domain. These residues were replaced by alanines (cloned by Jerome Boudin) to analyze their impact in the interaction. Further, two residues potentially involved in RNA backbone binding (K319A and R316A) were also mutated. The constructs were cloned in the 6XHis-GST-tag encoding plasmid pET-M-30. WT constructs of the MIWI PAZ domain (constructed by Elsa Rocha) and the PAZ domain of hAGO2 (constructed by Maartje Luteijn) were used.

The proteins were expressed in Rosetta^{TM} 2 *E. coli* cells and purified via a Ni-column. After optimization of the purification protocol, all constructs were soluble and homogeneous by SDS-PAGE (see Figure 19). As a negative control for further experiments, GST alone was expressed. The negative control protein GST migrates as higher molecular weight construct than the GST-tagged PAZ domains and mutants, because it contains a 6XHis-tag, a TEV-cleavage site and a *DCoH* (an unrelated protein).

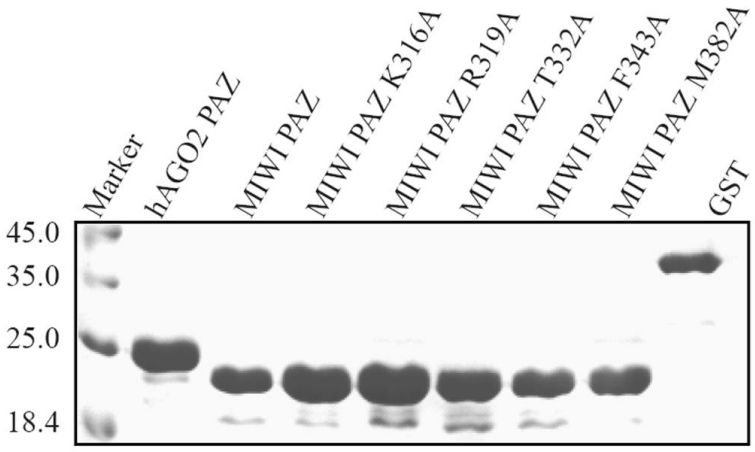


Figure 19: Purified PAZ domains and mutants
6XHis-GST, 6XHis-GST-hAGO2 PAZ, 6XHis=GST-MIWI PAZ and the 6XHis-GST-MIWI PAZ mutants were purified and run on a 15 % SDS gel, which was coomassie stained afterwards.

2.4.3.4.2 Biophysical investigation of the influence of the 3`O methylation of RNAs in the PAZ binding

The purified 6XHis-GST-hAGO2 PAZ, the 6XHis-GST-MIWI PAZ and the 6XHis-GST-MIWI PAZ mutant domains were used for Surface Plasmon Resonance (SPR) experiments against methylated and unmethylated single stranded RNAs to determine the affinity of each construct. The K_D of the binding to unmethylated RNAs is 2 µM and for methylated RNAs 0.9 µM so the ratio is $K_{D,meth}/K_{D,unmeth}$=0.5. Despite the fact that the NMR structure suggests that the M382 is a key residue for the recognition of the 2`O methyl-group on the RNA, the mutation M382A just increases the K_D of the MIWI PAZ M382A domain binding for the methylated RNA complex by a factor of two (K_D=2 µM). The K_D for the MIWI PAZ to unmethylated RNA complex increased by 5 fold (K_D=9.7 µM) indicating that the M382A mutation probably has a destabilizing effect on the hydrophobic core of the PAZ domain. Mutations of other residues predicted to influence the binding of the methylated RNA, like MIWI PAZ T332A, MIWI PAZ F343A and MIWI PAZ M382A did not change the binding.

Although it is not yet known how the biogenesis of piRNAs works, it is assumed that piRNAs arise from single stranded precursors, independent of Dicer activity. In this context it is expected that Piwi proteins bind preferentially ssRNAs and the sliding of the methylated RNA by one nucleotide in the MIWI PAZ domain should be accepted, because the ssRNA has no structural limitations. In Dicer dependent pathways, like the plant si- and miRNAs and the siRNAs in *Drosophila*, the RNAs carry a 3` 2nt overhang, where the last nucleotide can be methylated (Horwich *et al.* 2007; Yu *et al.* 2005). Due to the similarities of the structure and recognition of RNAs of the PAZ domains in the different pathways it was proposed that the PAZ domain has a similar affinity to methylated RNA with a 1 nt overhang analogous to unmethylated RNAs with a 2 nt overhang. To follow this idea, an SPR experiment was performed in which the MIWI PAZ domain was exposed to an RNA with a 10 nt overhang, simulating ssRNA ($K_{Dunmethylated}$=0.9 µM, $K_{Dmethylated}$=2 µM), to an RNA with a 1 nt overhang ($K_{Dunmethylated}$=134 µM, $K_{Dmethylated}$=105 µM) and an RNA with 2 nt overhang ($K_{Dunmethylated}$=55 µM, $K_{Dmethylated}$=27 µM). The interaction of the MIWI PAZ domain is 30 times more stabile with ssRNAs than with 2 nt overhang dsRNAs. This indicates that the domain was optimized during evolution to bind ssRNAs. Comparing RNAs with 1 nt overhang to 2 nt overhang RNAs shows that the affinity of the MIWI-PAZ domain to the RNA with 1 nt overhang is 4 times lower, irrespectively of the methylation. In conclusion,

the binding of the methyl-group on the 3'end of the RNA does not allow enough flexibility to compensate for another unpaired nucleotide.

2.4.3.4.3 RNA binding to the PAZ domains

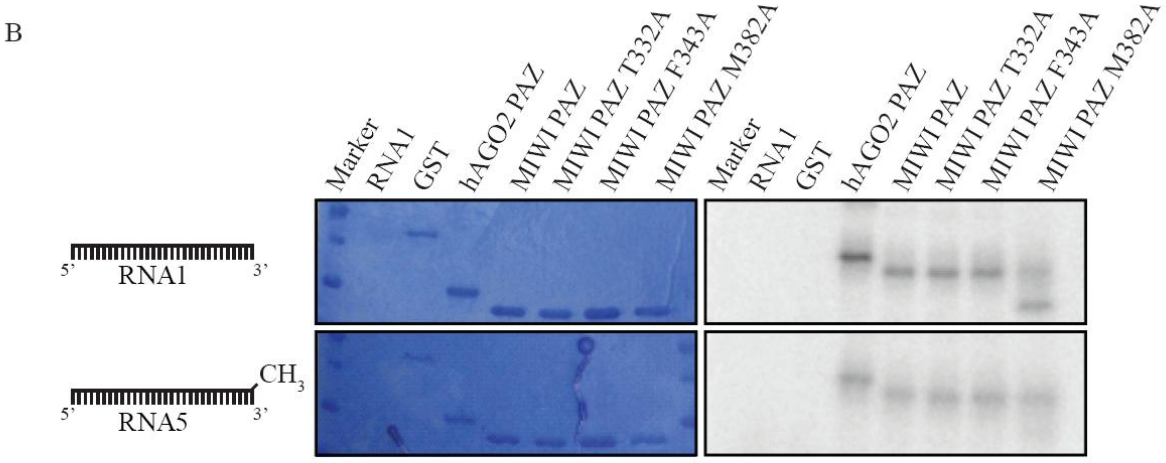
In parallel to the SPR experiments, the influence of different amino acids of the MIWI PAZ domain on the RNA-binding stability via UV cross-linking experiments was investigated. UV crosslinking is a common method to identify direct binding of RNAs to RNA-binding proteins. Upon exposure to ultraviolet light (UV cross-linking), proteins cross-link nucleic acids at their contact points.

In order to use this approach the purified 6XHis-GST-hAGO2 PAZ, the 6XHis-GST-MIWI PAZ and the 6XHis-GST-MIWI PAZ mutant domains were incubated with ^{32}P labeled methylated (RNA 5) and unmethylated (RNA 1) single stranded synthetic RNAs (see Figure 20 A, table) and UV-cross-linked. The samples were then separated by a 15% SDS-PAGE (see Figure 20 B right column) and the intensity of the radioactive bands on the dried gel (see Figure 20 B left column) were analyzed via the Typhoon™ 9410 Variable Mode Image (see Figure 20 C and D).

The 6XHis-GST-hAGO2 PAZ domain binding to non-methylated RNAs is slightly higher than to methylated targets. In contrast, the binding of the 6XHis-GST-MIWI PAZ domain and the 6XHis-GST-MIWI PAZ mutants has higher affinity for the methylated RNAs than for unmethylated RNAs. A significant difference in MIWI PAZ and MIWI PAZ mutant binding was not observed, possibly due to the sensitivity of the assay.

A

RNA		Sequence
RNA1	5' 3'	UGACAUGAACACAGGUGCUCAGAUAGCUUU
RNA5	CH_3 5' 3'	UGACAUGAACACAGGUGCUCAGAUAGCUU**U** (whereby **U** was 2'-O-methyl modified)



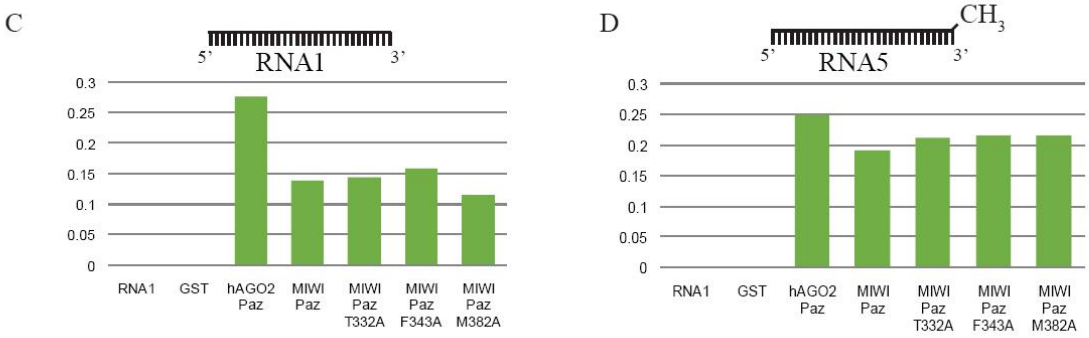


Figure 20: UV-cross-linking experiments of PAZ domain and mutants to RNAs.
(A) The RNAs used during the UV-cross linking experiment are indicated in the table.
(B) GST, 6XHis-hAGO2 PAZ, 6XHis-MIWI PAZ and the 6XHis-MIWI PAZ mutants were UV-cross-linked with unmethylated (RNA 1) and 2`O methylated RNAs. The gels were coomassie stained (left part) and exposed to a phosphor screen (right part).
The intensities of the non-methylated **(C)** and methylated **(D)** RNA cross-linked to the different proteins were measured against the background and blotted in a column diagram.

2.4.3.5 Discussion

Our collaborators of the Carlomagno group at the EMBL Heidelberg solved the NMR structure of the MIWI PAZ domain bound to a methylated small RNA. The structure is highly similar to the known PAZ domains (Lingel *et al.* 2003, 2004; Ma *et al.* 2004; Song *et al.* 2003; Yan *et al.* 2003). The interaction of the MIWI PAZ domain with RNAs were biophysically analyzed, indicating that the MIWI PAZ domain prefers single stranded methylated small RNAs.

piRNAs carry a 2`-O-methylation mark at their 3` terminal nucleotide (Horwich *et al.* 2007; Kirino & Mourelatos 2007c; Ohara *et al.* 2007a; Saito *et al.* 2007a; Zhou *et al.* 2010). It is not clear yet, what role this modification has. In fly and zebrafish mutants of the RNA methyltransferase *hen1*, Piwi proteins still correctly incorporate the unmethylated piRNAs, suggesting that the modification of the RNA is not needed for the loading process (Horwich *et al.* 2007; Kamminga *et al.* 2010). On the other hand, piRNAs in zebrafish, siRNAs in flies and si- and miRNAs in plants lacking the modification are exposed to an unknown enzyme adding a short Poly U-tail which consequently leads to trimming by an unidentified nuclease (Ameres *et al.* 2010; Kamminga *et al.* 2010; Li *et al.* 2005). This indicates that the methyl group at the 3`end is needed for stabilization of small RNA.

In mouse, the function of the post-transcriptional modification of the piRNAs is unknown, but even if it is assumed that the methylation protects the piRNAs from destabilizing events, the methylated 3` end of the piRNAs has to be recognized and to fit into the PAZ domain of the Piwi proteins. Here, the influence of the 2`-O-methylation of the 3'end of piRNAs on the binding to the MIWI PAZ domain was investigated.

In comparison to the hAGO1 PAZ structure, the binding of the methylated 3`end of the RNA to the MIWI PAZ domain shifts the RNA by one nucleotide out of the binding pocket, so that the first nucleotide takes the space, which is normally taken by the second nucleotide bound in an hAGO1 PAZ. Taken this shift by one nucleotide in the MIWI PAZ domain into account it could be that the methylation is needed for a tight fit into the protein. The PAZ domain itself can recognize the RNA and bind it, but without the methylation mark on the RNA the binding is loose and there is a risk of losing the piRNA due to enzymatic activities. To avoid this, the methylation mark is recognized and fits the RNA in the correct

tight position in the domain. The PAZ domain of the human Piwi protein HIWI shows as well that the RNA binding pocket is able to accommodate the methyl group of the RNA while AGO1 is unable to bind the modified RNAs due to steric clashes with the surfaces of amino acid side chains (Tian *et al.* 2010).

The key-residue of the specific interaction of the MIWI PAZ domain with methylated RNAs is the long side-chain of M382, which bends around from the β7 on the opposite side of the RNA binding surface to contact the RNA. Surprisingly, this methionine is not conserved in all Piwi proteins, but in MIWI, HIWI and Aubergine. For the HIWI PAZ domain it was shown that this methionine contacts the methyl group at the 3` end of RNAs as well (Tian *et al.* 2010). Despite these contacts, the methionine seems not to be required for the binding to methylated RNAs per se, but has perhaps evolved to optimize the interaction.

However, taken the cross-linking and Surface Plasmon Resonance (SPR) results into account, it was shown here that the recognition of methylated and unmethylated RNAs by the MIWI PAZ domain is highly similar. The K_D of the binding to unmethylated RNAs is 2 µM and for methylated RNAs 0.9 µM. Recently, it was published that the PAZ domain human Piwi proteins HIWI, HIWI2 and HILI bind to a methylated 12 nt self-complementary duplex with 2 nt overhangs 2.5 to 6-fold higher affinity than to unmethylated RNAs measured by isothermal titration calorimetry (Tian *et al.* 2010), confirming the observed data. The only significant change in the affinity to methylated RNAs was determined in M382 mutants, arguing again for an important role of this residue in the methyl group recognition of the MIWI PAZ domain. In contrast, double mutants of the methionine (M381Y) and a Proline (P379H) in HIWI PAZ revealed no significant difference in the binding to methylated or unmethylated RNA substrate (Tian *et al.* 2010).

In conclusion, the MIWI PAZ domain binds preferentially to single stranded RNAs over double stranded RNAs with 3` overhangs. The methyl-group on the 3` end of ssRNAs is not essential for the binding, but buries the RNA deep into the binding pocket. All data suggest that the PAZ domain has a high flexibility to adapt to RNAs with different sequences, chemical modifications and structures.

2.4.3.6 Methods

2.4.3.6.1 Protein expression and purification

Generation of the expression constructs and preliminary protein expression of the PAZ domain of MIWI (276-425 aa) and hAGO2 (228-371 aa) were carried out by Elsa Rocha and Maartje Luteijn, respectively. The mutations of the MIWI PAZ expression construct T332A, F343A and M382A (performed by Jerome Bourdin) and additional mutation constructs (K316A and R319A) were designed by directed point mutagenesis.

Samples for structural work were prepared in Dr. Teresa Carlomagno's lab at EMBL Heidelberg essentially based on the protocol set up by Elsa Rocha. Protein samples used for UV-crosslinking and some of the SPR measurements were prepared by myself as follows: the proteins were expressed in Rosetta^{TM} 2 in 4 L TB medium and induced by 1 mM IPTG. The cell pellets were lysed (50 mM Na2HPO4, 300 mM NaCl, 10 mM Imidazole, 4 U/mL DNase I, 0,3 µg/mL RNase A, 1 tablet of Complete EDTA free protease inhibitor cocktail; pH 7.5) by using a microfluidizer with a chamber pressure of 15-18 kPSI. The purification took place on a nickel column (2 CVs of a low salt solution: 50 mM Na2HPO4, 300 mM NaCl; then 10 CVs of a low Imidazole solution: 40 mM Imidazole; 10 CVs of a high salt solution: 50 mM Na2HPO4, 750 Mm NaCl, 750 mM of KCl, 40 mM Imidazole and, finally, 2 CVs of low salt solution again; eluted with 8 times 0.5 mL of elution buffer: 50 mM Na2HPO4, 300 mM NaCl, 250 mM Imidazole and dialyzed in 1x PBS, 0.03 % Tween20, 1 mM DTT).

2.4.3.6.2 Surface Plasmon Resonance (SPR) experiments

The experiments were performed at 25 °C on a Biacore T100 machine (Biacore AB, Uppsala, Sweden) by Dr. Peter Sehr at the Chemical Biology Facility, EMBL Heidelberg.

The RNAs were design to consist of an anchor strand with a coupled biotin at the 3` end (5` - GAGCACCUGUGUUCAUGUCA - 3` biotin, 100 µM) and a complementary strand, which was annealed to it, leading to 1, 2 or 10 nt overhangs. Used RNAs were RP RNA 1 (5`- UGACAUGAACACAGGUGCUCAGAUAGCUUU-3`, 100 µM), RP RNA 5 (5`-UGACAUGAACACAGGUGCUCAGAUAGCUU**U**-3` where **U** was 2`-O-methyl modified, 2nt overhang, 5` UGA CAU GAA CAC AGG UGC GC **UU** 3`; 1 nt overhang, 5`-UGA CAU GAA CAC AGG UGC GC **U**-3`. 25 µM biotinylated RNA and 25 µM of the

experimental required complementary RNA were incubated 2 min at 90 °C and cooled down to RT within 45 min in 20 mM Tris pH 7.5, 100 mM NaCl, 2 mM EDTA to form duplexes. The annealed duplex mixture was diluted to 25 nM in PBS pH 7.4, 0.05% Tween20, 1 mM DTT and bound to a streptavidin-coated biosensor chip. As a reference, the sensor surface without modification was used. The Biotin Capture Template was run without ligand with 50 mM NaOH + 1 M NaCl as conditioner. A manual run with 30 µl/min was performed to stabilize the baseline. The Biotin Capture Template was then run without conditioning, with the aim for a target level of 120 RU. The wild type and mutant PAZ proteins were twofold ranging from 100-0.098 µM. The proteins were injected as analytes at a flow rate of 30 µL/min and the equilibrium dissociation constants determined by steady state affinity evaluation.

2.4.3.6.3 Binding assay via UV-cross-linking

These studies were carried out at the EMBL Grenoble: 5 µM protein (6XHis-GST, 6XHis hAGO2 PAZ, 6XHis MIWI PAZ or 6XHis MIWI PAZ mutants) and 100 µM 30 nt P^{32}-labelled ssRNA (5`-UGACAUGAACACAGGUGCUCAGAUAGCUUU-3`) with (RP RNA5) or without (RP RNA1) 2`-O-methylation on the 3' end were mixed in 1 x PBS (10 µL final volume) and incubated at 30 ˚C for 10 min. The samples were exposed to ultraviolet light (300 mJ). The reactions were stopped by adding 10 µL 2x SDS loading dye and heated to 94 ˚C for 5 min (Adapted from (Lingel *et al.* 2003)). Then the samples were resolved on 15 % SDS gels, coomassie stained, dried and exposed to a storage phosphor screen. The screen was read out at the Typhoon™ 9410 Variable Mode Image, or Typhoon Trio imager GE healthcare, reference: 63-0055-87.

Chapter 3: Analysis of the MOV10L and TDRD1 homologs in zebrafish

3.1 Contribution remark

The results in this chapter are entirely my work. Help for the initial experimental set up was provided by the Gilmour group at the European Molecular Biology Laboratory, Heidelberg, Germany.

3.2 Remarques concernant ma contribution

Les résultats de ce chapitre proviennent uniquement de mon travail. L'équipe de Darren Gilmour, EMBL Heidelberg, m'a aidée pour la mise au point des expériences de base.

3.3 Analysis of the MOV10L and TDRD1 homologs in zebrafish

Stephanie Eckhardt[1], Andreea Gruia[2], Darren Gilmour[2] and Ramesh S. Pillai[1]

[1]European Molecular Biology Laboratory, 6 Rue Jules Horowitz, BP 181, 38042 Grenoble, France.

[2]European Molecular Biology Laboratory, Meyerhofstraße 1, 69117 Heidelberg, Germany.

Correspondence to:

Ramesh Pillai

Phone: xx-33-4-76 20 7446

Fax: xx-33-4-76 20 7199

E-mail: pillai@embl.fr

3.3.1 Abstract

Piwi-interacting RNAs (piRNAs) are responsible for silencing of transposons in the germline, but their biogenesis and the factors involved are, as yet, not fully understood. The inaccessibility of the germ cells is a major hindrance to study piRNAs. However, Tudor domain-containing protein 1 (TDRD1) and the putative RNA helicase MOV10L have been implicated in the mouse piRNA pathway, while their exact mechanism of action is not known. Homologs of these proteins are found in many other organisms ranging from flies to humans, indicating their conserved roles in the piRNA pathway. To analyze germline pathway components zebrafish (*Danio rerio*) embryos were utilized as a model system. *Dr*TDRD1 and *Dr*MOV10L were identified based on sequence analysis and their co-expression with fish Piwi proteins was determined. In localization experiments, the testes-specific *Dr*TDRD1 was found to be in primordial germ cells. It was shown by reporter studies that the tissue-specific localization of *Dr*TDRD1 is dependent on its 3'UTR. Knockdowns of *Dr*MOV10L desilences transposons, implicating a role in the fish piRNA pathway.

3.3.2 Resume

Les ARNs interagissant avec Piwi (piARNs) sont responsables de la répression des transposons dans la lignée germinale, mais leur biogénèse et les facteurs impliqués sont, à ce jour, mal connus. Limpossibilité d'étudier directement les cellules germinales constitue un obstacle majeur à l'étude des piARNs. En revanche, la protéine à domaine Tudor TDRD1 et la potentielle ARN hélicase MOV10L ont toutes deux été impliquées dans la voie des piARNs chez la souris, alors que leur mode d'action n'est pas connu. Les homologues de ces protéines sont retrouvés dans de nombreux autres organismes, des mouches aux humains, ce qui montre leur rôle conservé dans la voie des piARNs. Afin d'analyser les différents composants de la voie de signalisation dans les cellules germinales, des embryons de poisson zèbre (*Danio rerio*) ont été utilisé comme modèle. *Dr*TDRD1 et *D*rMOV10L ont été identifié à partir de l'analyse de leur séquence primaire et leur coexpression avec les protéines Piwi de

poisson zèbre a été établie. Dans des expériences de localisation, la protéine *Dr*TDRD1, spécifique des testicules, a été retrouvée dans les cellules germinales primordiales. Il a encore été montré, par gène rapporteur, que cette localisation tissu-spécifique de *Dr*TDRD1 est dépendante de son extrémité 3'UTR. Des KO de DrMOV10L dérépriment les transposons, ce qui suggère un rôle dans la voie des piARNs chez le poisson.

3.3.3 Introduction

The gonad-specific Piwi proteins of the Argonaute family associate with 26-31 nt RNAs, called piRNAs, to protect the genome of the germline from transposable elements in all animals studied so far (Ghildiyal & Zamore 2009) (see section 1.3.5-1.3.7).

Mice and *Drosophila* are the favored model systems for investigating the piRNA pathway components. However, both of these systems have their limitations with a suitable alternative model system being *Danio rerio* (zebrafish). The advantages of the zebrafish are rapid generation time (12-16 weeks), prolific egg-layers, visibility of the cells during larval stages and simplicity and high throughput nature of manipulations. The accessibility of the germline facilitates rapid work on the piRNA pathway. Only one other group uses the zebrafish system to investigate questions concerning the piRNA biogenesis, factors associated with this pathway and the functions of these components.

To date it has been shown that in zebrafish, the Piwi clade consists of two members: ZILI and ZIWI (Houwing *et al.* 2008; Houwing *et al.* 2007). ZIWI is initially maternally provided and later expressed strongly in mitotic and early meiotic stages of the germ cell differentiation (Houwing *et al.* 2007) (see Figure 21). ZILI was first described as expressed from 3 days post fertilization (dpf) onwards and not detectable as a maternally contributed product (Houwing *et al.* 2008). However, Sun and colleagues (2010) reported that ZILI can be detected at the one-cell stage of the embryo (see Figure 21). At 3 dpf, the ZILI protein is localized in the nucleus as well as in the cytoplasm. In adult zebrafish, ZILI is present in mitotic and early meiotic cells in the cytoplasm in male and female gonads like ZIWI (Houwing *et al.* 2008). *ziwi* or *zili* knockout fish show a loss of germ cells. While this loss is due to excessive apoptosis in *ziwi* mutants (Houwing *et al.* 2007), *zili* mutants show an arrest

in the germ cell development (Houwing *et al.* 2008). In both cases the fish are phenotypical male (Houwing *et al.* 2008).

ZIWI and ZILI-associated piRNAs (26-27 nt) are found in both the male and female germline, with greater abundance in ovaries compared to testis. The zebrafish piRNAs originate mainly from transposons (mainly from LTR elements) and repetitive sequences, while most of the non-repetitive sequences are in intergenic regions (Houwing *et al.* 2008; Houwing *et al.* 2007). As with flies, zebrafish piRNAs can be passed maternally to the embryo (Houwing *et al.* 2007). piRNAs binding to ZIWI have a 5` U bias and are preferentially antisense regarding the transcript. ZILI- associated piRNAs show a bias for an A at the 10 nt position and have sense orientation (Houwing *et al.* 2008). These are typical signatures required for the ping-pong amplification cycle. This model was analyzed in *Drosophila* and describes how piRNAs guide a Piwi protein-containing complex to a target RNA, which is cleaved, defining a 5`end of a new piRNA, while silencing transcripts of transposable elements (Brennecke *et al.* 2007; Gunawardane *et al.* 2007) (see section 1.3.6).

To gain deeper insights into the piRNA pathway and to identify associated factors, zebrafish was used as a model system. New tools had to be developed and methods used in other organisms adapted to the requirements of the zebrafish system. Screening methods involving fluorescence reporter or analysis of transposon activity via qPCRs could shed light on piRNA pathway components. To test the reliability of the methods and to gain first insights into piRNA-associated components, the homologs of two Piwi-interacting factors analyzed in mouse, MOV10L and TDRD1 (Reuter *et al.* 2009; Zheng *et al.* 2010), were chosen to be investigated.

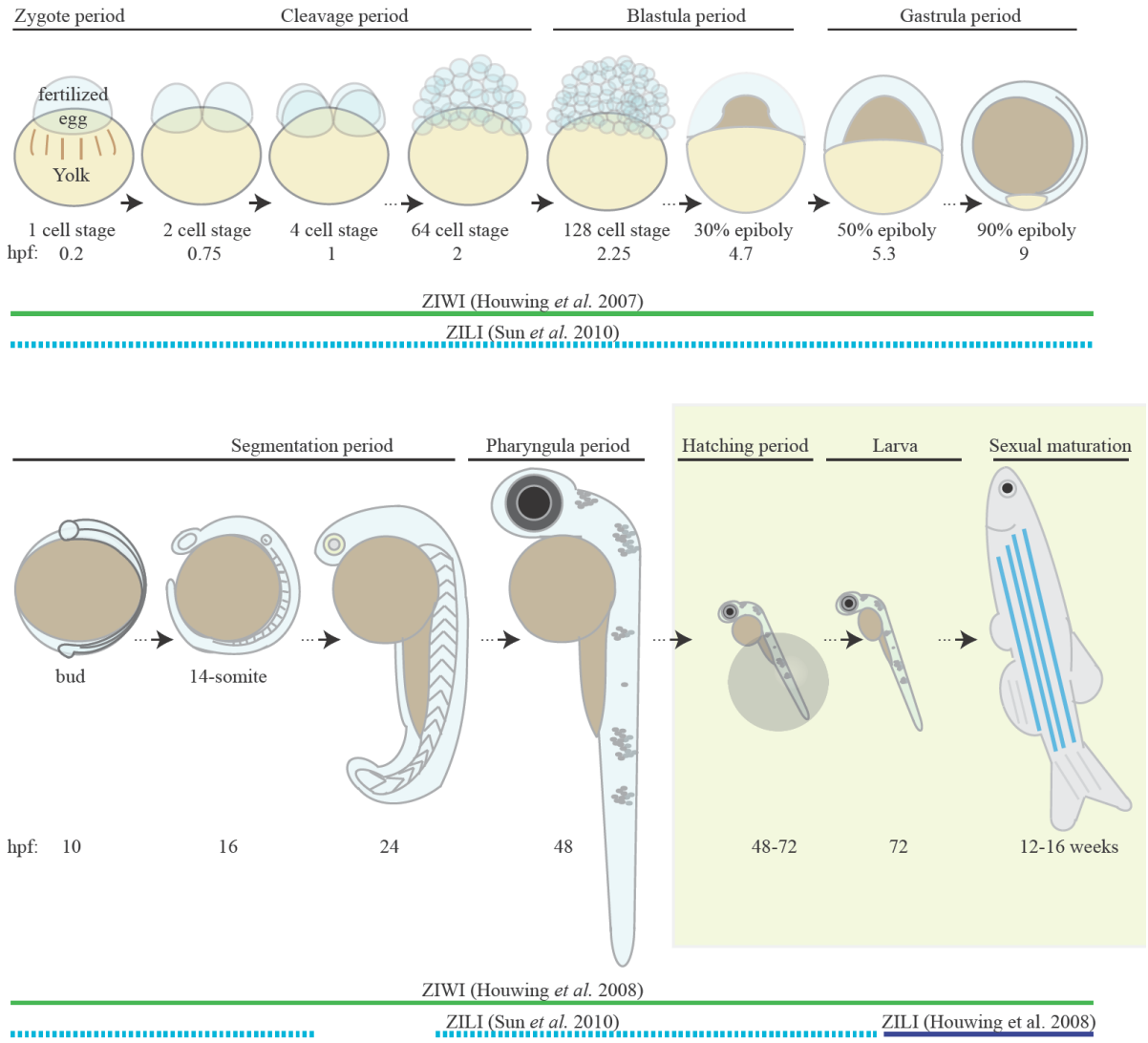


Figure 21: Piwi expression during zebrafish development
The embryonic development of zebrafish takes 48-72 hours post fertilization (hpf) and can be roughly divided in six periods: zygote, cleavage, blastula, gastrula, segmentation, pharyngula and hatching. After leaving the egg, the larva develops to sexual maturity after 12-16 weeks (green box, 1:10 scale compared to previous images). The Piwi protein ZIWI can be observed throughout development (Houwing *et al.* 2007). ZILI was found to be expressed from 72 hpf onwards (Houwing *et al.* 2008) (dark blue line), but Sun and described ZILI already in the one-cell stage of the embryo (Sun *et al.* 2010) (light blue dashed line). However, Sun describes a cap of expression of ZILI around 24 hpf. The image was adapted from (Kimmel *et al.* 1995).

3.3.3.1 The putative helicase MOV10L is involved in the primary piRNA biogenesis

One factor shown to be associated with the primary piRNA biogenesis is the fly helicase Armitage (Olivieri *et al.* 2010; Saito *et al.* 2010) and its mouse homolog MOV10L (named after its homology to protein Moloney Leukaemia Virus 10 (MOV10)) (Frost *et al.* 2010; Zheng *et al.* 2010), which are both expressed in the gonads (see section 1.4.4).

The testes-specific MOV10L contains a helicase domain and two highly conserved regions of unknown function (Frost *et al.* 2010; Vagin *et al.* 2009; Zheng *et al.* 2010). The same domain architecture is shown for ubiquitously expressed helicase MOV10, with which MOV10L shares 45 % amino acid identity (Zheng *et al.* 2010). MOV10 localizes in the cytoplasmic P-bodies (Meister *et al.* 2005; Wulczyn *et al.* 2007), where it associates with the mature-miRNA-containing RISC by interacting directly with Argonaute proteins. This interaction is required for miRNA-guided mRNA cleavage (Chendrimada *et al.* 2007; Meister *et al.* 2005).

Armitage is expressed in germline cells (Cook *et al.* 2004) and can also be detected in somatic follicle cells (Olivieri *et al.* 2010). MOV10L is only expressed in testis with low abundance in oocytes of mice (Frost *et al.* 2010; Lovasco *et al.* 2010; Wang *et al.* 2001b; Zheng *et al.* 2010). While *armitage* mutants are sterile in both genders, *mov10l* mutants impairs only the male fertility (Cook *et al.* 2004; Frost *et al.* 2010; Zheng *et al.* 2010), indicating their importance for the germline development. A direct interaction of Armitage and Piwis is supported by the result of Zheng *et al*. that show the mouse Armitage homolog MOV10L interacts with all three mouse Piwi proteins (Zheng *et al.* 2010) linking the helicases to the piRNA pathway.

piRNAs analyzed from MOV10L immunoprecipitation reveal that most of the piRNAs were associated with MILI and less were bound in MIWI complexes. This is confirmed by the co-localization of MOV10L with MILI in different stages of testes development (Zheng *et al.* 2010). It is suggested that MOV10L is required for piRNA biogenesis (Zheng *et al.* 2010), their stability or for the loading of them into Piwi proteins, as in the *mov10l* knockout neither MILI nor MIWI2 are loaded with piRNAs (Frost *et al.* 2010; Zheng *et al.* 2010). The loss of piRNAs results in activation of retrotransposons like LINE and IAP elements; this is accompanied by a loss of DNA methylation marks on their promoter elements. Additionally it was observed that the predominantly nuclear localization of MIWI2 is disturbed in *mov10l* knockout mice, in which it localizes in the cytoplasm

(Zheng *et al.* 2010). Similar effects were observed in *armitage* mutants, where the level of PIWI is reduced and the protein is mis-localized (Olivieri *et al.* 2010).

How MOV10L facilitates piRNA loading into Piwi proteins or production of small RNAs is not clear. In contrast to fly and mice, the homolog of MOV10L in zebrafish is not described yet. The protein was identified based on sequence alignments and its expression profile was investigated via RT-PCR analysis, indicating the RNA to be maternally derived. Knockdown experiments of *Dr*MOV10L were performed via morpholinos revealing that *Dr*MOV10L is required for transposon silencing in zebrafish as it was shown in mouse (Frost *et al.* 2010; Zheng *et al.* 2010).

3.3.3.2 The Tudor domain-containing protein TDRD1 in the piRNA pathway

Specific arginine residues in the N-terminus of the Piwi proteins are symmetrically dimethylated by the methyltransferase PRMT5 (see Figure 7 and section 1.4.2) and are necessary for the interaction with Tudor domain-containing proteins (Chen *et al.* 2009; Kirino *et al.* 2009; Nishida *et al.* 2009; Reuter *et al.* 2009; Vagin *et al.* 2009; Vasileva *et al.* 2009; Wang *et al.* 2009a).

Tudor domains are present in many protein sequences for most species (see section 1.4.3). The name Tudor (Ponting 1997), originate from the *Drosophila* TUDOR protein (2515 aa), which is required for the assembly of germ plasm. The F1 offspring of *tudor* mutant homozygous females are sterile (Grandchild-less family of genes) (Boswell & Mahowald 1985), which was the curse for the English royal Tudor family, hence the protein's name. The TUDOR protein contains 11 repeated Tudor domains (Talbot *et al.* 1998) which function in the germline-specification and polar granular architecture (Arkov *et al.* 2006; Thomson & Lasko 2004). TUDOR binds to the *Drosophila* Piwi protein Aubergine recognizing the symmetrically dimethylated arginines and is essential for the localization of Aubergine to the germ plasm (Kirino *et al.* 2010b) (see section 1.4.3).

In mice, 10 members of the Tudor domain-containing protein family are expressed (Chuma *et al.* 2003; Chuma *et al.* 2006). Of these, seven (TDRD1-9, except TDRD3 and 5) are found to associate with Piwi proteins (Chen *et al.* 2009; Kirino *et al.* 2010a; Kojima *et al.* 2009; Shoji *et al.* 2009; Vagin *et al.* 2009; Vasileva *et al.* 2009; Wang *et al.* 2009a). All analyzed mutants of the mentioned TDRDs are male sterile (Aravin *et al.* 2009; Shoji *et al.*

2009; Vasileva *et al.* 2009). TDRD1 (129 kDa) contains a N-terminal myeloid–nervy–DEAF-1 (MYND) domain (163-199 aa) and four Tudor domains (307-361 aa, 536-589 aa, 756-809 aa and 982-1034 aa) (Chuma *et al.* 2003). These Tudor domains are responsible for protein interactions, recognizing symmetrically dimethylated arginines (Chen *et al.* 2009; Kirino *et al.* 2009; Nishida *et al.* 2009; Reuter *et al.* 2009; Vagin *et al.* 2009; Vasileva *et al.* 2009; Wang *et al.* 2009a).

TDRD1 interacts with MILI in an RNA-independent manner throughout the germ cell development; from the primordial germ cells through to haploid post meiotic germ cells (Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a). It recognizes also the post-translational modifications of MILI and MIWI (and perhaps MIWI2) (Kojima *et al.* 2009; Reuter *et al.* 2009; Vagin *et al.* 2009).

tdrd1 mutant testes are smaller, because the spermatogenesis is blocked after the pachytene-diplotene stage leading to apoptosis (Chuma *et al.* 2006). The levels of the MILI protein and the piRNAs are not modified in *tdrd1*-null spermatocytes (Reuter *et al.* 2009; Wang *et al.* 2009a), but the identity of the MILI bound piRNA is changed towards genic piRNAs, showing an overrepresentation of sense-stranded exonic reads and a lack of MIWI2 bound piRNAs (Reuter *et al.* 2009; Vagin *et al.* 2009). MIWI2 is not only unloaded in *tdrd1* mutants, its nuclear localization is shifted to the cytoplasm (Reuter *et al.* 2009; Vagin *et al.* 2009), as it is in *mili* mutants (Aravin *et al.* 2007b). Furthermore, the loss of TDRD1 activates LINE1 elements due to a decrease of DNA methylation during the developmental window when MILI is required to silence transposons by *de novo* DNA methylation (Aravin & Bourc'his 2008; Kuramochi-Miyagawa *et al.* 2008; Reuter *et al.* 2009; Vagin *et al.* 2009).

TDRD1 is a cytosolic protein localized to the intermitochondrial cement in male and female germ cells and to the chromatoid bodies in the male mice. It interacts with the chromatoid body marker mouse VASA homolog (MHV) (Chuma *et al.* 2003; Chuma *et al.* 2006). The medaka TDRD1 localizes to granules in the cytoplasm in primordial germ cells in the male and female germline. When the primordial germ cells arrive in the primordium, the medaka TDRD1-containing granules expand, creating a hollow area, which stays visible during the sex determination. The oocytes undergo cystic division and enter then meiosis. During this time the granules shrink and localize in close proximity to the nucleus (Aoki *et al.* 2008), similar to the chromatoid body in mice.

Interestingly, the localizations of proteins specifically expressed in primordial germ cells in zebrafish, like TDRD7, Vasa and Nanos1, are dependent on the 3'UTR in their mRNAs. These 3`UTRs of testes-specific mRNAs contain a miRNA binding site which is flanked by two Dead end protein 1 (Dnd1) binding motifs. The Dnd1 protein is only expressed in the germline, while in somatic cells the miRNA can bind to the mRNA binding site of the 3`UTR recruiting the post-transcriptional silencing complex, which leads to repression. In the germline the Dnd1 binds to the flanking regions, leading to an inaccessibility of the miRNA binding site and the proteins are expressed (Kedde *et al.* 2007).

In contrast to TDRD1 in mice, the zebrafish homolog is not described in the literature yet. Also unknown is whether ZILI and ZIWI carry symmetrically methylated arginine/glycine repeats in their N-termini like Piwi proteins in *Drosophila* and mouse.

Within this study, the zebrafish homolog of TDRD1 (*Dr*TDRD1) was identified via sequence alignments. Expression profiling of *Dr*TDRD1 indicated that it is maternally contributed and persists throughout the lifecycle. In localization experiments, *Dr*TDRD1 was identified in the primordial germ cells, whereby this localization is dependent on its 3`UTR. Although, *Dr*TDRD1 is a testes-specific protein, expressed similar to Piwi proteins, it was not observed to interact with ZILI or ZIWI, or associate with piRNAs.

3.3.4 Results

3.3.4.1.1 Identification of the homolog of MOV10L in zebrafish

MOV10L is a putative helicase involved in the primary piRNA biogenesis in mice (Frost *et al.* 2010; Vagin *et al.* 2009; Zheng *et al.* 2010) (see section 1.4.4 and 3.3.3.1). To identify the zebrafish homologs, a pBLAST search was performed and a single sequence was found. The putative zebrafish *Dr*MOV10L comprises 1106 aa with a predicted molecular weight of 122.6 kDa. It is 47.9 % identical and 68.6 % similar at the amino acid level to the mouse MOV10L. The helicase domain of the zebrafish *Dr*MOV10L (668-1106 aa) accommodates an ATP binding site (701-705 aa) and a putative Mg^{2+} binding site (814-816 aa) (see Figure 22). *Dr*MOV10L contains two domains of unknown functions (~34-212 aa and ~222-512 aa), which are highly conserved in homologs of other species.

3.3.4.1.2 Identification of *Dr*TDRD1

The mouse TDRD1 is a Tudor domain-containing protein which is likely to interact with all three murine Piwi proteins and is needed for transposon silencing in the mouse germline (Kojima *et al.* 2009; Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a) (see section 1.4.3 and 3.3.3.2). The zebrafish homolog of *Dr*TDRD1 consists of 1176 aa and has a calculated molecular weight of 129.2 kDa. It showed 33.7 % identity (63.7% similarity) at the amino acid level to the mouse TDRD1 and was identified to carry the same domains (see Figure 23): These domains include the MYND domain (83-119 aa) with highly conserved cysteine repeats (see Figure 23) and four predicted Tudor domains (173-293 aa, 405-523 aa, 627-745 aa and 876-995 aa, respectively). Furthermore, the binding pocket responsible the recognition of symmetrically dimethylated arginines on other proteins (*e. g.* Piwis, Vasa-like proteins) is also present in *Dr*TDRD1 (see Figure 23).

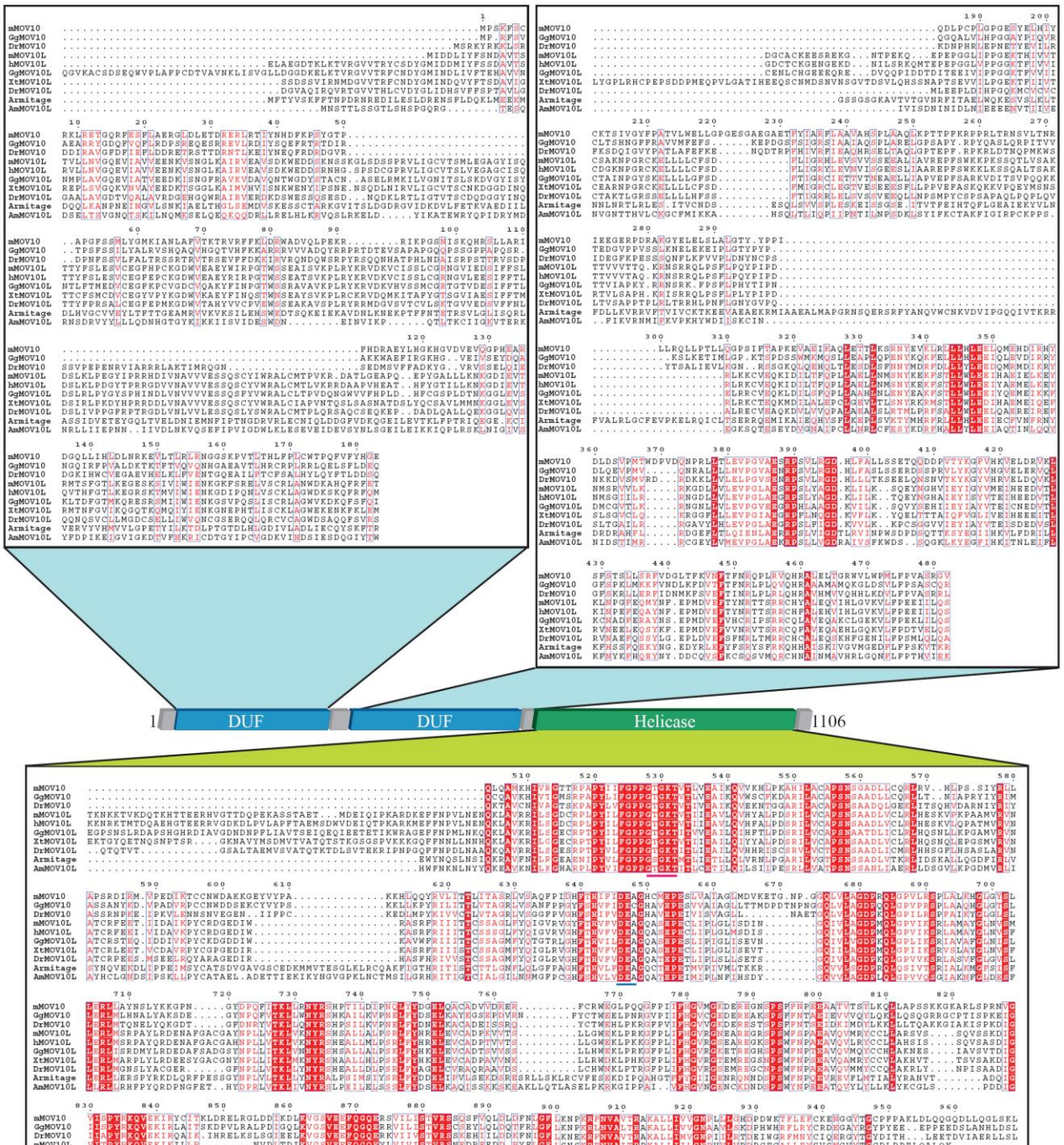


Figure 22: *Dr*MOV10L alignment

ClustalW alignment of MOV10and MOV10L homologs from mouse (mMOV10), *Gallus gallus*, and *Danio rerio* were aligned with MOV10L of mouse, human, *Gallus gallus*, *Xenopus tropicalis*, *Danio rerio*, *Apis mellifera* and *Drosophila* Armitage were aligned using. The image was generated with ESPript. Identical residues in all aligned sequences are highlighted with red boxes and conserved substitutions in red letters. Two domains of unknown function (DUF) and the helicase domain are highlighted. The violet line indicates the ATP binding site and the dark blue line the Mg^{2+} binding site.

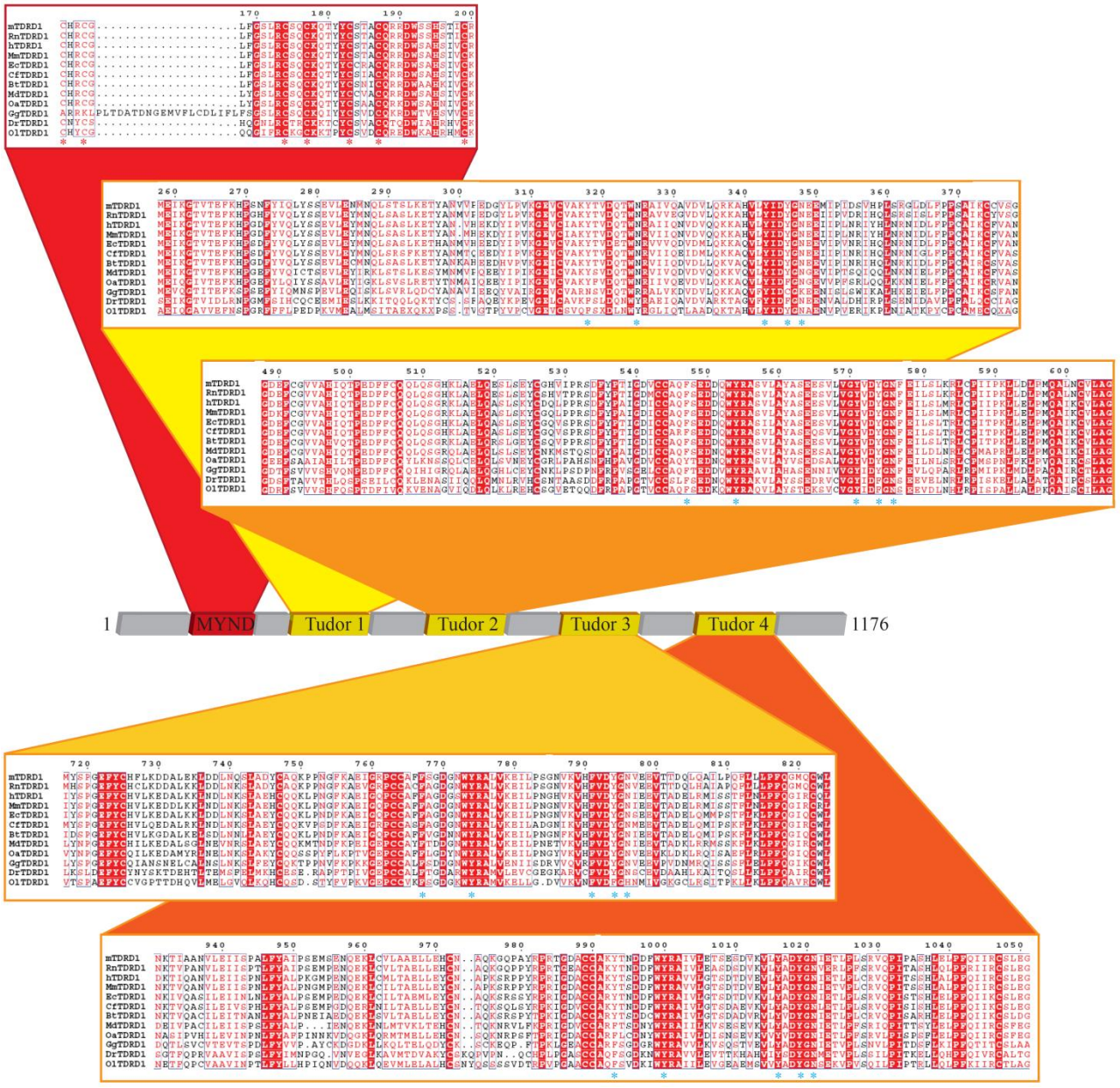


Figure 23: *Dr*TDRD1 alignment

TDRD1 from mouse, *Rattus norvegicus*, human, *Macaca mulatta*, *Equus caballus*, *Canis familiaris*, *Bos taurus*, *Monodelphis domestica*, *Ornithorhynchus anatinus*, *Gallus gallus*, *Danio rerio* and *Oryzias latipes* were aligned and analyzed as in (A). The MYND domain is marked with a red box and the conserved Cysteine repeats are indicated with red stars (*). The Tudor domains are indicated with yellow (Tudor domain 1), dark orange (Tudor domain 2), light orange (Tudor domain 3) and orange-red (Tudor domain 4). The putative aromatic cage building residues, based on similarity to the Tudor domain 11 of TUDOR of Drosophila (Liu *et al.* 2010) indicated with blue stars (*).

3.3.4.2 Expression profiles of *Zili*, *Drtdrd1* and *Drmov10l* during embryonic development

Due to the contradictory literature regarding the onset of ZILI expression as well as the open question of the expression profiles of the putative helicase *Dr*MOV10L and the *Dr*TDRD1, RT-PCRs were carried out across the lifecycle. Therefore, RNA was extracted from eggs (0 hpf) and embryos at different developmental stages (2, 4, 6, 8, 12, 24, 48, 72, 96 hpf). RT-PCRs with specific primers for *Ziwi*, *Zili*, *Drmov10l*, *Drtdrd1* and actin (as a normalization control) were performed and the expression of these genes analyzed by agarose gel electrophoresis (see Figure 24). As expected the *Actin* control was present throughout the embryonic development, indicating that similar amounts of RNAs were used at each time point (see Figure 24). All remaining transcripts (*Ziwi*, *Zili*, *Drmov10l*, and *Drtdrd1*) were detected at 0 hpf, indicating a maternal contribution of these mRNAs. Interestingly, the amount of *Zili* continually decreased from this point, reaching the minimum at 48 hpf, followed by a slight increase. *Drmov10l*, *Drtdrd1* and *Ziwi* (see Figure 24) are expressed throughout our experimental time points. The RT-PCR signal of *Drtdrd1* and *Drmov10l* decreases slightly with time, but to a lesser extent than *Ziwi*, while the *Actin* signal increases slightly.

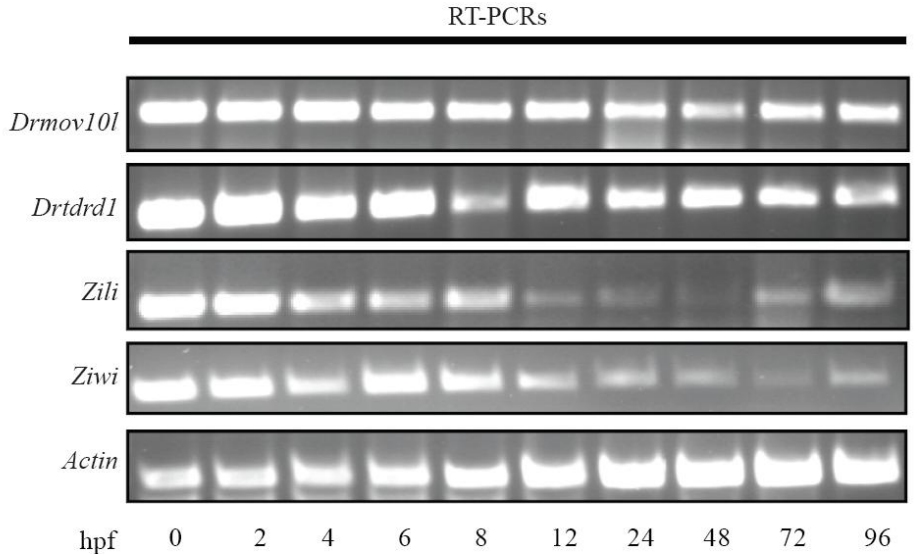


Figure 24: RT-PCRs to determine the expression of fish piRNA-pathway related mRNAs.
RT-PCRs using specific primers for *Drmov10L*, *Drtdrd1*, *Zili*, *Ziwi* and *Actin* mRNAs from unfertilized eggs and embryos of different developmental age (time in hours post fertilization; hpf). All *Drmov10l*, *Drtdrd1*, *Ziwi*, *Zili* and *Actin* are expressed from the unfertilized egg onwards throughout development. The *Zili* expression gets weaker between 12 and 48 hpf to increase at 72 hpf again. *Drmov*10l, *Drtdrd1*and *Ziwi* appear to decease during the time frame. However, this is not normalized for cell differentiation and is therefore an artifact.

3.3.4.3 Development and testing of anti- *Dr*TDRD1 and *Dr*MOV10L antibodies

The biochemical characterization of the endogenous zebrafish homologs of TDRD1 and MOV10L required tools to detect and isolate these proteins. Antisera against the C-terminal region of *Dr*TDRD1 (1011-1176 aa) and the N-terminus of *Dr*MOV10L (1-143) (see Figure 25 A) were raised in rabbits. Western blotting analyses with the affinity-purified antibodies detected the respective tagged proteins expressed in HEK293 cell (Myc-tagged *Dr*MOV10L: ca. 125 kDa and HA-*Dr*TDRD1: ca. 130 kDa) (see Figure 25 B).

To verify if *Dr*TDRD1 and *Dr*MOV10L are expressed in the gonads as their mouse homologs are, fish testes extract was obtained from frozen zebrafish, testing if working with frozen tissues would allow samples to be stored. The proteins were blotted to a membrane and analyzed using the affinity-purified antibodies against *Dr*TDRD1 and *Dr*MOV10L. It was not possible to detect good signal from these extracts (see Figure 25 B).

Therefore, the experiment was repeated with fresh ovary, testes and muscle extracts. In this experimental set up α-*Dr*TDRD1 detects a strong band of about 130 kDa in testes extract, but not from ovaries or muscles (see Figure 25 C), suggesting that the expression of *Dr*TDRD1 is testes-specific and sensitive to freeze-thaw treatment. Thus, all further experiments were performed on fresh material. α-MOV10L was not able to detect a specific band in the extracts (see Figure 25 C). The membrane was re-stained with an α-mouse Tubulin antibody and observed that the amount of the zebrafish Tubulin (ca. 50 kDa) in the different extracts is comparable.

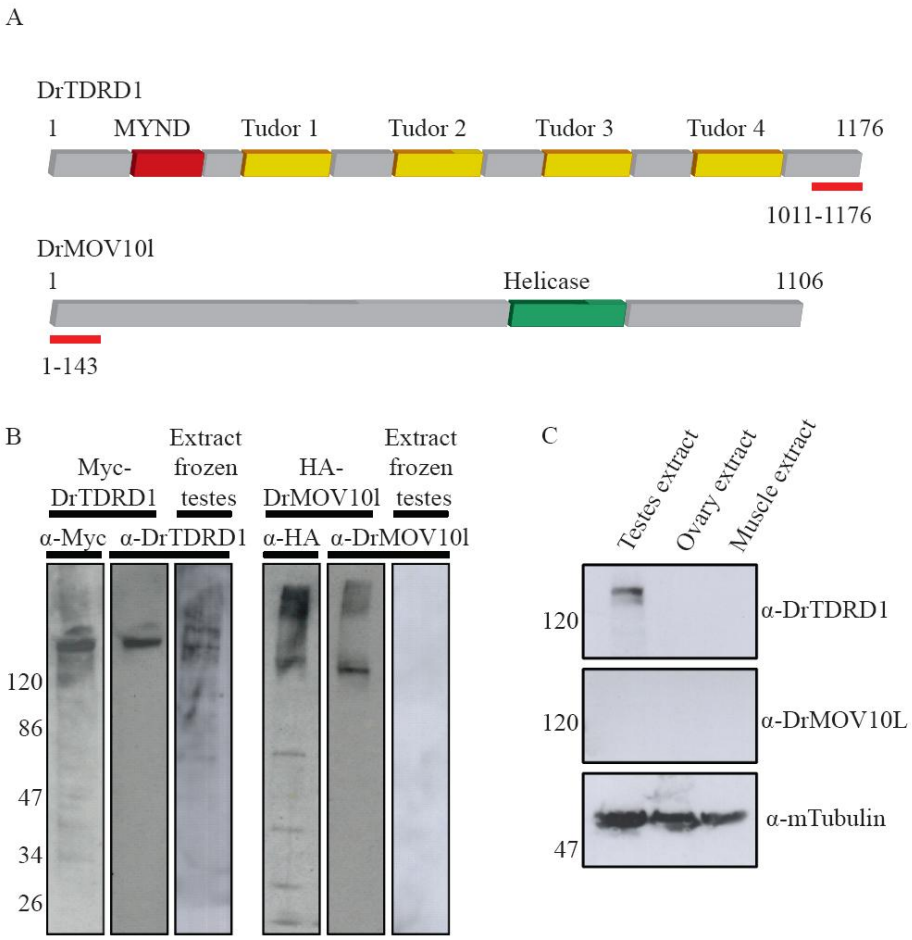


Figure 25: Tests of α-*Dr*TDRD1 and α-*Dr*MOV10L antibodies

(A) Antibodies against *Dr*TDRD1 and *Dr*MOV10L were designed to bind in a non conserved region (*Dr*TDRD1 peptide 1011-1176 aa; *Dr*MOV10L 1-143 aa). In the schematic view the domain architecture of *Dr*TDRD1 and *Dr*MOV10L is shown and the antibody binding regions are marked with a red line.

(B) The antibodies were extensively tested. The α-*Dr*TDRD1 was tested against Myc-tagged TDRD1 expressed in HEK293 cells, which is detected by the α-Myc and α-*Dr*TDRD1 antibodies. In testes extracts of frozen fish α-*Dr*TDRD1 could only provide a weak signal. The α-*Dr*MOV10L antibody detects the HEK293 expressed HA-tagged protein, like the α-HA antibody, but in testis extract from frozen fish no signal could be observed. The indicated size of the marker is given in kDa.

(C) Testes, ovaries and muscle extract were prepared from fresh zebrafish and analyzed by Western blot using the α-*Dr*TDRD1, α-*Dr*MOV10L and as a control mouse α-Tubulin antibody as indicated. While Tubulin was present in all samples as expected, endogenous *Dr*TDRD1 was detected only in testes extract. The α-*Dr*MOV10L antibody did not indicate the endogenous *Dr*MOV10L. The indicated size of the marker is given in kDa.

3.3.4.4 Attempt to link the endogenous *Dr*MOV10L and *Dr*TDRD1 directly to the piRNA pathway in zebrafish

*Dr*TDRD1 is a testes-specific protein (see section 3.3.4.3), which could be involved in the piRNA pathway by recognizing potential symmetrically dimethylated arginines on the zebrafish Piwi proteins by its Tudor domains. It is unknown whether the zebrafish ZIWI and ZILI carry similar post-translational modifications.

To analyze the piRNA profile of the zebrafish Piwi proteins, endogenous ZILI and ZIWI complexes were immunopurified from adult zebrafish ovary extract using specific affinity purified antibodies. To investigate if the Piwi proteins carry symmetrically di-methylated arginines a specific antibody recognizing the symmetrically dimethylated arginines but not the Piwi protein itself, called Y12 (Reuter *et al.* 2009), was used for immunoprecipitations as well. To control the experiment, endogenous MILI complexes were immunoprecipitated from mouse testes extract. The complex-associated RNAs were isolated and after 5`-end-labelling separated on a denaturating gel.

In comparison to the MILI immunoprecipitation (see Figure 26 A lane 6), the amount of ZILI and ZILI bound piRNAs is significant lower (see Figure 26 A lane 3 and 4). No RNAs were observed using beads without specific antibodies for the immunoprecipitation (see Figure 26 A second panel). Immunoprecipitations of ZIWI and ZILI co-purified RNAs of 27 nt and 26 nt, respectively (see Figure 26 A lane 3 and 4). Immunoprecipitations using the Y12 antibody indicated that at least one zebrafish protein must carry symmetrically dimethylated arginines, due to the fact that RNAs of 26 nt were present (see Figure 26 A). In comparison to the immunoprecipitations of ZIWI (27 nt) and ZILI-bound RNAs (26 nt), the size of the Y12 pull down RNAs are 26 nt, which indicates that they originate from ZILI-containing complexes. However, some contribution from ZIWI cannot be excluded.

Assuming that *Dr*TDRD1 is an interaction partner for the zebrafish Piwi proteins, piRNAs bound in this complex could be co-purified. Similarly, if *Dr*MOV10L interacts with ZIWI or ZILI bound piRNAs this could indicate complex formation. Therefore, the *Dr*TDRD1 complex was immunoprecipitated from adult zebrafish testes and ovaries using affinity purified antibodies. Complex bound RNAs were extracted and, after 5` labeling, analyzed on a denaturating gel. In the *Dr*TDRD1 immunoprecipitations no specific small RNAs were enriched, indicating that no piRNAs were co-purified (see Figure 26 B ^{32}P gel).

*Dr*MOV10L immunoprecipitations also did not indicate any bound piRNAs (see Figure 26 C). In summary, at least one of the zebrafish Piwi proteins is potentially post-translational modified via symmetrically dimethylated arginines as indicated by immunoprecipitations using an antibody recognizing these modifications pulling down RNAs in the size range of piRNAs. Immunoprecipitations with α-*Dr*TDRD1 and α-*Dr*MOV10L did not indicate any complex formation with piRNAs.

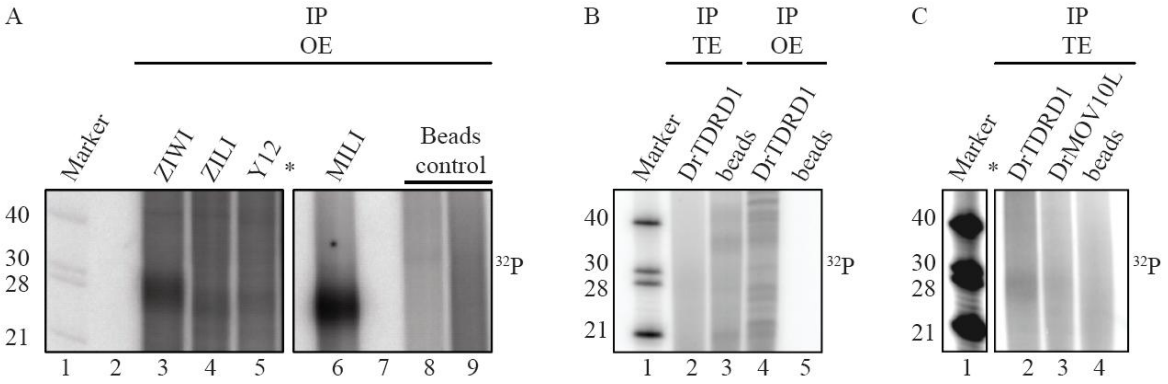


Figure 26: Immunoprecipitations of zebrafish piRNA pathway componments

(A) Analysis of co-purified RNAs from immunoprecipitations using α-ZIWI, α-ZILI and Y12 antibodies on adult zebrafish ovary extract (OE) and monoclonal α-MILI on adult mouse testes extract.. As controls beads without coupled antibodies were used, incubated with the fish ovary extract (left lane) or mouse testes extract (right lane). Lane 2 and 7 are empty

(B) Immunopurified and extracted RNAs from the *Dr*TDRD1 complex from adult zebrafish testes extracts (TE). No RNAs in the size range of 26-27 nt, indicating bound piRNAs, were detected.

(C) Immunopurified and extracted RNAs from the *Dr*TDRD1 and *Dr*MOV10l complexes treated as in the other experiments of the figure. No piRNAs were detected.

(*) denotes empty lanes deleated for display purpose, intensities can be directly compaired in sub-panels as they are taken from the same image without manipulation.

3.3.4.5 *Dr*TDRD1 and *Dr*MOV10L interaction studies

Interaction studies of *Dr*MOV10L and *Dr*TDRD1 with the zebrafish Piwi proteins ZIWI and ZILI were performed, to link the Tudor domain-containing protein and the helicase directly to the piRNA pathway. In a heterologous system using HEK293 cells, HA-tagged constructs of full-length *Dr*TDRD1 and *Dr*MOV10L were co-expressed with Myc-ZIWI and Myc-ZILI. The HA-tagged proteins were purified with α-HA beads from HEK293 cell lysates prepared 48 hours post-transfection and subjected to Western analyses using α-Myc antibodies to detect the possible co-purified Piwi proteins. Unfortunately, the HA-tagged *Dr*TDRD1 and *Dr*MOV10L were lowly expressed (see Figure 27 A, α-HA). The Myc-tagged ZILI and ZIWI proteins are present in the input sample, but they were not co-immunoprecipitated by HA-*Dr*TDRD1 or HA-*Dr*MOV10L (see Figure 27 A). Mutual interaction experiments of Myc-*Dr*TDRD1 and HA-*Dr*MOV10L did not indicate any association (see Figure 27 B).

MILI and the mouse TDRD1 are known to interact via a symmetrical dimethylation on the argenine 74 in the N-terminus of MILI (Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a) (see section 1.4.3). ZILI shows similar arginine/glycine repeats in the N-terminal region like many orthologs in different species (see Figure 27 C). To test the possible recognition of a homologous region containing arginine/glycine in the N-terminus of ZILI by *Dr*TDRD1, *in vitro* translated *Dr*TDRD1 were incubated with biotinylated synthetic peptides corresponding to the amino acids 158-168 of ZILI, where R163 was either unmodified or symmetrically dimethylated (see Figure 27 green mark). Precipitations with Streptavidin beads revealed that *Dr*TDRD1 was only co-purified when the peptide carried a methylation mark (see Figure 27 D), but the achived signal was weak. However, in comparison with *in vitro* translated mouse TDRD1 binding to the methylated MILI peptide, the signal of *Dr*TDRD1 recognizing the methylated ZILI peptide is hardly detectable. Even the signal of mouse TDRD1 interacting with methylated ZILI peptides is higher than the *Dr*TDRD1/ZILI peptide association (see Figure 27 E), but the signal of input sample of TDRD1 is much stronger than the one for *Dr*TDRD1 as well.

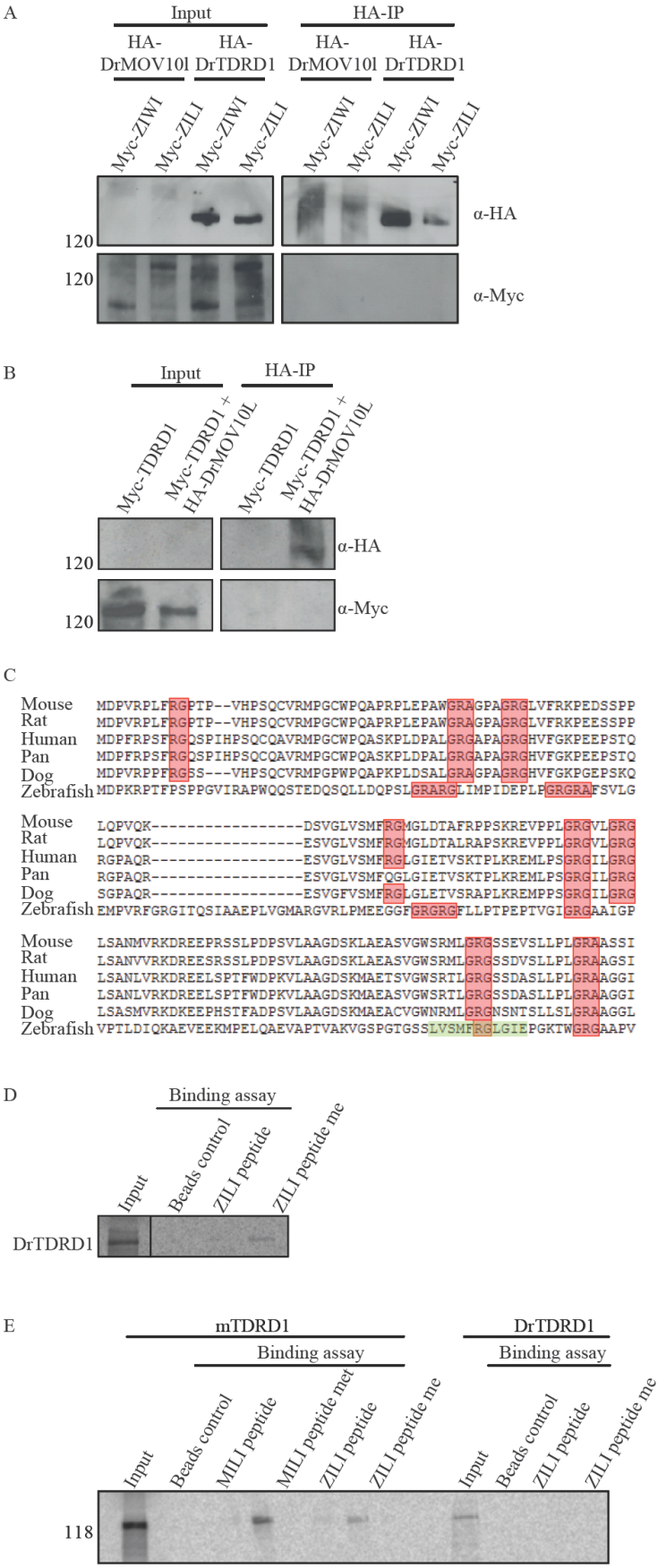


Figure 27: Binding studies of *Dr*TDRD1 and *Dr*MOV10L

(A) HA-*Dr*TDRD1 and HA-*Dr*MOV10L were co-expressed with Myc-ZILI and Myc-ZIWI in HEK293 cells. The HA-tagged proteins were immunoprecipitated using an α-HA matrix, after 1/20th of the soluble part of the cell extract was taken as a input sample. The Myc-tagged zebrafish Piwi proteins were not co-purified by HA-*Dr*TDRD1 or HA-*Dr*MOV10L, as shown in the Western blot analysis using α-Myc antibodies.

(B) Co-expressions of Myc-*Dr*TDRD1 and HA-*Dr*MOV10L using the same method as before did not reveal any association of the two proteins.

(C) Comparison of the N-terminal region of MILI orthologs. Marked in red are RG/GR repeats. In green highlighted is the sequence of the peptide used in the binding assay.

(D) *In vitro* translated full-length *Dr*TDRD1 was incubated with biotinylated peptides, which were unmodified or carried a symmetrical dimethylation at R163. After purification, the radioactive marked *Dr*TDRD1 could only be detected in context with the methylated peptide and not with the unmodified one.

(E) The binding assay shown in **(D)** was repeated in comparison with mTDRD1. mTDRD1 associated with a MILI peptide containing the symmetrically di-methylated R74 and recognized even the methylated ZILI peptide, while the interaction of *Dr*TDRD1 with the ZILI peptides was too weak to detect.

3.3.4.6 Localization of *Dr*TDRD1

The medaka and the mouse TDRD1 are mainly expressed in the male germline. In both species, the TDRD1 was found to localize in cytoplasmic granules (Aoki *et al.* 2008; Chuma *et al.* 2003) (see section 3.3.3.2). The localization of *Dr*TDRD1 was investigated using fluorescent reporter constructs. Therefore, one-cell stage zebrafish embryos were injected with RNA of GFP-tagged full-length *Dr*TDRD1. After 24 h the embryos were examined using fluorescence microscopy. The fluorescence signal of GFP-*Dr*TDRD1 was detected in the primordial germ cells, which are distinguishable in the tail region close to the yolk (see Figure 28 A).

The reason why *Dr*TDRD1 is just expressed in the primordial germline and not in the other tissues of the embryo was intriguing. Based on the findings of Kedde *et al.* in 2007, that the 3'UTRs of testes-specific mRNAs of proteins are responsible for germline-specific expression due to a miRNA-guided silencing event in somatic cells (see section 3.3.3.2), the sequence of the 3`UTR of *Dr*TDRD1 was examined to identify potential miRNA target sites. The online search machine microcosm provided by the EBI identified 6 potential binding sites for miRNAs (*dr*e-miR-145, *dr*e-miR-182, *dr*e-let-7e, *dr*e-miR-29a, *dr*e-miR-146a and *dr*e-miR-193a). Additionally, based on sequence comparison of the 3`UTR of *Dr*TDRD1 with the 3`UTR of Nanos and p27 revealed potential Dnd1 binding sites (Figure 28 B).

To determine if the 3`UTR of *Drtdrd1* alone is sufficient to localize the protein expression in the germline, the sequence was isolated and cloned downstream of a GFP coding region. As controls, a GFP-construct which contained the 3`UTR of Vasa (Kedde *et al.* 2007) was used as positive germline expression marker and a GFP construct without 3`UTR was created to serve as negative control (see Figure 28 C, cartoon).

The RNAs of these construct were injected into zebrafish one-cell stage embryos. 24 hpf the embryos were analyzed using a fluorescent microscope: While the GFP-construct without 3`UTR expressed the fluorescence protein throughout the embryo, embryos injected with GFP constructs containing the 3`UTR of Vasa or the 3`UTR *Dr*TDRD1 expressed the fluorescence protein mainly in the primordial germ cells (see Figure 28 C), indicating that the 3` UTR of *Dr*TDRD1 is responsible for the specific expression of the proteins in the germline.

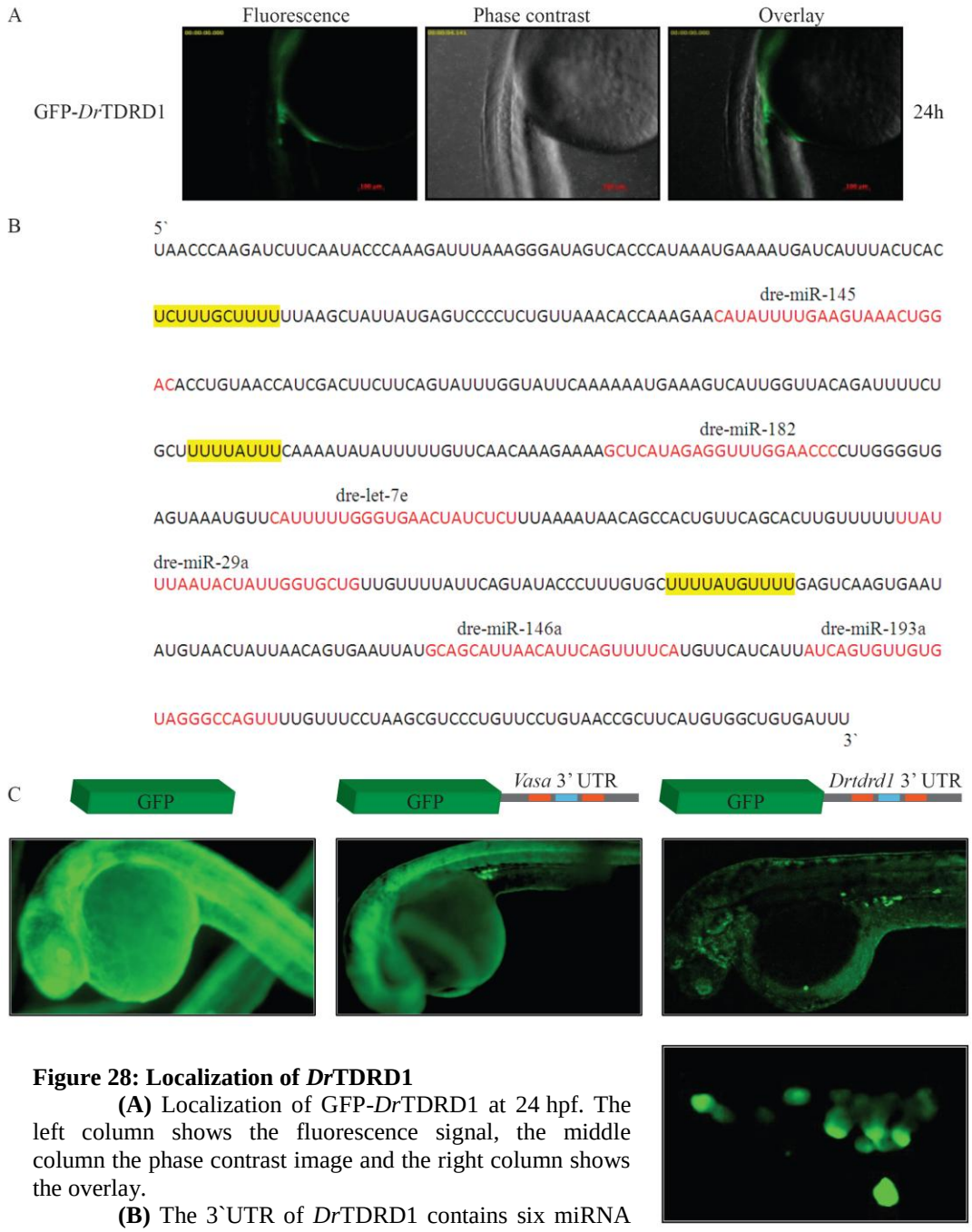


Figure 28: Localization of *Dr*TDRD1

(A) Localization of GFP-*Dr*TDRD1 at 24 hpf. The left column shows the fluorescence signal, the middle column the phase contrast image and the right column shows the overlay.

(B) The 3`UTR of *Dr*TDRD1 contains six miRNA binding sites (red letters). Potential Dnd1 binding sites are boxed in yellow.

(C) 24 hpf embryos injected with RNAs encoding GFP (left schema), GFP with the 3`UTR of Vasa (middle schema) and GFP with the 3`UTR of *Dr*TDRD1. GFP without 3`UTR showed expression in the entire embryo (right image). The *Dr*Vasa 3`UTR shuttles the expression of the GFP to the germline (left image) as shown by (Kedde *et al.* 2007). The same effect can be achieved with a construct containing GFP with the 3` UTR of *Dr*TDRD1. Below the primordial germ cells the GFP expression of the construct with the 3` UTR of *Dr*TDRD1 is show in higher resolution.

3.3.4.7 Knockdowns of *Dr*TDRD1 and *Dr*MOV10L

*Dr*TDRD1 is a testes-specific protein, as indicated by western blot analysis of fish tissue extracts and localization experiments (see section 3.3.4.3 and 3.3.4.4), but its implication in the zebrafish piRNA pathway was not yet shown. Furthermore, a proof for a functional involvement of *Dr*MOV10L as a factor in the piRNA biogenesis was still missing in zebrafish.

To investigate if *Dr*TDRD1 and *Dr*MOV10L are involved in piRNA mediated gene silencing, morpholinos for knockdown experiments in zebrafish were designed. For *Dr*TDRD1 knockdowns, two morpholinos were used: Morpholino 1 (MO1) was constructed to lead to an inclusion of intron1 in *Dr*TDRD1, leading to a pre-mature stop codon in the translation (see Figure 29 A). The second morpholino (MO2) is targeting the exon3-intron3 junction and should result in exon3 deletion (see Figure 29 B). To knockdown the expression of *Dr*MOV10L, morpholino 3 (MO3) was designed to bind the translational start site and block the translation of the protein (see Figure 29 C). All morpholinos were injected into one-cell stage zebrafish embryos. After 24 h the embryos were harvested and total RNA was extracted to be analyzed in RT-PCRs with specific primers. To determine the MO1 activity specific primers complementary to the exon 1 and 2 of *Dr*TDRD1 were used. In case of active MO1 a larger band was expected than in the uninjected wild type (WT). For MO2 primers before and after exon 3 were designed, which would lead to a shorter RT-PCR product if the morpholino is active. Unfortunately, both morpholinos against *Dr*TDRD1, MO1 and 2, showed mainly RT-PCR products of the WT sequence and an insufficient amount of the knockdown construct (see Figure 29 D), at all volumes of morpholinos injected.

Given that MO3 against *Dr*MOV10L is a translation inhibitor, it was not possible to use this method for the activity control. Lacking an active antibody it was impossible to control the activity of Morpholino 3 directly. So, the activity of MO3 was tested indirectly by analyzing a potential change in transposon activity, cause by the reduced amount of *Dr*MOV10L. Hundreds of one-cell stage embryos were injected with MO3 to knockdown the expression of *Dr*MOV10L. Already after 24h the injected embryos are phenotypically different to WT or embryos injected with a control morpholino (MOC) (see Figure 29 E). MO3 embryos display a shorter tail and develop slower than the WT. The survival rate of WT embryos was 85 %, for embryos injected with a control morpholino (MOC) 75 %, while

only approximately 10 % of the MO3 injected embryos survived the first 24 hpf. These results indicated already the importance of *Dr*MOV10L in the embryonic development.

Equal numbers of the surviving embryos were collected and the RNAs extracted. After DNase treatment equal amounts of total RNA were used to perform RT-PCRs using random primers. This cDNAs were applied to preliminary qPCR experiments, using primers to determine the activity of the transposable element *Dr*Gypsy2. The values of the qPCR were normalized against *Dr*Actin and the normalized values of RNAs of the MO3 injected embryos were compared against RNAs of WT embryos. The activity of *Dr*Gypsy2 was just slightly higher than the WT control (0.2 fold difference, see Figure 29 E). The qPCR experiments were repeated in a bigger screen using primers against the transposons EnSPmN1, Ngaro and DIRS1a. To ensure the effect is due to the morpholino and not caused by an overall increase of the gene activity in injected embryos, the activity of Vasa was measured as well. While *Dr*Vasa (see Figure 29 G yellow bar) and *Dr*Actin (see Figure 29 G red bar) did not indicate any change in their activity, the expression of EnSPmN1 was 5 fold higher than in WT embryos (see Figure 29 G light blue bar), the activity of Ngaro (see Figure 29 G royal blue) increased by 18 fold and DIRS1a (see Figure 29 G dark blue bar) was even 24 fold higher expressed than in the WT.

In conclusion, *Drmov10l* knockdowns display increased activity of transposable elements, indicating that the piRNA silencing mechanism is disturbed.

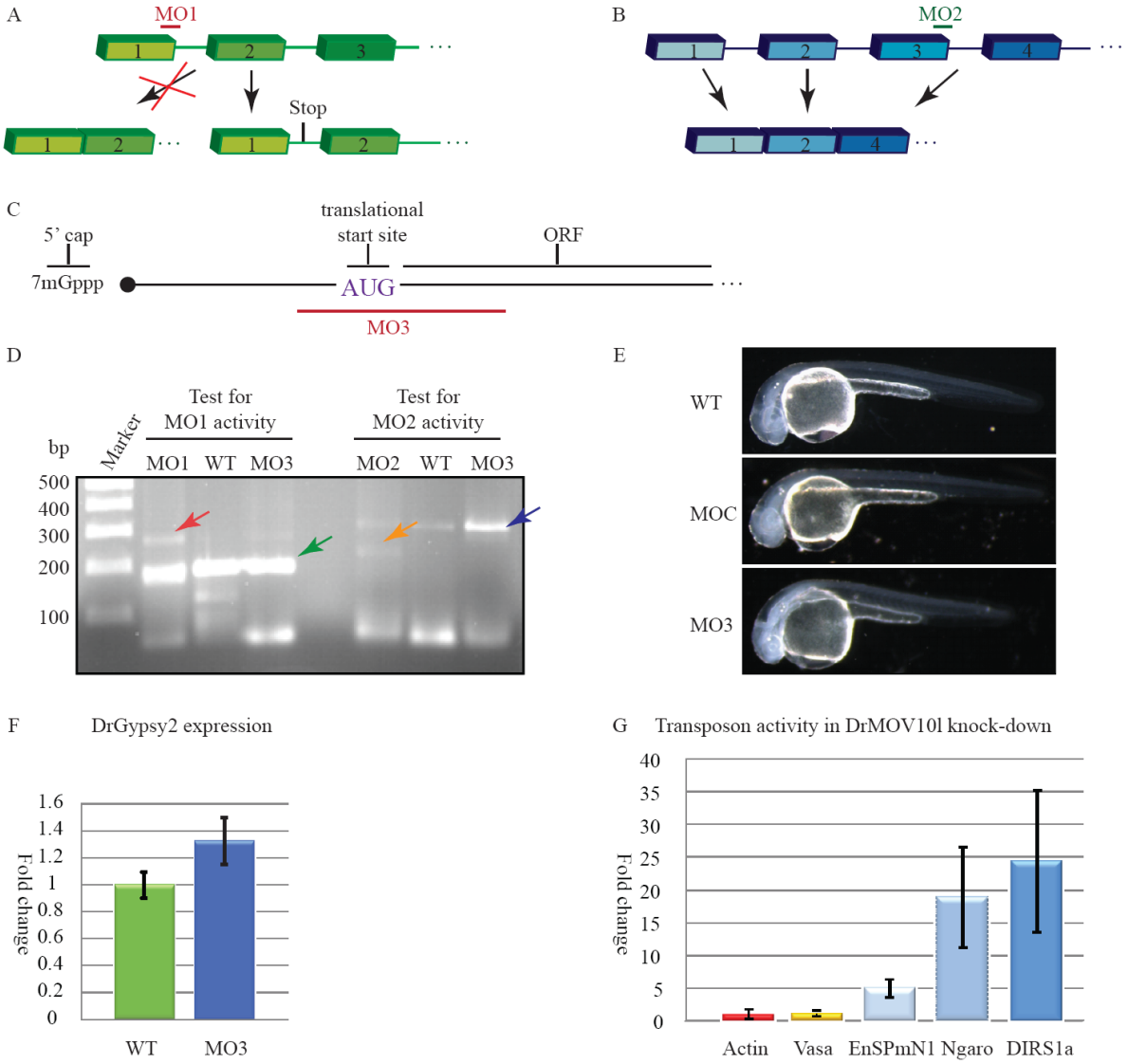


Figure 29: Morpholino knockdowns

(A) Morpholino 1 (MO1) binds at the exon 1 intron junction of *Dr*TDRD1 and leads to an inclusion of the intron 1 into the mRNA of *Dr*TDRD1

(B) Morpholino 2 (MO2) coresponds to the exon3 intron junktion and results in an exon3 deletion in the mRNA of *Dr*TDRD1

(C) Morpholino 3 (MO3) recognizes the translational start site of *Dr*MOV10L and inhibit the translation of the mRNA.

(D) RT-PCRs to reveal the knockdown capasity of MO1 and MO2. Primers flanking the regions of interest were used. For active MO1 a larger band was expected (red arrow) than in the WT (green arrow left side of the gel). For MO2 a shorter RT-PCR product (orange arrow) indicates the morpholino activity in difference to the longer RT-PCR fragments of the WT (blue arrow).

(E) WT embryos (upper row) and embryos injected with a morpholino control (MOC) display a normal development in size. Embryos injected with MO3 are shorter at the same developmental stage.

(F) Preliminary qPCR result of *Dr*Gypsy2 activation in MO3 treated embryos and in WT normalized against the expression of *Dr*Actin. The expression is given in fold differnce.

(G) qPCRs of MO3 treated embryos analysed against the WT and normalized against *Dr*Actin, showed that in difference to *Dr*Vasa, the activity of transposable elements is increased. Data presented as mean ± SD.

3.3.5 Discussion *Dr*MOV10L and *Dr*TDRD1

In this chapter, zebrafish homologs of the mouse piRNA pathway components TDRD1 (*Dr*TDRD1) and MOV10L (*Dr*MOV10L) were identified and analyzed. Transcripts of *Drtdrd1* and *Drmov10l* are detected from the unfertilized egg stage onwards throughout development, as shown by RT-PCR experiments (see section 3.3.4.2). *Dr*TDRD1 is a testes-specific protein, shown via Western blot analysis (see section 3.3.4.3) and localizes in primordial germ cells of zebrafish embryos, where this tissue-specific expression is dependent upon its 3`UTR (see section 3.3.4.6). Unfortunately, interactions of *Dr*TDRD1 and *Dr*MOV10L with Piwi proteins could not be shown. However, to identify a possible link between *Dr*TDRD1, *Dr*MOV10L and the piRNA pathway, knockdown experiments were performed. While the *Drtdrd1* knockdown was ineffective to elucidate its function, *Drmov10l* (see section 3.3.4.7) knockdowns showed an increased activity of piRNA repressable transposable elements (Houwing *et al.* 2008), linking *Dr*MOV10L function to the piRNA pathway.

RT-PCR experiments indicated that the RNA of *Drtdrd*1, *Drmov*10l, as well as from the zebrafish Piwi genes *Zili* and *Ziwi* are all expressed from the one-cell stage onwards throughout embryonic development. The identified *Ziwi* expression is in agreement with earlier studies, showing that the *Ziwi* is maternally distributed (Houwing *et al.* 2007). The RT-PCR results indicate that *Zili* is already present in the one-cell stage, verifying studies of Sun and colleagues (Sun *et al.* 2010). Interestingly, the amount of *Zili* decreased in the RT-PCRs from 12 to 48 hpf with a significant recovery from 72 hpf onwards. Similar results were obtained by Sun *et al.* showing the decrease of *Zili* around 24 hpf, when the protein is no longer expressed in the axis of the embryo but changes towards expression in the primordial germ cells (Sun *et al.* 2010). The decreasing effect of *Zili* is may be caused by the reduction of maternally contributed and stored *Zili* mRNAs until the embryo itself starts to transcribe it in the primordial germ cells. The RT-PCR results of *Drtdrd1* and *Drmov10l* display two clues. First, both mRNAs are maternally contributed and second, the expression pattern of *Drtdrd1* and *Drmov10l* is similar to the ones of *Zili* and *Ziwi*. The observed decrease of the *Drtdrd1* and *Drmov10l* signal is probably due to the cell differentiation occurring in the embryo: while in the early stage embryos, a few cells express the genes of

interest, in later stages the expression is probably localized in specific tissues (primordial germ cells as shown for Zili and Ziwi (Houwing *et al.* 2008; Houwing *et al.* 2007)). The RNAs were extracted from the entire embryo or larva, leading to the impression of a decreasing signal in the total RNA amount.

To be able to perform biochemical experiments with the endogenous *Dr*MOV10L and *Dr*TDRD1, antibodies against these proteins were developed and extensively tested. Both antibodies were able to detect full-length *Dr*MOV10L and *Dr*TDRD1 expressed in HEK293 cells, respectively. The endogenous *Dr*TDRD1 was mainly detected in fresh testes extracts, not in ovary extract and never in muscle extracts. These results are in agreement with the studies of Aoki and colleagues, who showed that the medaka TDRD1 is primarily expressed in testes and to a much smaller extent in ovaries (Aoki *et al.* 2008). *Dr*TDRD1 was only detected in freshly harvested material and not in extracts prepared from frozen zebrafish, attributed due to oxygen-dependent degradation of the material and/or damage by repetitive freeze-thaw cycles. Unfortunately, *Dr*MOV10L was not detected with the specifically designed antibody in Western blots regardless of the origin of the extracts.

Using the α-*Dr*TDRD1 and α-*Dr*MOV10L antibodies immunoprecipitations were performed to determine if the two proteins co-immunoprecipitate piRNAs. In both cases no small RNAs were identified. In terms of the α-*Dr*MOV10L immunoprecipitation it became clear that the antibody was inactive. The absence of piRNAs in the α-*Dr*TDRD1 immunoprecipitation, (although that it was shown that TDRD1 in mice co-purify piRNAs (Reuter *et al.* 2009)) could be due to the small amounts of biological material. Another possibility is that *Dr*TDRD1 interacts with mainly unloaded zebrafish Piwi proteins. In case of such an interaction, it is likely that the Tudor domains of *Dr*TDRD1 would recognize potentially symmetrically dimethylated arginines at ZILI or ZIWI, as the mouse homolog TDRD1 recognizes R74 at the N-terminus of MILI (Reuter *et al.* 2009). To test if ZIWI or ZILI carrying similar post-translational marks, immunoprecipitations recognizing the symmetrically dimethylated arginines were performed and the RNAs analyzed. The presence of piRNAs indicates that at least one Piwi protein is modified with methylation marks. The size of the piRNAs indicates ZILI as a primary modified protein. To analyze potential interactions of *Dr*TDRD1 and *Dr*MOV10L with ZILI and ZIWI, epitopic tagged proteins were co-expressed in HEK 293 cells. *Dr*MOV10L did not express well and no interactions with the ZILI and ZIWI were detected. *Dr*TDRD1 did not interact with Piwi proteins nor

with *Dr*MOV10L. It was speculated that either the proteins are not folded correctly due to a lack of the suitable chaperone proteins in the expression system or that ZILI or ZIWI are not recognized by the PRMT5 protein in the human cells. This would result in Piwi proteins lacking the symmetric dimethylation moiety recognized by the Tudor domains of *Dr*TDRD1.

GFP-*Dr*TDRD1 was found in the migrating primordial germ cells. A GFP-construct containing only the 3`UTR of *Drtdrd*1 injected into *Danio rerio* embryos localized the GFP expression into the primordial germ cells as expected. In most of the localization experiments a background in other tissues, mainly in the back bones was visible. This effect is probably caused by the high amounts of RNA constructs, which were injected. The inhibition of the expression of germ cell-specific proteins in vegetative tissues by miRNAs, which recognize the miRNA binding site in the 3`UTR of these proteins, is efficient, but the artificial constructs overwhelms the capacity of regulation. However, the signals of the primordial germ cells have higher intensity, which leads to the conclusion that the expression in testes of *Dr*TDRD1 is guided by its 3`UTR.

To determine the function of *Drtdrd1* and *Drmov10l*, knockdown experiments using morpholinos against *Drtdrd1* and *Drmov10l* were performed. The morpholinos against *Drtdrd1* disturbed the splicing, but the observed silencing of the mRNA was inefficient. More than 50% of the *Drtdrd1* mRNA was still wild type, while only a minority showed the knockdown variant. It might be possible that the WT construct was still abundant enough to silence the transposable elements, so that the change of the transposon activity was not measurable by qPCR. In knockdowns of *Drmov10l* transposable elements like EnSPmN1, Ngaro and DIRS1a display a significantly higher abundance than in the wild type. In contrast, the activity of the germ cell-specific expressed Vasa and Actin is not affected. The values obtained for the transposon activities in the experiments showed a comparable activity pattern like in knockouts of *zili* (Houwing *et al.* 2008) and is consistent with studies of mouse MOV10L (Zheng *et al.* 2010). In conclusion, *Dr*MOV10L is needed for piRNA-guided transposon silencing.

3.3.6 Methods

3.3.6.1 Identification of *Dr*TDRD1 and *Dr*MOV10L

The known mouse protein sequences of TDRD1 and MOV10L were used in a BLAST search to identify the zebrafish homologs. The sequences identified in different species were aligned using ClustalW2 (*Multiple sequence alignment with the Clustal series of programs. (Chenna et al. 2003)).* The Alignments were analyzed with ESPript (Gouet *et al.* 1999).

3.3.6.2 *Dr*TDRD1 and *Dr*MOV10L expression test by RT-PCR

To analyze the expression of *Dr*TDRD1 and *Dr*MOV10L in compared to zebrafish Piwi protein, RT-PCRs were performed. DNA fragments of ~300 nt in length of

*Dr*MOV10L (primers: 5`-GCGAATTCATGGTGAGTGTGGTGCAGTGTTTAATGTCG-3`,
5`-CCGCTCCACCCGTATGGCTCTCC-3`);

*Dr*TDRD1 (primers: 5`-CTCTCGAGACACCATGAATCCAGCCTTCGCTCAGCC-3`,
5`-CACCGAGTGCACCTTAAGTTGCCC-3`);

ZILI (primers: 5`-GCGAATTCATGGATCCAAAACGACCAACC-3`,
5`-CCT CCC TCT TCC ATA GGT AAC CTC-3`);

ZIWI (primers: 5`-GAATTCATGACAGGACGAGCAAGAGC-3`,
5`-GGATTGTTTCACGTGTCCATAAG-3`),

*Dr*Actin (primers: 5`-ATGTGTGACGACGACGAGACTAC-3`,
5`-GGGGCAACACGGAGCTCATTG-3`)

were analyzed on a 2% agarose gel containing ethidium bromide.

3.3.6.3 Antibodies

For the production of antibodies, GST-fusions of specific protein fragments were produced in *E. coli* strain BL21 and injected into rabbits. The fragments used were anti-*Dr*TDRD1 (1011-1176 aa) and anti-*Dr*MOV10L (1-143 aa). All antibodies were affinity-purified and used at a dilution of 1:200. Anti-ZILI and anti-ZIWI antibodies were kindly provided by the Ketting lab (Houwing *et al.* 2008) or produced based on peptides:

ZILI: MDPKRPTFPSPPGVI+C

ZIWI: MTGRARAR+C

Other antibodies used are: anti-HA antibodies (Santa-Cruz) at 1:250 dilutions and anti-Myc at 1:250 dilutions.

3.3.6.4 Mouse and zebrafish tissue extracts and RNA analysis

CD1 mouse testes extract preparation, immunoprecipitation and 5`-end-labelling of piRNAs were performed as described in the literature (Aravin *et al.* 2006; Girard *et al.* 2006). The same method was used for *D. rerio* AB testes, ovaries and muscle extract preparation and RNA analysis. Dried gels were exposed to a storage phosphor screen and scanned with a Typhoon scanner (Amersham).

3.3.6.5 Clones and cell culture

For mammalian cell transfections, full-length ORFs of Piwi proteins (ZIWI and ZILI) or *Dr*MOV10L or *Dr*TDRD1 were obtained by RT-PCR and were inserted into a pCIneo-NHA vector (Pillai *et al.* 2004), into pcDNA3-Myc (Invitrogen), or in a pSP64T-mGFP or pSP64T-dsRedEX (Kedde *et al.* 2007) vector as required.

For the HEK293 Ebna interaction experiments, 2 µg each of each construct were transfected into 6 cm dishes at 80% confluence using Lipofectamine (Invitrogen) following manufacturer's protocol, and harvested 48 hours post-transfection. Cells from one 6 cm plate were lysed and mixed with 10 µL α-HA-beads (anti-HA affinity matrix beads (Roche)) following the purification in the manufacturer's protocol. Western analysis was performed with affinity-purified antibodies to specific proteins. An aliquot of 5% input material was taken to control for expression in all reactions.

3.3.6.6 Localization studies in zebrafish

The reporter construct pSP64T-mGFP-Vasa-3`UTR (Kedde *et al.* 2007) was kindly provided by the laboratory of Dr Erez Raz. We cloned the full-length *Dr*TDRD1 sequence after the mGFP sequence. We removed or replaced the 3`UTR of Vasa by the 3` UTR of *Dr*TDRD1.

The constructs were used to synthesis capped sense RNA using the mMessageMachine kit from Ambion. 210pg/embryo RNAs were injected into one-cell stage *D. rerio* embryos.

The primordial germ cells were imaged 24 hpf with a Zeiss Axiovision Z1 wide-field microscope. Images were taken with Axiocam camera, processed with manufacturer's software. To analyze potential miRNA binding sides MicroCosm Targets V5 was used.

3.3.6.7 Knockdowns of *Dr*TDRD1 and *Dr*MOV10L

0.1 mM dilutions of Morpholinos were injected into one-cell stage of *D. rerio* embryos.

Morpholinos: MO1: 5'-GTATATCATTCATACCATCCAGAAC-3'
MO2: 5'-AATGCAGCAGCATATCTTACCCTGG-3'
MO3: 5'-ACACTGCACCACACTCACCATCCTC-3'
MOC: 5'-CCTCTTACCTCAGTTACAATTTATA-3'

The analysis of the Morpholino injected embryos were done after 24 h. MO1 and MO2 were controlled by RT-PCR.

The qPCR was performed using specific primers for the transposable elements, Vasa (Houwing *et al.* 2008) and Actin (see above) and the Brillian®II SYBER®Green QPCR Master Mix kit from (Stratagene) in the Stratagene Mx3005P QPCR System (Program: 10 min 95 °C; 40 cycles: 30 s 95°C, 30 s 51°C, 30 s 72°C; 1 min 95°C; 30 s 55°C; 30 s 95°C)

3.4 Additional experiments

3.4.1 Development of a reporter assay for functional studies of piRNAs

3.4.1.1 Aim

Piwi-interacting RNAs are the largest class of small RNA molecules expressed in animals, specifically in the gonads. Together with the Piwi protein they form complexes linked to transcriptional gene silencing of transposable elements in the animal germline (Lau 2010) (see sections 1.3.5-1.3.7). Most of the factors involved in the piRNA field and mechanistic studies were performed in flies due to the possibility to manipulate the genome easily. Biochemical analyses of the mouse homologs of these fly proteins and the discovery of new factors in this system has led research to a basic understanding of the piRNAs. However, a fast tool to identify and manipulate piRNA pathway components is still missing.

D. rerio is relatively easy and fast to manipulate (see section 3.3.3). The accessibility of the germline and the embryonic development of the fish in a Petri dish provide the possibility to use *D. rerio* embryos to develop a screening method involving fluorescence reporter for a fast functional studies of piRNA pathway components. Therefore, a suitable construct to be injected into one-cell stage embryos, which would then expresses a fluorescence protein only in the germline was needed. In 2007 Kedde *et al.* reported that the 3`UTRs of germline-specific proteins are responsible for their specific localized expression due to the inaccessibility of a miRNA binding site, which is blocked by the germline-specific expressed Dnd1 protein (Kedde *et al.* 2007) (see section 3.3.3.2). The design of a fluorescence based reporter assay using *D. rerio* embryos was based on a construct encoding GFP followed by the 3`UTR of Vasa, which was kindly provided by Kedde and colleagues (see Figure 30 A).

Based on the assumption that piRNAs (like miRNAs and siRNAs) guide the piRNA silencing complex to a target mRNA and silence the translation either by slicing or a translational block (Hutvagner & Zamore 2002), the fluorescent signal of a construct containing piRNA target sites, should decrease. In case that the piRNA pathway is disturbed by knockdowns of piRNA pathway components, the signal intensity should increase.

3.4.1.2 Results

3.4.1.2.1 Design and analysis of fluorescence piRNA reporters

The construct encoding GFP followed by the 3`UTR of Vasa (Kedde *et al.* 2007) was modified to serve the purpose of a fluorescence reporter for piRNA activity. Therefore, a cloning site (linker) was inserted between the stop codon of the GFP and the 3`UTR of Vasa to be able to insert different piRNA target sites (see Figure 30 B, first construct). In this linker construct perfect complementary sequences to piRNAs which are known to be abundant were cloned (piRNA30, piRNA31 and piRNA80). It is still unknown if piRNAs bind in perfect complementarity to their target mRNAs and silence them by cleavage like siRNAs, or if they bind only partial complementary and inhibit the translation like most of the animal miRNAs (Jackson & Standart 2007). To answer this question, mismatched (bulges) binding sites of piRNAs were inserted into the linker construct as well. miRNAs are shown to perform translational silencing better on target mRNAs which contain more than one miRNA binding site (Doench *et al.* 2003). Therefore bulge constructs containing three piRNA target sites were created (see Figure 30 B, construct 5 and 6). To be able to quantify the fluorescence intensity of the different constructs and to compare them, independent of the actual amount of injected RNAs (because that is technically difficult to control accurately), a normalization factor was needed. A second construct encoding another fluorescent protein, Cherry, followed by the 3`UTR of Nanos without piRNA binding site was designed (see Figure 30 B, construct 7). This construct co-injected with the GFP-reporter enables the experimentator to normalize the obtained values for the intensity of the GFP signal against the intensity of the Cherry signal to quantitatively compare different biological samples injected simultaneously with both constructs. To test if the constructs are expressed, equal amounts of RNA of the GFP-linker-3`UTR Vasa construct and the normalization construct Cherry-3`UTR Nanos were co-injected into one-cell stage embryos. 24 hpf, using fluorescence microscopy, the fluorescence of both constructs was obtained from primordial germ cells of dechorinated embryos. This confirmed the result of Kedde *et al.* showing that the 3`UTRs lead the expression of the fluorescence marker into the germ cells (see Figure 30 C). To analyze quantitatively changes in the GFP expression, the microscopy settings, once

optimized, were unchanged for each sample throughout the experiment and the images were analyzed with the image processing program ImageJ from the EMBL (see Figure 30).

One of the difficulties in analyzing the images is the unbiased determination of the background (noise). Therefore, selecting only the signal of the region of interest on the image was necessary by applying the automated ImageJ threshold in one of the images (the signals in the Cherry image were easier to determine, so it was used for this procedure, see Figure 30 D step 1). Thereby, a grayscale Cherry image was converted to a binary image by defining a grayscale cutoff point (values below the cutoff become black and those above become red (see Figure 30 D step 2). From this image a mask (see Figure 30 D step 3) was created, given the value 1 to the red (over cutoff) pixels of the image and 0 to the black pixels (below cutoff). The mask was then inverted (see Figure 30 D step 4); so that the signals below the cutoff take the value 1, while the signals above the cutoff are 0 (this step is necessary, because the ImageJ program is not able to create an inverted mask directly). The original GFP and Cherry images were multiplied with the inverted mask (pixel by pixel), giving images of the noise without the signals (see Figure 30 D step 5). The average noise intensity of each image was calculated with ImageJ. Then this calculated value of the background noise was subtracted from the original GFP and Cherry images (see Figure 30 D step6). To be able to compare the signal intensity of different cells, embryos and fluorescent signal of the different constructs, the GFP signals had to be normalized by the Cherry signals. The ImageJ automated threshold was determined for both images with background subtracted to define the region of interest (cells where the fluorescence signals are expressed) (see Figure 30 D step 7). Then the signal intensity for the GFP was divided by the Cherry signal (pixel by pixel). In this normalized image (see Figure 30 D step8) the cells were determine by eye. Therefore the cell shape was analyzed in the original images and the cell edges transferred to the normalized image. To analyze in an unbiased manner, the entire cell was taken into account, although some regions in the cell, for example where the nucleus is located have a weaker signal then just cytoplasmic regions. Given that these regions are comparable in each cell, they should not lead to an increase in the standard deviation. The mean of GFP/Cherry intensity was determined for each cell by ImageJ, excluding background signals and the intracellular space. Values were blotted in an excel file.

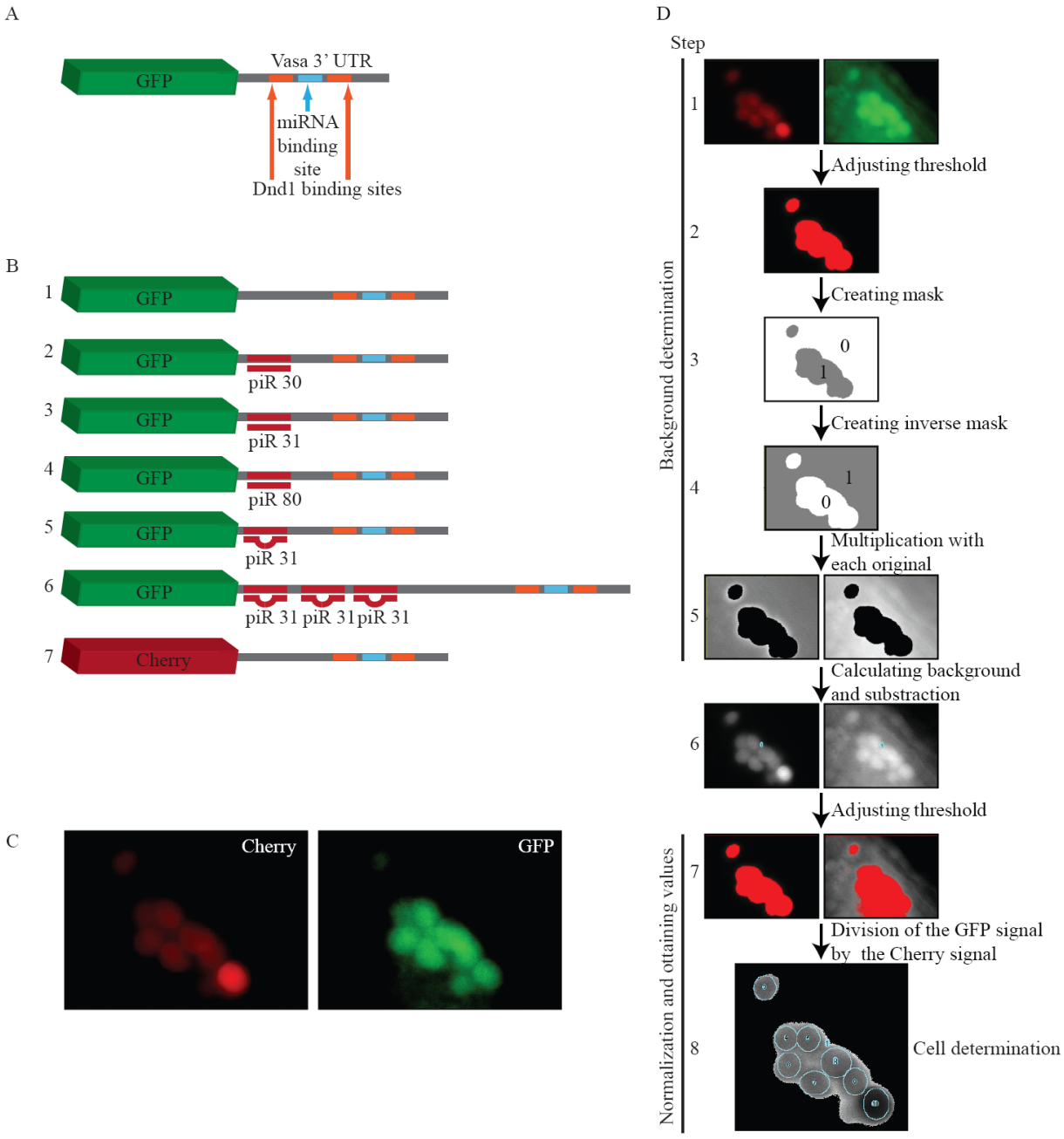


Figure 30: Reporter constructs and image processing

(A) A schematic view of the reporter construct as described by Kedde *et al.*. The sequence encoding GFP is followed by the 3`UTR of Vasa, containing a miRNA-binding site (blue), flanked by two Dnd1 binding sites (orange).

(B) The construct of Kedde was used and a linker region was inserted behind the GFP stop codon in the 3`UTR of Vasa (construct1). Different piRNA binding sites with perfect sequence recognition (piR30: construct 2, piR31: construct 3, piR80: construct 4) or bulged forms (piR30 1x bulge: construct 5, piR31 3x bulge: construct 6) are cloned into the construct. As a control and normalization construct we used a construct encoding Cherry (red construct) followed by the 3`UTR of Nanos. In this construct no piRNA recognition site was inserted.

(C) Expression of reporter constructs. Cherry-Nanos construct and the GFP-reporter co-localize in the primordial germ cells.

(D) The image processing pathway of the reporter assay experiments.

3.4.1.2.2 Application of the fluorescence reporter assay

To determine if endogenous piRNAs down regulate the fluorescence signal of the described reporter constructs, containing perfect piRNA binding sites (piR30,31, 80) or bulge sites (1x piR31 bulge, 3x piR31 bulge) and also which of the GFP-reporters is most efficient, the fluorescence reporter assay was performed. Therefore, one-cell stage embryos were injected with equal amounts of *in vitro* transcribed RNAs encoding for the GFP-reporter constructs and the Cherry-normalization construct. 24 hpf, the dechorinated embryos were analyzed using fluorescence microscopy. The microscopy settings were unchanged for each sample throughout the experiment. The obtained images were analyzed as described in section 3.4.1.2.1. The mean of the GFP/Cherry signal of each cell was transferred into an Excel sheet and combined into a column diagram (see), whereby the GFP/Cherry value of the GFP-linker construct without piRNA binding site was set to 1. The error bars represent the standard deviation.

Although a trend was observed, indicating that the GFP-3`UTR Vasa piR31 construct was the least expressed. The standard deviations are too high. Thus the trend cannot be confirmed to be statistically significant without further testing.

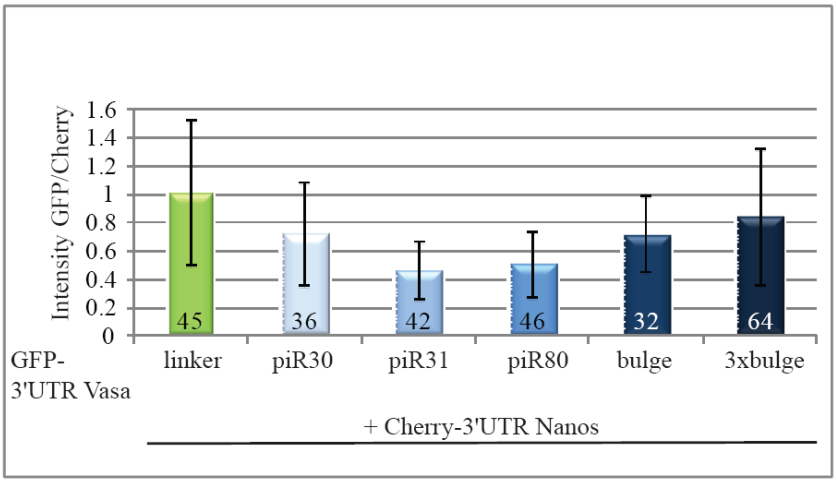


Figure 31: Analysis of the reporter constructs
The columns represent the Intensity of the GFP signal divided by the intensity of the Cherry signal. The standard deviation is indicated by bars, and the number of samples (n) analyzed indicated by the numbers in the columns.

3.4.1.3 Discussion

In this study *D. rerio* embryos were used as a system to develop a screening method based on fluorescence reporter for a fast functional insight into the piRNA pathway. Therefore, constructs were created encoding a fluorescence protein followed by the 3`UTR of Vasa or Nanos, containing piRNA binding sites and the RNAs of these constructs were injected into one-cell stage embryos. In all cases fluorescence in primordial germ cells was observed, indicating that the reporter constructs with and without piRNA binding site are tissue-specific expressed, as expected. It is known that the 3`UTR of Vasa or Nanos guide the expression of the proteins to the germ cells (Kedde *et al.* 2007). Adding the piRNA binding sites into the constructs did not disrupt this feature of the 3`UTR.

The intensity of the fluorescence proteins from constructs with piRNA binding sites, or empty cloning site was measured and the values normalized against the intensity of co-injected constructs containing another fluorescence protein, without the piRNA binding sites. Therefore, the noise in the image was determined from the whole image and subtracted. The normalization was performed by dividing the fluorescent intensity of the construct of interest (GFP) by the signal of co-injected normalization construct without piRNA binding site (Cherry). By this method the biased definition of the background was minimized and the normalized values are comparable, independent of the actual amount of injected RNAs. In default of a program defining the cells in the image, they had to be distinct by eye. Although, in most cases the definition what is cell and what is background is rather easy, it is admitted that an automatic recognition would reduce biased observations by the human eye.

In the design of this assay it was assumed that the RNA translation into a fluorescent protein of constructs containing piRNA binding sites would be down regulated by the endogenous piRNAs in the primordial germ cells of the zebrafish embryo. Unfortunately, constructs containing piRNA binding sites (with perfect complementarity or one or 3 bulge sites) did not down regulate the fluorescence signal with statistical significance and the calculated standard deviation is too high to even consider the values of the performed piRNA reporter assay. The variability of the values is possibly caused by the fluorescence background due to auto-fluorescence of the cells and the leaking expression of the GFP construct (outside of primordial germ cells). Second, the depth of field of the confocal plane

using multi-color TIRF microscope is small giving well defined images. Unfortunately, the fluorescence signals measured are from the entire sample thickness, which gives rise to variations in the signals between samples. A possible solution could be the use of confocal microscopy, although this technique is more time consuming and inhibits fast high through-put measurements.

However, a preliminary impression of the observed fluorescence intensity changes due to piRNA binding indicated that binding sites for the recognition by the zebrafish piRNAs 31 and 80 represses the expression of the fluorescence protein more than piR30 or bulge sites. The results using the assay show a high variability and require confocal microscopy to improve the statistical significance for identification of piRNA pathway components.

3.4.1.4 Method

The reporter constructs pSP64T-mGFP-Vasa-3`UTR and pSP64T-dsRedEX-nos1-3`UTR (Kedde *et al.* 2007) were kindly provided by the laboratory of Erez Raz. The pSP64T-mGFP-Vasa-3`UTR was equipped with perfect complementary binding sites for piR-30, piR-31, or piR-80. Additionally, constructs with one bulge binding site or three bulge binding sites for piR-31 were created. The constructs were used to synthesize capped sense RNA using the mMessageMachine kit from Ambion. 210pg/embryo RNA were injected into one-cell stage *D. rerio* embryos.

As a normalization factor the same amount of RNA (210pg/embryo) of a Cherry-Nanos construct was always mixed with the RNAs reporter constructs in the master mixture to be co-injected. The primordial germ cells were imaged with an Olympus Biosystems Cell^R multi-color TIRF microscope with a 40x water objective. Images were analyzed using ImageJ (EMBL).

3.4.2 Identification of possible interaction partners of the MYND domain of TDRD1 using a yeast two hybrid

3.4.2.1 Aim

TDRD1 contains four Tudor domains at its C-terminus and a N-terminal myeloid–nervy–DEAF-1 (MYND) domain (Chuma *et al.* 2003). It was shown that the Tudor domains recognize symmetrically dimethylated arginines in the N-termini of the murine Piwi proteins and the mouse Vasa homolog (MVH) (Kirino *et al.* 2009; Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a) (see section 1.4.3), while potential interaction partners of the MYND domain, although MYND domains are known to be a potential protein interaction domain (Gross & McGinnis 1996; Lutterbach *et al.* 1998), are unknown.

The structure of the MYND domain was solved from the human ETO protein (Myeloid translocation protein 8), showing homologies to proteins of the PHD and RING finger families. The MYND domain consists of an α-Helix and two β-sheets and contains a highly conserved cysteine repeat, which form two zinc fingers (Liu *et al.* 2007) (see Figure 32). To identify potential interaction partners of the murine TDRD1 MYND domain, yeast two hybrid experiments against a murine testis library were performed.

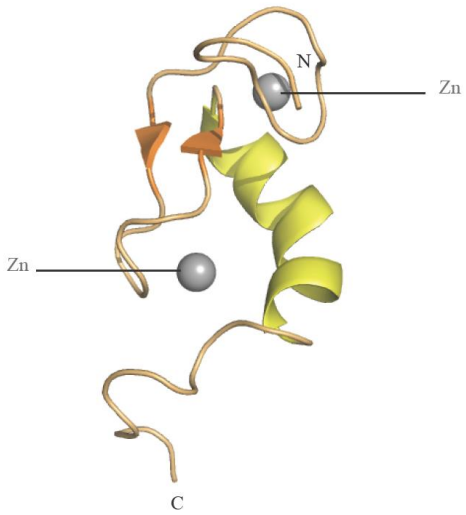
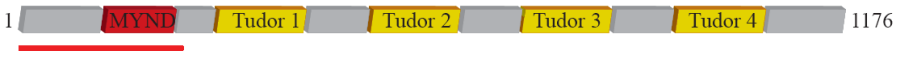


Figure 32: Structure of the ETO MYND domain
Ribbon display of the structure of the ETO MYND domain. The bound Zinc are indicated as spheres. Structure adapted from (Liu *et al.* 2007) using PyMOL

3.4.2.2 Results

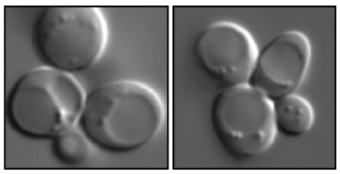
To determine the interacting proteins, the N-terminus of TDRD1 (see Figure 33 A) was cloned as bait for performing a yeast two-hybrid screen against a pre-transformed mouse testes library. After exclusion of auto activity and toxicity of the bait, a yeast mating was performed over 24h with a mating efficiency of 11.73 % (see Figure 33 B). 435 colonies grown on selective media were selected positive as containing potential interactors for the MYND domain. Of these 76 colonies indicate an interaction by turning blue on X-Gal containing medium. Clearly dark blue colonies were restreaked and selected via a 3-AT-test. The prey-vectors from the positive tested 32 remaining colonies were rescued and send for sequencing. Additionally, 350 colonies over Ø 2 mm were observed. These colonies were used for 3AT tests as well and 72 colonies were tested positive as potential interactors. After rescuing the pray-vectors, the inserts were obtained by digests which showed that except of 3 clones all represented a similar pattern. The 3 clones which showed differences were sent for sequencing as well. All together 29 clones were found containing the FK506 binding protein 8 (Fkbp8), 2 clones of RAN binding protein 9 (Ranbp9) and one clone which contained the epidermal Langerhans cell protein (LCP) 1.

A



1-240 aa
bait

B



C

TDRD1 interacting library clones	Found partners
29 clones	*Mus musculus* FK506 binding protein 8 (FKBP8)
2 clones (two different preys)	*Mus musculus* RAN binding protein 9 (RanBPM)
1 clone	*Mus musculus* Epidermal Langerhans cell protein 1 (LCP1)

Figure 33: Yeast-two-hybrid

(A) Schematic view of TDRD1. The bait was constructed containing the 1-240 aa of TDRD1 including the MYND domain (red).

(B) Mating yeast at 100 x. The mating was performed with an efficiency of 11.73 %.

(C) Obtained results of the yeast two hybrid experiment using the N-terminus of TDRD1 against a pre-transformed mouse testes library. 29 clones containing the FK506 binding protein 8 (Fkbp8), 2 clones of Ran binding protein 9 (RanBPM) and one clone which contained the epidermal Langerhans cell protein (LCP) 1 as potential interaction partners of the MYND domain of TDRD1

3.4.2.3 Discussion

Potential interaction partners of the MYND domain were analyzed in a two hybrid screen using the TDRD1 N-term as bait and a pre-transformed testis library. The results indicated that the MYND domain potentially interacts with FK506 binding protein 8 (Fkbp8), RanBPM (RAN binding protein 9, Ranbp9) and/or the epidermal Langerhans cell protein (LCP) 1.

A direct functional relevant interaction of the TDRD1 MYND domain with LCP1 is unlikely given that LCP1 is expressed in the nucleus of mainly epidermal Langerhans cells, while TDRD1 is a cytoplasmic protein in germ cells. Furthermore, the piRNA pathway is restricted to germline as well and a functional connection of LCP1 is unlikely.

RanBPM was initially described to bind to the human RAN, a small GTP binding protein belonging to the RAS superfamily that is essential for the translocation of RNA and proteins through the nuclear pore complex and microtubule assembly (Nakamura *et al.* 1998). RanBPM localizes in the cytoplasm as well as in the nucleus surrounding the centromers, where it is required for correct nucleation of microtubules (Nakamura *et al.* 1998; Nishitani *et al.* 2001). In mouse RanBPM is ubiquitously expressed with a high expression in testes from late spermatocytes to round spermatids with a peak at the pachytene stage (Shibata *et al.* 2004), overlaying with the expression profile of TDRD1. RanBPM is mainly localized in the cytoplasm and a component of the chromatoid body in round spermatids (Shibata *et al.* 2004) like TDRD1. In agreement to the similar expression patterns and the localizations, it was found that RanBPM interacts with the N-terminal of MVH (Fujiwara *et al.* 1994; Shibata *et al.* 2004). The role of RanBPM/MVH interaction during spermatogenesis could be the transfer of MVH proteins to an anchoring complex during the chromatoid body arrangement, mediating microtubule assembly. Another possibility is that the complex associates with RanGTP and could be involved in catching and transferring RNAs from the nucleus into the chromatoid body (Shibata *et al.* 2004). Another indication of a possible involvement of RanBPM in the piRNA pathway is that it was found to interact with MAEL, another component of the chromatoid body, which interacts with mi- and piRNA pathway components (Costa *et al.* 2006). Further, it interacts with the chromatoid body member GASZ. *Gasz* knockout mice are male sterile due to a developmental block after the zygotene

stage. These mutants show retrotransposon activation due to a decrease of the piRNA amount (Ma *et al.* 2009).

Taken the similarities of the expression, localization and known interaction partners of RanBPM to TDRD1 into account, it is possible that the MYND domain interacts directly with RanBPM, maybe to guide the localization of TDRD1 into the chromatoid body. It would be interesting to confirm the association of RanBPM and TDRD1 endogenously and to analyze the function of this complex in the future.

The search for interaction partners of the MYND domain of TDRD1 revealed FKBP8 as a potential candidate. FKBP8 is involved in the regulation of apoptosis in neuron cells but also in other cell types (Bulgakov *et al.* 2004; Okamoto *et al.* 2006; Rosner *et al.* 2003; Wang *et al.* 2006). It is a member of the FK506-binding, immunophilin protein family. It contains a peptidyl-prolyl cis-trans isomerase domain that may be involved in protein folding or as a scaffold to facilitate protein-protein interactions. Additionally, it contains a C-terminal tripartite tetratricopeptide repeat protein-protein interaction (TPR) domain, calcium/calmodulin-binding motif and a transmembrane motif (TM) (Pedersen *et al.* 1999; Shirane & Nakayama 2003). FKPB8 is mainly expressed in prefrontal cortex and early erythroids. It is also expressed in testis but not at a significantly level. Given the expression pattern of Fkbp8, it is probably not a direct interaction partner of the TDRD1 MYND domain and not of high interest concerning the piRNA pathway studies. On the other hand, members of the Fkbp class (FKBP6, 5 and 3) were recently uncovered as components of the MIWI, MILI and MIWI2 complexes (Vagin *et al.* 2009). FKBP3 and FKBP5 are cis-trans prolyl isomerases that bind to the immunosuppressants FK506 and rapamycin (Jin *et al.* 1992; Nair *et al.* 1997). FKBP3 is mainly expressed in brain and immune system cell. It has a higher affinity for rapamycin than for FK506 and is maybe an important target molecule for immunosuppression by rapamycin (Jin *et al.* 1992). Interestingly, FKBP3 interacts with the two Histone deacetylases HDAC1, HDAC2, the HDAC-binding transcriptional regulator YY1(Yang *et al.* 2001) and the E3 ubiquitin Ligase, mouse double minute 2 (MDM2) (Ochocka *et al.* 2009), indicating a role in chromatin remodeling and gene expression regulation. Given FKBP3s weak expression in the germline, it is unlikely but not excluded to be an interaction partner of the TDRD1 MYND domain and component of the piRNA pathway. FKPB5 is involved in post-traumatic stress disorder, depression and anxiety (Gillespie *et al.* 2009). It is expressed in a wide range of tissues but mainly in the skeletal

muscles and adipocytes (Nair *et al.* 1997). FKPB5 is as well not a primary candidate for an association with TDRD1.

An interesting interaction partner for the MYND domain of TDRD1 would be prolyl isomerase and TPR domain-containing protein FKBP6 (Meng *et al.* 1998; Pedersen *et al.* 1999; Shirane & Nakayama 2003). The protein is expressed in male and females germlines, but the loss of FKBP6 causes an arrest in the spermatogenesis during the meiotic prophase in mice and rats, leading to male sterility. The female fertility is not affected (Crackower *et al.* 2003; Noguchi *et al.* 1993). In spermatocytes FKBP6 localizes in the cytoplasm and in the nucleus. In the pachytene stage the protein was found to be in the synaptonemal complex. In *fkbp6* knockouts the autosomes and sex-chromosomes mis-align or show non-homologous paring in the pachytene stage (Crackower *et al.* 2003; Noguchi *et al.* 2008). It would be interesting to investigate, if the piRNA pathway is influencing the alignment of chromosomes by regulating and guiding FKBP6.

In conclusion, the identified interaction partners of the TDRD1 MYND domain may be directly involved or are members of protein families, which could be associated to the piRNA pathway (except LCP1). Based on the yeast two hybrid results and the literature search, it would be interesting to confirm direct interactions and to perform functional studies of the potential complexes to gain deeper insights into the piRNA mechanism.

3.4.2.4 Method

The MYND domain of mTDRD1 (1-240 aa) was cloned as a bait into pGBKT7. After auto-activation and toxicity tests, the bait was used with a Matchmaker™ mTestis in Yeast library (638872) following the manufacturer's protocol. Used yeast strains were AH109 (bait) and Y187 (library). Preliminary positive tested colonies were restreaked and used for a 3-AT-test according to the manufacturer's protocol, to select strong interactions. Positive clones were digested and unique digests sizes were sent for sequencing.

Chapter 4: General conclusions and future perspectives

4.1 General conclusions and future perspectives

Metazoan germlines express germline-specific non-coding RNAs, the so called piRNAs. These RNAs associate with Piwi proteins and ensure maintenance of germline stability by silencing transposable elements (Aravin *et al.* 2003; Brennecke *et al.* 2007; Gunawardane *et al.* 2007; Houwing *et al.* 2007; Saito *et al.* 2006; Vagin *et al.* 2006). piRNA biogenesis and mechanisms of their action are poorly understood. piRNAs originate from a single-stranded precursor transcribed from genomic piRNA clusters (Aravin *et al.* 2006; Lau *et al.* 2006; Watanabe *et al.* 2006) and the piRNA biogenesis was found to be Dicer-independent in flies and fish (Houwing *et al.* 2007; Vagin *et al.* 2006). Genetic screens in *Drosophila* and biochemical purifications of macromolecular complexes in mice are just beginning to uncover factors involved in the pathway. This thesis focused on the analysis of three such factors: the methyltransferase mHEN1 in the mouse system and the Tudor domain-containing protein *Dr*TDRD1 and the putative helicase *Dr*MOV10L in the *Danio rerio* model system.

All small RNAs in plants are methylated at the 3` terminal nucleotide. Some miRNAs and siRNAs in *Drosophila* carry similar modifications. piRNAs in all animals analyzed to date are modified by a 2`-O-methyl mark. This modification protects the small RNA 3` termini from uridylation, which usually acts as a signal for destabilization (Ameres *et al.* 2010; Chen 2005; Ebhardt *et al.* 2005; Horwich *et al.* 2007; Kamminga *et al.* 2010; Li *et al.* 2005; Saito *et al.* 2007a; Yang *et al.* 2006b). The methylation mark on the small RNAs is catalyzed by HEN1 and its homologs in different species (Akbergenov *et al.* 2006; Horwich *et al.* 2007; Kamminga *et al.* 2010; Katiyar-Agarwal *et al.* 2007; Kirino & Mourelatos 2007c; Li *et al.* 2005; Ohara *et al.* 2007a; Saito *et al.* 2007a; Yang *et al.* 2006a; Yang *et al.* 2006b; Yu *et al.* 2005; Zhou *et al.* 2010). Although it was shown that the *Dm*HEN1 is able to interact with the *Drosophila* Piwi proteins (Saito *et al.* 2007a), it was not clear how HEN1 is recruited to the piRNA pathway. Despite the fact that the mechanism was not understood, in mice only the recombinant mHEN1 was shown to methylate RNAs *in vitro* (Kirino & Mourelatos 2007a, b). Therefore, in this thesis the mechanism by which the enzyme is recruited to participate in the mouse piRNA pathway was investigated. The results demonstrate that the endogenous mHEN1 in mouse testes extracts is an active methyltransferase capable of methylating small RNAs in the size range of piRNAs. In addition, it shows that mHEN1

interacts with one of the murine Piwi proteins, MILI. These findings link the enzyme directly to the piRNA pathway.

Mapping studies indicated that the N-terminal region of MILI interacts with the C-terminal region of mHEN1. This region on MILI was mapped close to the PAZ domain, which binds the methylated 3` end of piRNAs. It was shown that the N-terminus of MILI is required, but not sufficient to mediate interaction with mHEN1. Therefore, it is believed that a composite interaction site formed from the N-terminus and PAZ facilitates this interaction. This suggests that the 3` RNA binding site of MILI, the PAZ domain, and the catalytic domain of mHEN1 are in close proximity to facilitate the methylation of the piRNAs and its transfer to the PAZ domain. Furthermore, it was shown here that this interaction occurs in an RNA-independent manner and is not influenced by the methylation status of the arginines of MILI, which are recognized by Tudor domain-containing proteins. This suggests that MILI need not to be loaded to interact with mHEN1 and that this interaction can occur early in piRNA biogenesis, before MILI is post-translationally modified. Further investigations towards the structure of the MILI/mHEN1 complex together with bound RNAs could lead to a definition of the binding surface and could address the specific determinant of mHEN1 for Piwi proteins. Dynamic studies using NMR techniques could reveal how the recognition of the methylated 3` end by the PAZ domain alters this interaction surface, as mHEN1 is never found in complex with MILI loaded with piRNAs. This could be a mechanism for separation of the enzyme from the product after the methylation reaction.

In mice, spermatogenesis proceeds through different stages of mitotic and meiotic divisions. A strong MILI expression is observed in pachytene stage spermatocytes during meiosis (Aravin *et al.* 2008; Kuramochi-Miyagawa *et al.* 2001). The round spermatids are characterized by a dense peri-nuclear structure called the chromatoid body that was shown to contain MILI (Aravin *et al.* 2008; Wang *et al.* 2009a). mHEN1 was localized to a separate perinuclear granule (HEN-body), suggesting either that mHEN1 leaves the MILI-containing piRNA complex after performing the methylation of the RNA, or that the piRNA formation (Piwi slicer activity) and methylation by mHEN1 are spatially separated in the chromatoid and HEN bodies, respectively. Recombinant mHEN1 is shown to methylate small RNAs both in the size range of miRNAs and piRNAs (Kirino & Mourelatos 2007b). So it is believed that the distinct localization of mHEN1 might be a mechanism to spatially restrain a potentially promiscuous enzyme that might methylate miRNAs and other mRNAs present in the CBs.

Future research should reveal the importance of this unique localization of mHEN1 and the identity of other mHEN1-interacting factors that dynamically links it to the piRISC. Future research should reveal the exact site of the small RNA methylation; either at the site of piRNA biogenesis (which is currently unclear) or in the HEN bodies where Piwi-bound RNAs transiently reside.

In contrast to mammalian Ago proteins which bind miRNAs and siRNAs, the fact that piRNAs carry a modified 3` end requires the Piwi proteins to accommodate the methyl group in their PAZ domain. To further understand this recognition process, biophysical studies were performed in the laboratory which demonstrated that the MIWI PAZ domain affinity for ssRNAs which is thirty times higher than for double stranded RNAs, consistent with the absence of dsRNA intermediates in the piRNA pathway. However, MIWI PAZ showed only a 2-fold higher affinity for a methylated 3` end compared to an unmethylated RNA. It is possible that further interactions in the context of the full-length protein (that can accommodate longer small RNAs) might facilitate a stronger interaction with the methylated piRNAs. To obtain molecular insight of this interaction, in collaboration with the group Dr. Teresa Carlomagno (EMBL Heidelberg), the NMR structure of the MIWI PAZ domain in complex with a piRNA mimic was solved (Simon *et al.* 2011). The solution structure reveals that the methyl-group is inserted deeply into a pre-formed hydrophobic pocket of conserved residues. The binding pocket contains a hydrophobic cleft consisting of F333, A381 and M382. In the complex, the long hydrophobic side-chain of the M382 at the back of the cavity makes contact with the methyl group. Mutation of this residue results in proportional reduction in the affinity for methylated 3' ends. However, the methionine at this position is not conserved across phyla, indicating that this contact is not necessary for the methylation mark recognition. What is interesting in this structure is the fact that although the MIWI PAZ structure resembles very much those of Ago proteins, subtle differences in the amino acid composition of the pocket allows the recognition of the extra methyl group on the 3` end of piRNAs. These differences were not evident in simple sequence alignments. Similar results were obtained for the human HILI protein as well (Tian *et al.* 2010). It would be interesting to identify the binding capacity and the structures of PAZ domains without the conserved methionine and measure the affinity to single- and double-stranded, methylated and non-methylated RNAs.

Coming back to the task of describing piRNA pathway components, the restrictions using mouse as a model system had to be overcome by establishing a new, more flexible model system using *Danio rerio*. Two piRNA pathway factors in mice are the putative helicase MOV10L, which is thought to be involved in the stabilization or loading of piRNAs into Piwi proteins (Frost *et al.* 2010; Lovasco *et al.* 2010; Wang *et al.* 2001b; Zheng *et al.* 2010) and the Tudor-domain-containing protein TDRD1, which is an essential component of the piRNA silencing complex (Kojima *et al.* 2009; Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a). Here, this study was aimed to identify and functionally characterize the mechanism of the two proteins in the zebrafish system. Once the proteins were identified in the fish, the expression pattern of these proteins revealed that the mRNAs of *Drtdrd1* and *Drmov10l* are expressed from the egg stage throughout development of the zebrafish similar to zebrafish Piwi proteins, suggesting the mRNAs for all these proteins are maternally contributed. Using an affinity purified antibody against *Dr*TDRD1, it was shown that the protein is expressed in testes. This result is in agreement with the expression pattern of the medaka TDRD1 (Aoki *et al.* 2008). The expression and localization of *Dr*TDRD1 in this specific tissue is similar to other germ cell-specific proteins (Kedde *et al.* 2007) and dependent on its 3`UTR. Future research could localize *Dr*TDRD1 as well as *Dr*MOV10L in the PGCs with high resolution to determine the sub-cellular targeting.

In contrast to the mouse TDRD1 protein, no bound piRNAs were detected in immunoprecipitations of *Dr*TDRD1. Either, the amount of piRNAs was insufficient for detection, or *Dr*TDRD1 binds to unloaded Piwi proteins, or it does not bind Piwi proteins at all. The mouse TDRD1 is shown to detect symmetrically dimethylated arginines in the N-termini of at least MILI and MIWI (it is not clear for MIWI2) (Kojima *et al.* 2009; Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a). Unfortunately, an interaction of the zebrafish Piwis and *Dr*TDRD1 was not observed in endogenous tissue extracts or in mammalian cell culture experiments. Immunoprecipitations using better antibodies will answer the question of the interaction in the future.

To examine the function of *Dr*MOV10L, knockdown experiments were performed using morpholino against the translational start site of *Dr*MOV10L, which were injected into the one-cell stage embryo. Like in mouse MOV10L (Zheng *et al.* 2010) and zebrafish *zili* knockouts (Houwing *et al.* 2008), the activity of transposable elements increased as a consequence of morpholino injection. The similarity to the other phenotypes indicates that

*Dr*MOV10L is directly involved in the piRNA pathway and required for silencing of retrotransposable elements. The knockdown method via morpholinos could be a powerful tool to screen for piRNA biogenesis components which influence the activity of transposable elements. Future experiments involving complementation of the knockdown phenotype with either full-length or mutagenised (deletion, point-mutant) versions of the protein should reveal the functional role of the different domains and its putative helicase activity.

Overall, my thesis work has identified a biochemical basis for recruitment of the piRNA methyltransferase to the mouse piRNA pathway. This work is now being expanded into a structural analysis of the interaction that was described here. Together, these studies will be of great interest to the field. Regarding the development of the zebrafish embryos as a model system to study piRNA factors in the lab, the tools necessary for addressing these questions were set up and it was demonstrated their utility by studying two factors. Morpholino knock-downs followed by detection of endogenous transposons could be used as a rapid functional assay to validate candidate vertebrate piRNA factors identified in biochemical purifications elsewhere.

4.2 Conclusions générales et perspectives

Les lignées germinales des Métazoaires expriment des ARN non codant spécifiques aux lignées germinales appelés piARN. Ces ARN s'associent à des protéines Piwi et assurent le maintien de la stabilité des cellules germinales en inhibant les éléments transposables (Aravin *et al.* 2003; Brennecke *et al.* 2007; Gunawardane *et al.* 2007; Houwing *et al.* 2007; Saito *et al.* 2006; Vagin *et al.* 2006). La biogenèse des piARN et leur mécanisme d'action sont mal connus. Les piARN proviennent d'un précurseur ARN simple brin transcrit à partir de clusters génomiques de piARN (Aravin *et al.* 2006; Lau *et al.* 2006; Watanabe *et al.* 2006). Leur biogenèse est indépendante de Dicer chez la drosophile et les poissons (Houwing *et al.* 2007; Vagin *et al.* 2006). Des criblages génétiques chez la drosophile et des purifications biochimiques des complexes macromoléculaires chez la souris commencent à peine à identifier les facteurs impliqués dans cette voie. Cette thèse s'est concentrée sur l'analyse de trois facteurs de ce type: la méthyltransférase mHEN1 dans le système de la souris et les deux protéines comportant des domaines Tudor, *Dr*TDRD1 et l'hélicase putative *Dr*MOV10L dans le système *Danio rerio*.

Tous les petits ARN des plantes sont méthylés au niveau du nucléotide 3' terminal. Certains miRNA et siRNA chez la drosophile portent des modifications similaires. Les piARN de tous les animaux analysés à ce jour sont modifiées par une marque 2'-O-méthyl. Cette modification protège l'extrémité 3` des petit ARN de l'uridylation qui agit habituellement comme un signal de déstabilisation (Ameres *et al.* 2010; Chen 2005; Ebhardt *et al.* 2005; Horwich *et al.* 2007; Kamminga *et al.* 2010; Li *et al.* 2005; Saito *et al.* 2007a; Yang *et al.* 2006b). La marque de méthylation sur les petits ARN est catalysée par HEN1 et ses homologues dans différentes espèces (Akbergenov *et al.* 2006; Horwich *et al.* 2007; Kamminga *et al.* 2010; Katiyar-Agarwal *et al.* 2007; Kirino & Mourelatos 2007c; Li *et al.* 2005; Ohara *et al.* 2007a; Saito *et al.* 2007a; Yang *et al.* 2006a; Yang *et al.* 2006b; Yu *et al.* 2005; Zhou *et al.* 2010). Bien qu'il ait été démontré que *Dm*HEN1 est capable d'interagir avec les protéines Piwi de drosophile (Saito *et al.* 2007a), le recrutement de HEN1 dans la voie des piARN n'était pas clair. En dépit du fait que le mécanisme d'action n'était pas compris, chez la souris uniquement la protéine recombinante mHEN1 était connue pour méthyler les ARN *in vitro* (Kirino & Mourelatos 2007a, b). Par conséquent, dans cette thèse le mécanisme

par lequel l'enzyme est recrutée pour participer à la voie des piARN a été étudiée chez la souris. Les résultats démontrent que la protéine mHEN1 endogène dans les extraits de testicules de souris est une méthyltransférase active, capable de méthylation des petits ARN dans la gamme de taille des piARN. En outre, ceci montre que mHEN1 interagit avec l'une des protéines Piwi murin, MILI. Ces résultats lient directement l'enzyme mHEN1 à la voie des piARN.

Des études de cartographie ont indiqué que la région N-terminale de MILI interagit avec la région C-terminale de mHEN1. Cette région de MILI a été cartographiée à proximité du domaine PAZ qui lie le groupement méthyle des extrémités 3` des piARN. Il a été démontré que la partie N-terminale de MILI est requise, mais ne suffit pas à relayer l'interaction avec mHEN1. Par conséquent, il est admis qu'un site d'interaction composite formée à partir du domaine N-terminal et de PAZ facilite cette interaction. Ceci suggère que le site 3` de liaison à l'ARN de MILI, le domaine PAZ, et le domaine catalytique de mHEN1 sont localisés à proximité les uns des autres pour faciliter la méthylation des piARN et leur transfert au domaine PAZ. En outre, il a été montré ici que cette interaction ne repose pas sur la présence d'ARN et n'est pas influencée par la méthylation des résidus arginines de MILI, qui sont reconnus par les protéines contenant des domaines Tudor. Ceci suggère que MILI ne nécessite pas d'être chargée pour interagir avec mHEN1 et que cette interaction peut se produire au début de la biogenèse des piARN, avant que MILI ne soit modifiée post-traductionnellement. D'autres investigations vers la structure du complexe MILI/mHEN1 associé à l'ARN pourraient conduire à une définition de la surface d'interaction et pourraient identifier les déterminants spécifiques de mHEN1 pour les protéines Piwi. Des études dynamiques reposant sur des techniques de RMN pourraient révéler la façon dont la reconnaissance, par le domaine PAZ, de l'extrémité 3` méthylée modifie cette surface d'interaction, étant donné que mHEN1 ne se trouve jamais dans un complexe avec MILI chargé en piARN. Ce pourrait être un mécanisme permettant l'élimination du produit par l'enzyme après la réaction de méthylation.

Chez la souris, la spermatogenèse se fait par différentes étapes de divisions mitotiques et méiotiques. Une forte expression de MILI est observée dans les spermatocytes au stade pachytène de la méiose (Aravin *et al.* 2008; Kuramochi-Miyagawa *et al.* 2001). Les spermatides rondes sont caractérisées par une structure dense péri-nucléaire appelée le corps chromatoïde qui est connu pour contenir MILI (Aravin *et al.* 2008; Wang *et al.* 2009a).

mHEN1 a été localisé dans un granule périnucléaire distinct (corps HEN), ce qui suggère soit que mHEN1 quitte le complexe MILI chargé en piARN après avoir effectué la méthylation de l'ARN, ou que la formation des piARN (activité « Slicer » de Piwi) et la méthylation par mHEN1 sont séparés spatialement dans le corps chromatoïde et les corps HEN respectivement. La protéine recombinante mHEN1 est connue pour méthyler de petits ARN dans la gamme de taille des miARN et des piARN (Kirino & Mourelatos 2007b). Donc, on peut imaginer que la localisation distincte de mHEN1 pourrait être un mécanisme permettant d'éviter la promiscuité d'une enzyme qui pourrait potentiellement méthyler des miARN et autres ARNm présent dans les corps chromatoïdes. Les futures recherches devraient révéler l'importance de cette localisation unique de mHEN1 ainsi que l'identité des autres facteurs interagissant avec mHEN1, les reliant de manière dynamique au piRISC. Les futures recherches devraient révéler le site exact de la méthylation de petits ARN; soit sur le site de la biogenèse des piARN (encore peu connu) ou dans les corps HEN où les ARN liés à Piwi résident transitoirement.

Au contraire des protéines Ago de mammifères qui s'associent aux miRNA et siRNA, le fait que les piARN portent une modification en 3` nécessite des protéines Piwi qu'elles accommodent le groupement méthyle dans leur domaine PAZ. Pour mieux comprendre ce processus de reconnaissance, des études biophysiques ont été réalisées dans le laboratoire qui a démontré que l'affinité du domaine PAZ de MIWI pour les ARN simple brin qui est trente fois supérieur à celle des ARN double brin, est en accord avec l'absence d'intermédiaires d'ARN double brin dans la voie des piARN. Toutefois, le domaine PAZ de MIWI n'a montré qu'une affinité deux fois plus élevée pour une extrémité 3`méthylée par rapport à un ARN non méthylé. Il est possible que d'autres interactions dans le contexte de la protéine non raccourcie (qui peut accueillir plus de petits ARN) pourraient faciliter une interaction plus forte avec le piARN méthylé. Pour obtenir un aperçu de cette interaction moléculaire, en collaboration avec le groupe du Dr Teresa Carlomagno (EMBL Heidelberg), la structure RMN du domaine PAZ de MIWI en complexe avec un ARN mimant un piARN a été résolue (Simon *et al.* 2011). La structure en solution révèle que le groupement méthyle est profondément inséré dans une poche hydrophobe préformée comportant des résidus conservés. La poche de liaison hydrophobe contient une fente composée de F333, A381 et M382. Dans le complexe, la longue chaîne latérale hydrophobe de M382 à l'arrière de la cavité, entre en contact avec le groupement méthyle. La mutation de ce résidu résulte en une

réduction proportionnelle de l'affinité pour les extrémités 3` méthylées. Cependant, la méthionine à cette position n'est pas conservée à travers les embranchements, ce qui indique que ce contact n'est pas nécessaire pour la reconnaissance de la marque de méthylation. Ce qui est intéressant dans cette structure réside dans le fait que, bien que la structure du domaine PAZ de MIWI ressemble beaucoup à celles des protéines Ago, de subtiles différences dans la composition en acides aminés de la poche permet la reconnaissance du groupement méthyle supplémentaire présent à l'extrémité 3` des piARN. Ces différences n'étaient pas évidentes dans les simples alignements de séquences. Des résultats similaires ont aussi été obtenus pour la protéine humaine HILI (Tian *et al.* 2010). Il serait intéressant d'identifier la capacité de liaison et les structures des domaines PAZ sans le résidu méthionine conservé et de mesurer l'affinité pour les ARN simple et double brin, méthylés et non méthylés.

Pour en revenir à la description des composants de la voie des piARN, les restrictions dues à l'utilisation de la souris comme modèle devaient être dépassées en établissant un nouveau modèle plus flexible, ce qui a été fait en utilisant *Danio rerio*. Deux facteurs de la voie des piARN chez la souris sont l'hélicase putative MOV10L, que l'on pense être impliquée dans la stabilisation ou le chargement des piARN dans les protéines Piwi (Frost *et al.* 2010; Lovasco *et al.* 2010; Wang *et al.* 2001b; Zheng *et al.* 2010) et la protéine Tudor-domain-containing TDRD1, qui est une composante essentielle du complexe d'inhibition impliquant les piARN (Kojima *et al.* 2009; Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a). Ici cette étude visait à identifier et à caractériser fonctionnellement le mécanisme d'action des deux protéines en utilisant le modèle de poisson zèbre. Une fois que les protéines ont été identifiées dans le poisson, le profil d'expression de ces protéines a révélé que les ARNm de *Drtdrd1* et *Drmov10l* sont exprimés à partir du stade de l'oeuf tout au long du développement du poisson zèbre poisson de façon similaire aux protéines Piwi du poisson zèbre, suggérant que les ARNm de toutes ces protéines sont héritées de la mère. En utilisant un anticorps purifié par affinité contre *Dr*TDRD1, il a été montré que la protéine est exprimée dans les testicules. Ce résultat est en accord avec le profil d'expression de TDRD1 chez le poisson médaka (Aoki *et al.* 2008). L'expression et la localisation de *Dr*TDRD1 dans ce tissu spécifique est similaire à d'autres protéines spécifiques des cellules germinales (Kedde *et al.* 2007) et dépend de sa partie 3` non traduite. Les futures recherches pourraient localiser

*Dr*TDRD1 ainsi que *Dr*MOV10L dans les cellules germinales primordiales avec une haute résolution pour déterminer le ciblage sub-cellulaire.

Contrairement aux protéines TDRD1de souris, aucun piARN lié n'a été détecté au cours d'immunoprécipitations de *Dr*TDRD1. Soit la quantité de piARN était insuffisante pour la détection, soit *Dr*TDRD1 se lie à des protéines Piwi non chargées, ou soit elle ne s'associe pas du tout aux protéines Piwi. Il a été montré que la protéine TDRD1 de souris permet la détection des arginines symétriquement diméthylées dans le domaine N-terminal d'au moins MILI et MIWI (ce n'est pas clair pour MIWI2) (Kojima *et al.* 2009; Reuter *et al.* 2009; Vagin *et al.* 2009; Wang *et al.* 2009a). Malheureusement, aucune interaction entre protéines PIWI du poisson-zèbre et *Dr*TDRD1 n'a pas été observée dans les extraits tissulaires endogènes ou dans des expériences reposant sur des cultures cellulaires de mammifère. Des immunoprécipitations qui utiliseraient de meilleurs anticorps pourraient répondre à la question de l'interaction dans le futur.

Pour examiner la fonction de *Dr*MOV10L, des expériences de knockdown ont été réalisées en utilisant des morpholinos dirigés contre le site d'initiation de la traduction de *Dr*MOV10L et qui ont été injectés dans un embryon au stade d'une cellule. Comme chez la protéine MOV10L de souris (Zheng *et al.* 2010) et le poisson zèbre knockout pour *zili* (Houwing *et al.* 2008), l'activité des éléments transposables a augmenté à la suite de l'injection de morpholinos. La similitude avec les autres phénotypes indique que *Dr*MOV10L est directement impliquée dans la voie des piARN et est nécessaire pour inhiber les rétrotransposons. La méthode de knockdown via l'utilisation de morpholinos pourrait être un outil puissant pour identifier les facteurs impliqués dans la biogenèse des piARN qui influencent l'activité des éléments transposables. Les futures expériences impliquant des complémentations des phénotypes de knockdown avec soit les protéines intégrales ou mutées (délétion, mutation ponctuel) devraient révéler le rôle fonctionnel des différents domaines et leur activité d'hélicase putative.

Dans l'ensemble, mon travail de thèse a identifié une base biochimique pour le recrutement de la piARN-méthyltransférase dans la voie des piARN de souris. Ce travail est maintenant étendu à une analyse structurale de l'interaction qui a été décrite ici. Ensemble, ces études seront de grand intérêt pour le domaine. En ce qui concerne le développement des embryons de poisson zèbre comme modèle pour étudier les facteurs piARN dans le laboratoire, les outils nécessaires pour répondre à ces questions ont été mis en place et il a été

démontré leur utilité en étudiant deux facteurs. Des expériences de knockdown par morpholinos suivies par la détection des transposons endogènes pourraient être utilisées comme un rapide test fonctionnel pour valider les facteurs candidats identifiés comme associés aux piARN vertébrés par purifications biochimiques.

Supplements

5.1 Eckhardt, S., Szostak, E., Yang, Z., Pillai, R. S. (2010). *Artificial tethering of Argonaute proteins for studying their role in translational repression of target mRNAs***.** Argonaute Proteins: Methods in Molecular Biology; series by Human Press.

5.2 Bernd Simon, John P. Kirkpatrick, Stephanie Eckhardt, Michael Reuter, Elsa A. Rocha, Miguel A. Andrade-Navarro, Peter Sehr, Ramesh S. Pillai, Teresa Carlomagno (2011). *Specific Recognition of 3'-end 2'-O-methylated piRNA by the PAZ domain of a piwi protein*. Structure

5.1 Artificial tethering of Argonaute proteins for studying their role in translational repression of target mRNAs

Stephanie Eckhardt[1,2], Emilia Szostak[1] , Zhaolin Yang[1] and Ramesh Pillai[1]

[1]European Molecular Biology Laboratory, 6 Rue Jules Horowitz, BP 181, 38042 Grenoble, France.

[2]EMBL International PhD Programme.

Correspondence to:
Ramesh Pillai
Phone: 00-33-4-76 20 7446
Fax: 00-33-4-76 20 7199
E-mail: pillai@embl.fr

i. Abstract

Small RNAs like microRNAs (miRNAs) and small interfering RNAs (siRNAs) associate with a member of the RNA-binding Argonaute family proteins. Together they participate in transcriptional and post-transcriptional gene silencing mechanisms. The fate of the target mRNA is determined, in part, by the degree of complementarity with the small RNA. To examine the exact role of the Argonaute protein in the silencing complex, human Argonautes were artificially recruited to reporter mRNAs in a small RNA-independent manner by the BoxB-N-peptide tethering system. Tethering of Argonaute proteins to a reporter mRNA leads to inhibition of translation, mimicking the repression seen with miRNAs. Similar tethering experiments were performed with fly and fission yeast Argonaute proteins and other components of the small RNP (ribonucleoprotein) complex, uncovering their specific roles in the silencing complexes containing them.

ii. Key Words: Tethering; BoxB; N-peptide; Argonaute; translational repression; miRNA

1. Introduction

Small RNAs like microRNAs (miRNAs) and small interfering RNAs (siRNAs) participate in a variety of gene regulatory pathways in most eukaryotes studied (1,2). Both classes of RNAs function as part of RNA-protein complexes which contain a member of the Argonaute family as a central component (3). The Argonaute (Ago) proteins recruit additional interacting factors to assemble a silencing complex called RNA-induced silencing complex (RISC)-containing siRNAs or a miRNA ribonucleoprotein (miRNP), when miRNA is the guide RNA (4). In most eukaryotes, the role of small RNAs in post-transcriptional silencing by either target mRNA degradation or repression of translation is well described. In the fission yeast, *S. pombe*, siRNAs associate with Ago1 to form a RNA-induced transcriptional silencing (RITS) complex that participates in assembly of heterochromatin at various genomic loci (5). Although transcriptional silencing by small RNAs in higher eukaryotes is reported, it is presently unclear if this is a wide-spread phenomenon (6).

Initial biochemical studies indicated that the RISC and miRNP are distinct complexes, yet were shown to share several key components (7). Also, it was revealed that the degree of complementarity between the small RNA in these complexes and the target has a decisive role in determining the fate of the target mRNA. Perfectly complementary targets are degraded by the endonuclolytic activity (slicer activity) of the Ago protein (8,9). In contrast, partially complementary targets are silenced by translational repression, with very little change in target RNA levels. In other instances, partially complementary targets are also subject to mRNA decay but via a pathway that is mechanistically distinct from slicer-mediated RNA cleavage (10). To appreciate the contribution of Ago proteins in the silencing process, they were artificially tethered to a reporter mRNA by a N-peptide- BoxB interaction system (11). This allows a small RNA-independent recruitment of Ago proteins to a target mRNA.

The first 22 amino acids (aa) of the transcriptional anti-termination protein N of the bacteriophage lambda (λN peptide) specifically recognize a short 19-nucleotide BoxB hairpin (12). N-peptide fusions of a protein of interest can be artificially recruited to a target RNA bearing one or more BoxB hairpins, allowing an independent examination of its functions (13). Similarly, human Argonaute proteins were modified by addition of the N-peptide

followed by a haemaglutinin (HA) tag at the N-terminus of the protein (Fig. 1). The HA-tag allows tracking the expression of the fusion protein by Western blot analysis. A luciferase reporter mRNA was modified by insertion of five BoxB hairpins in the 3'-untranslated region (UTR) of the RNA. Co-transfection of plasmids expressing the 5BoxB reporter mRNA and N-HA-Ago2 fusion protein into HeLa cell cultures results in repression of the reporter mRNA (Fig. 2). Presence of multiple BoxB hairpins leads to increased repression, which is very similar to the observation that multiple miRNA-binding sites on a target mRNA contributes to efficient repression. The experimental setup also allows mapping of functional domains in the Ago2 protein and the determination that all Ago proteins mediate repression of the reporter mRNA (Fig. 3). The observation that silencing is without any change in reporter mRNA levels leads to the conclusion that the default effect of Argonaute proteins is repression of translation of the target mRNA (Fig. 4). Sucrose gradient analysis indicates that translational inhibition of the RL-5BoxB reporter mRNA is at the initiation step of translation (14).

Similar approach was used by others in tethering the *S. pombe* Ago protein to nascent transcripts of a reporter, leading to recruitment of the RITS complex, and formation of heterochromatin on the target loci (15). Tethering experiments with Ago-interacting GW182 protein in fly and human cells revealed the essential role of Ago proteins in recruiting GW182 to mediate translational repression and mRNA decay (16,17). Thus, the RNA tethering assay can be used to dissect the molecular mechanism of individual components in the multiprotein small RNA complex that acts on RNA targets.

This chapter describes the protocols used in the tethering experiments with human Ago proteins as reported (11).

2. Materials

2.1. Preparation of reporter and protein expression constructs

1. The mammalian expression vector pCIneo (Promega) was modified by insertion of annealed oligonucleotides coding for the 22 aa λN-peptide, followed by

the 14 aa HA tag or carrying HA tag alone, giving rise to pCIneo-NHA and pCIneo-HA constructs (Fig. 1b).

2. The full-length human Argonaute proteins or their deletion versions, are cloned downstream of the N-HA tag in the EcoRI/NotI sites of pCIneo-N-HA construct. This results in the expression of N-HA-tagged Argonaute protein from the cytomegalovirus (CMV) promoter.

3. Vectors expressing the R. reniformis (phRL-TK; expressing the "humanized" Renilla luciferase) and P. pyralis or fire fly (pGL3 Promoter) luciferases are commercially available (Promega). Annealed oligonucleotides encoding five separate units of the BoxB hairpin sequences (5BoxB) were inserted downstream of the stop codon in the XbaI site of phRL-TK vector resulting in the phRL-TK-5BoxB construct (Fig. 1c). To achieve higher expression levels of the reporter required for Northern blot analysis, cassettes encoding the RL-5BoxB were subcloned into the pCIneo backbone to obtain pCMV-RL-5BoxB construct. All modified constructs are available from the authors (11).

2.2. Mammalian cell culture, transfections and luciferase assays

1. Dulbecco's Modified Eagle's medium (DMEM; Invitrogen) and DMEM supplemented with L-Glutamine (Invitrogen), pencillin/streptamycin (Invitrogen) and 10 % fetal bovine serum (Invitrogen). Media is warmed up to 37 ˚C before use.

2. Exponentially growing Hela cell cultures is seeded into 6-well tissue culture plates (Costar) and incubated at 37 ˚C.

3. Transfection reagents- Plus reagent (Invitrogen) and Lipofectamine (Invitrogen).

4. Sterile, autoclaved 1x PBS (phosphate-buffered saline) (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na2HPO4•7H2O, 1.4 mM KH2PO4, pH 7.3).

5. Dual-Luciferase Reporter Assay kit (Promega) and luminometer (Centro LB 960, Berthold Technologies).

2.3. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

1. Hoeffer SE-260 gel system (GE Health) and Hamilton needle (Hamilton)
2. Buffer A (4x) (resolving gel buffer): For 1 L, dissolve 181.64 g Tris base (MW 121.1) in water, adjust to pH 8.8. Add 20 mL 20% SDS (Sodium Dodecyl Sulfate).
3. Buffer B (4x) (stacking gel buffer): For 500 ml, dissolve 30.28 g Tris base (MW 121.1) in water, adjust to pH 6.8. Add 10 mL 20% SDS.
4. Running buffer (1x): 0.025 M Tris base, 0.19 M Glycine and 0.1% SDS (make a 10x stock).
5. SDS-PAGE sample loading buffer (2x): 100 mM Tris-HCl pH 6.8, 4 % (w/v) SDS, 0.2 % (w/v) bromophenol blue, 20 % (v/v) glycerol, 10 % (v/v) 2-mercaptoethanol.
6. Thirty percent (30 %) acrylamide/bis-acrylamide solution (37.5:1) (National diagnostics) (see Note 1).
7. Ammonium persulfate: prepare a 10 % solution in water, keep at 4˚C and TEMED (see Note 2).
8. Pre-stained protein molecular weight markers (Fermentas).

2.4. Western blot analysis

1. Transfer buffer (1x): 25 mM Tris, 190 mM Glycine, 20 % Methanol.
2. Reinforced nitrocellulose membrane (Whatman OPTITRAN BA-S 85) and Whatman 3MM chromatography paper (Whatman).
3. Semidry transfer apparatus from (Biorad; Transblot apparatus)
4. Blocking buffer (1x): 5 % non-fat dry milk powder in PBS-Tween (1x PBS with 0.1 % Tween-20).
5. Primary antibodies: Anti-HA rabbit polyclonal (Y-11, Santa Cruz, sc-805); rabbit IgG horse radish peroxidise-conjugate (HRP) (GE Health).
6. ECL Plus Western Blotting Detection system (GE Health). Hyperfilm ECL (GE Health) and film exposure Hypercassette (GE Health).
7. Developing machine (AGFA Curix 60).

2.5. Northern blot analysis

1. Absolutely RNA RT-PCR miniprep kit (Stratagene).
2. NanoDrop spectrophotometer (NanoDrop Technologies, Inc)
3. DEPC-treated water: mix 1 mL of DEPC with 1 L of double-distilled water and incubate overnight by vigorous shaking. Autoclave to remove DEPC and aliquote.
4. Water-saturated phenol (acidic phenol), chloroform and absolute ethanol.
5. Reagents for denature agarose gel: agarose, 37 % formaldehyde.
6. Running buffer for denaturing gel (10x): 0.2 M MOPS, pH 7.0, 20 mM sodium acetate and 10 mM EDTA) (see Note 3).
7. RNA sample loading buffer (2x): To obtain 3 mL, mix 1.5 mL formamide, 300 µL of 10x MOPS, 525 µL of 37% formaldehyde, 15 µL of ethidium bromide (10 mg/mL stock) and a speck each of bromophenol blue and xylene cyanol. Make up with water. Freeze in aliquots at -20 ˚C.
8. Nylon membrane (Hybond N+; GE Health), UV crosslinker (Stratalinker with 254 nm UV bulbs, Stratagene) and hybridization oven.
9. Random Primed DNA labeling kit (Roche) and microspin G-25 spin columns (GE Health). Radioactive nucleoside triphosphate [□-32P] dCTP, 3000 Ci/mmol (Perkin Elmer).
10. Hybridization buffer (1x; Church buffer): 0.25 M NaHPO4, 0.25 M NaH2PO4, 1 mM EDTA, 1 % BSA (Bovine Serum Albumin), 7 % SDS.
11. Prepare 20x SSC (Sodium chloride/sodium citrate): 3 M NaCl, 0.3 M sodium citrate, adjust pH to 7.0 with 1M HCl.
12. Northern wash buffer: 2x SSC with 0.1 % SDS (low stringency wash) and 0.2x SSC with 0.1 % SDS (high stringency wash). Buffers are warmed up to 65 ˚C before use.

3. Methods

3.1. Transfection of HeLa cell cultures with reporters

1. Seed HeLa cells into required number of 6-well plates at a dilution so that they are 60 - 70 % confluent at the time of use the next day.
2. For each well prepare a mixture of the following plasmids in a 1.5 mL eppendorf tube: 100 ng phRL-TK-5BoxB, 100 ng of pGL3 Promoter (for normalization of transfections) and 500 ng of plasmid expressing either HA or N-HA Argonaute protein fusions. When analysis by Northern blotting is required, prepare the following: 100 ng pCMV-RL-5BoxB, 100 ng of pGL3 Promoter (for normalization of transfections), 500 ng of plasmid expressing either HA or N-HA Argonaute protein fusions and 75 ng of a GFP expression plasmid for normalization of RNA levels. The strong CMV promoter gives considerably higher expression levels than the weaker Herpes Simplex Virus Thymidine Kinase (HSV-TK) promoter, required for detection of the RL-5BoxB mRNA by Northern blotting.
3. The DNA mixture is diluted in DMEM to a final volume of 100 µL.
4. Add 4 µL of Plus Reagent and mix by pipetting, incubate for 15 minutes (min) at room temperature.
5. In the meantime, mix 4 µL of Lipofectamine Reagent with 96 µL of DMEM media. Incubate for 15 min at room temperature.
6. Combine the DNA-Plus mix with the Lipofectamine solution and mix well by pipetting. Incubate a further 20 min at room temperature.
7. Each transfection in an experimental series is performed in triplicates, so a master-mix for three transfections can be prepared (steps 2-6). The experiment series is also reproduced at least three times.
8. During the incubation period, wash the cells sitting in the 6-well plates: remove culture media by aspiration and wash once with 2 mL of DMEM media. Remove the wash and replace with 800 µL of DMEM. Leave cells in the 37 ˚C incubator till required.

9. Remove cells from the incubator and gently pipette the transfection mix from step 5 into the wells. Return cells to incubator.
10. Replace the media after 4 hours (h) of incubation with fresh culture media containing supplements.
11. Change media once more after 24 h.

3.2. Luciferase assays

1. Around 24 h post-transfection, remove media by aspiration and wash cells with 2 mL of 1x PBS. Remove as much of the 1x PBS as possible to prevent dilution of the cell extract.
2. Add 300 µL of passive lysis buffer (1x PLB) provided in the Dual-Luciferase Reporter Assay kit (Promega) to the cells in each well of the 6-well plate. Incubate for 15 min at room temperature with gentle shaking on a tilting platform.
3. Luciferase measurements are made with a luminometer using 10 µL of the lysate from each well. The following amounts of the luciferase assay reagents are used: 50 µL of LAR II and 25 µL of Stop and Glo. A detailed protocol is provided by the manufacturer of the kit (Promega). Due to the high level of expression of hRL from the CMV promoter, the lysate might need to be diluted with 1x PLB to stay within the detection range of the luminometer.
4. Analysis of the luciferase measurements: the value obtained from RL measurements is divided by the fire fly measurements from the same sample. This normalized value for the triplicates of each transfection is then used to calculate the mean and the standard deviation (SD). The mean obtained in transfections with plasmid expressing HA-Argonaute fusion is set to 1 (Fig. 2a).

3.3. SDS-PAGE

To detect the expression of the HA or N-HA-tagged Argonaute proteins or their deletions in the transfection experiments, Western analysis is performed using anti-

HA antibodies with the cell lysate. The samples are first resolved by 10 % SDS-PAGE and then analysed by Western blotting.

1. The 10 % SDS-PAGE is performed with the Hoeffer SE-260 gel system (GE Health).
2. Glass plates are scrubbed clean with soap and water, and allowed to drain in a vertical position. Clean them with 70 % ethanol applied to a tissue paper to remove any eventual trace of contaminants that can trap air bubbles while pouring the gel.
3. Set up the glass plate and alumina plate separated by 1.5 mm thick spacers on the gel caster. If doing it the first time, check whether the set up is water-tight by filling the space between the plates with water. When satisfied, pour off the water into the sink.
4. Insert the comb into the space between the plates, and using a water-proof marker pen mark a level indicating 1 cm below the end of the teeth of the comb. This will be the level to which the resolving gel will be poured. Remove the comb.
5. Prepare the following solutions without TEMED. Resolving gel solution: 2.5 mL Buffer 4x A, 3.3 mL 30 % acrylamide solution, 4.2 mL water, 66 µL 10 % APS. Stacking gel solution: 1.9 mL Buffer 4x B, 1.5 mL 30 % acrylamide solution, 2.25 mL water, 50 µL 10 % APS.
6. Add 20 µL of TEMED to the resolving gel solution and pour into the space between the plates till the level marked in step 4. This procedure has to be done as rapidly as possible to avoid the solution gelling before it can be poured. Since polymerization is accelerated by higher temperatures, it is helpful to briefly place the solution (before adding TEMED) on ice. Overlay the gel solution with 70 % ethanol to a height of about 1 mm. This removes any potential air bubbles on the surface and also accelerates the polymerization process. Allow to set for 10 min.
7. Pour off the 70 % ethanol and soak up any remaining liquid with a tissue paper.
8. Add 20 µL TEMED to the stacking gel solution and fill up the space on top of the resolving gel. Quickly insert the comb to form the wells. Let the gel sit for another 10 min (see Note 4).

9. Remove the comb from the gel, fix it to the gel apparatus and fill up the top and bottom tanks with 1x SDS-PAGE running buffer.

10. Rinse the wells by pipetting running buffer into them using 18-guage needle attached to a 5 mL syringe.

11. Mix 10 µL of the lysate prepared for the luciferase assays with sample loading dye. Prepare a similar mixture for the pre-stained marker. Heat samples for 1 min at 95 ˚C and load into the wells using a Hamilton needle (Hamilton) or a pipette tip.

12. Connect the cables to a power supply source and run the gel at 15 mA into the stacking gel and then at 30 mA till the dye front is at the bottom of the gel. The migration of the gel can also be monitored by the separation of the pre-stained marker.

13. When finished, switch off the power supply and disconnect the cables before opening the gel apparatus.

3.4. Western blotting

1. After the samples have been resolved by SDS-PAGE, the plates are removed and the stacking gel is cut off. The resolving gel is then transferred to a container and soaked in transfer buffer.

2. During this time, cut four pieces of Whatman 3MM chromatography paper having same dimensions as the resolving gel and prepare a piece of reinforced nitrocellulose membrane of same size and cut on one corner to mark orientation of the gel.

3. Soak two 3MM paper sections in transfer buffer and place on the semi-dry transfer blotting apparatus. Use a 10 mL plastic pipette to roll out the air bubbles from underneath. Place the nitrocellulose membrane and then the gel on top, followed by two more sheets of soaked 3MM paper. Each time roll out any air bubbles that can cause inefficient transfer (Fig. 5a). Place the lid of the apparatus and attach cables. Run the transfer overnight at 5 V (see Note 5).

4. Switch off the power supply and disconnect the cables. The stack is slowly disassembled and the membrane removed. The transfer of the pre-stained marker will serve to indicate that the transfer worked properly. Since the Argonaute proteins are

~90 kDa, observe whether the molecular weight marker larger than 90 kDa is clearly visible.

5. Block the membrane in blocking buffer on a tilting shaker for 30 min at room temperature.

6. Discard the blocking buffer and replace with the minimum amount of blocking buffer required to cover the membrane (approximately 1-5 mL depending on the container used to hold the membrane).

7. Add the primary antibody at the required dilution and continue incubation with shaking for 1 h. Remove antibody solution and wash five times 5 min each with abundant amounts of 1x PBS-Tween.

8. Incubate the membrane with blocking buffer containing diluted secondary antibody- anti-rabbit IgG-HRP conjugate for 1 h. Wash as in step 7.

9. Prepare the ECL detection reagent as per manufacturer's instructions. Remove the membrane from the wash buffer, place on a sheet of SaranWrap and uniformly apply the ECL detection solution for 1 min. Quickly absorb the excess ECL liquid with a tissue paper and wrap the membrane with the SaranWrap.

10. Place the membrane in the exposure cassette and proceed to the dark room. Remove one X-ray Hyperfilm, cut or fold a corner to indicate orientation and place on the membrane. After about 2 min, develop in the developer machine. During this time replace a new film in the cassette to evaluate a second, longer exposure. The anti-HA reactive proteins are visible as dark bands on the film. Overlay on the membrane to mark out the pre-stained marker positions.

3.5. RNA extraction and Northern blotting

1. Total RNA from transfected HeLa cells can be prepared using the Absolutely RNA RT-PCR miniprep kit following the manufacturer's instructions. This kit allows DNase-treatment of the RNA bound on the purification columns, which is an advantage. Efficient removal of the transfected plasmid DNA is very important for the subsequent detection of the expressed reporter mRNAs. Measure the concentration of

the purified RNA using a NanoDrop spectrophotometer. Take extreme precaution while working with RNA (see Note 6).

2. The RNA samples are resolved in a formaldehyde denaturing gel and then transferred to a positively-charged Nylon membrane (Hybond N+).

3. Clean the agarose gel tray, comb and gel running apparatus with soap and water. Allow them to dry and then set them up in the casting stand in the fume hood. Prepare formaldehyde denaturing 1 % agarose gel by boiling 1 g agarose in 72 mL of DEPC-treated water double-distilled water. Allow to cool to 60 ˚C in a waterbath (see Note 7).

4. In the fume hood, add 10 mL of 10x MOPS and 18 mL of 37 % formaldehyde. Mix and pour into the setup for agarose gel. Place the comb and allow gel to solidify for approximately 30 min.

5. In the meantime, prepare 1x MOPS running buffer by diluting the 10x buffer in autoclaved double distilled water. Pour sufficient buffer into the gel tank. Remove the comb from the gel and place the gel into the apparatus, submerged in the buffer.

6. Mix the RNA samples with equal volume of 2x RNA loading dye. Heat the mixture for 5 min at 65 ˚C (see Note 8).

7. Load the samples into the gel with a pipette tip and run at a constant 5 V/cm of gel length. When the dye front is 2-3 cm away from the bottom of the gel, stop the gel. Disconnect all cables and remove gel for examination on a UV-transilluminator. The 28S and 18S ribosomal RNA (rRNA) bands should be clearly visible on all the lanes. Observe whether the intensity of the 28S rRNA band is approximately twice that of the 18S rRNA band. This should indicate that the RNA is of good quality. Degraded RNA samples will show either equal intensities for the two bands or a smeary pattern. To prevent RNase contamination during the imaging process, always place the gel on the transilluminator with a clean SaranWrap plastic sheet between them.

8. Place the gel in a container and wash with abundant amounts of autoclaved double distilled water to remove the formaldehyde. Change the solution twice. Dispose formaldehyde-containing solutions into dedicated waste containers.

9. Prepare a transfer stack by placing a long strip of Whatman 3MM paper of the same width as the gel on a raised platform. The two ends of the strip should be in contact with the transfer buffer (20x SSC) contained in a large tray at the bottom. The strip is soaked with 20x SSC and air bubbles rolled out with a 10-mL plastic pipette. Build the stack with four sheets of 3MM paper soaked in 20x SSC, place the gel, the nylon membrane, followed by four more sheets of 3MM paper. Each of these is soaked in transfer buffer and any air bubbles are rolled out. Place a stack of dry tissue paper and top it off with a glass plate. Weight down the stack with something modestly heavy (~300g; for example, a bottle of water with 250 mL water). Let the RNA transfer to the membrane by passive transfer overnight (Fig. 5b).

10. On the next day, remove the stack and carefully extract the membrane. Place the membrane on a dry tissue paper and fix the RNA to the membrane by a short UV-crosslinking exposure in a UV-crosslinker (Stratalinker). Push the AUTO CROSSLINK button on a Model 1800 UV crosslinker from Stratagene. This will deliver energy of 120,000 microjoules from the 254 nm UV lamps to form covalent bonds between uracils of the RNA and amino groups of the nylon membrane (see Note 9).

3.6. Northern analysis using radioactive probes

The expression from the transfected luciferase plasmids can be studied by probing the RNA bound to the nylon membrane with specific radioactive probes complementary to the mRNA of interest. Working with radioactivity requires specialized laboratory equipment and specialist training. Please consult the authorities at your institution before embarking on experiments involving radioactivity.

1. The coding region for the Renilla luciferase is released from the phRL-TK plasmid by restriction digestion, gel eluted, purified using a gel-extraction kit (QIAgen). The purified RL DNA fragment is then used for preparation of radioactive probes using the random-primed DNA labelling kit (Roche) using [γ-32P] dCTP as per manufacturer's instructions.

2. Labelled single-stranded DNA probes are seprated from free unincorporated deoxyribonucleoside triphosphates by the use of Microspin G-25 columns. Follow manufacturer's protocol. Store prepared probe at -20 ˚C until use.

3. Place the nylon membrane into a hybridization bottle and add sufficient hybridization buffer prewarmed to 65 ˚C. Prehybridize for 1 h at 65 ˚C in a rotating hybridization oven. This allows efficient blocking of the membrane to prevent background signals.

4. Denature radioactive probe at 95 ˚C for 1 min and immediately place on ice.

5. Remove the hybridization solution and replace with fresh hybridization solution. Add probe at 1 million counts per minute (cpm)/mL. Hybridize at 65 °C overnight.

6. On the following day, remove radioactive hybridization solution into proper waste container. Wash twice with low-stringency wash buffer (2x SSC containing 0.1 % SDS), each for 10 minutes. Wash once with high-stringency wash buffer (0.1x SSC containing 0.1 % SDS) for 5 min. Check for presence of counts on the membrane using a Geiger-Muller counter.

7. Seal the membrane in plastic sheeting and expose to a storage phosphor screen for 5 h. Scan the screen in a Typhoon Scanner (GE Health) to obtain the signals arising from radioactive probe hybridized to the RL mRNA present in the different samples on the membrane.

8. The membrane can be stripped to remove the radioactive probe for RL and re-probed with a probe for GFP to allow normalization of transfection efficiency.

9. Signals are quantified using the software provided with the Typhoon Scanner (Image Quant).

4. Notes

1) Wear gloves while handling this reagent as it is a neurotoxin.
2) Ammonium persulfate: Renew the stock every 3 months.
3) Running buffer for denaturing gel (10x): Store in dark up to 3 months at 4 ˚C.

4) At this step the gel can be stored for later use by wrapping in tissue paper soaked with 1x running buffer. Then place the gel into a plastic packet or wrap with SaranWrap. Store at 4 ˚C.

5) Increasing the voltage can allow reduced transfer times. However, better transfer efficiency is observed with longer duration.

6) Always wear gloves and use RNase-free plasticware (tips, reaction tubes etc).

7) Formaldehyde is toxic and has to be handled with gloves and only under the fume hood.

8) Be carefully as the loading buffer contains ethidium bromide, which is a carcinogen and has to be handled with proper gloves. Do not dilute the loading buffer more than 50 %, to maintain the required denaturing conditions for the sample.

9) The UV-crosslinked membrane can be stored dry between two Whatman 3MM sheets until further use.

5. References

1. Filipowicz, W., Bhattacharyya, S.N. and Sonenberg, N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet, 9, 102-114.
2. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell, 116, 281-297.
3. Carmell, M.A., Xuan, Z., Zhang, M.Q. and Hannon, G.J. (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev, 16, 2733-2742.
4. Meister, G. and Tuschl, T. (2004) Mechanisms of gene silencing by double-stranded RNA. Nature, 431, 343-349.
5. Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I. and Moazed, D. (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. Science, 303, 672-676.
6. Kim, D.H., Villeneuve, L.M., Morris, K.V. and Rossi, J.J. (2006) Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. Nat Struct Mol Biol, 13, 793-797.
7. Siomi, H. and Siomi, M.C. (2009) On the road to reading the RNA-interference code. Nature, 457, 396-404.
8. Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G. and Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell, 15, 185-197.
9. Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L. and Hannon, G.J. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. Science, 305, 1437-1441.
10. Eulalio, A., Huntzinger, E. and Izaurralde, E. (2008) Getting to the root of miRNA-mediated gene silencing. Cell, 132, 9-14.
11. Pillai, R.S., Artus, C.G. and Filipowicz, W. (2004) Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. Rna, 10, 1518-1525.
12. Legault, P., Li, J., Mogridge, J., Kay, L.E. and Greenblatt, J. (1998) NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. Cell, 93, 289-299.
13. Baron-Benhamou, J., Gehring, N.H., Kulozik, A.E. and Hentze, M.W. (2004) Using the lambdaN peptide to tether proteins to RNAs. Methods Mol Biol, 257, 135-154.

14. Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E. and Filipowicz, W. (2005) Inhibition of translational initiation by Let-7 MicroRNA in human cells. Science, 309, 1573-1576.
15. Buhler, M., Verdel, A. and Moazed, D. (2006) Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell, 125, 873-886.
16. Eulalio, A., Huntzinger, E. and Izaurralde, E. (2008) GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. Nat Struct Mol Biol, 15, 346-353.
17. Chekulaeva, M., Filipowicz, W. and Parker, R. (2009) Multiple independent domains of dGW182 function in miRNA-mediated repression in Drosophila. Rna, 15, 794-803.

Acknowledgments

The protocols were originally developed with the active help and advice of Caroline G. Artus and Witold Filipowicz at the Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. Research in R.S.P's group is supported by the European Molecular Biology Laboratory.

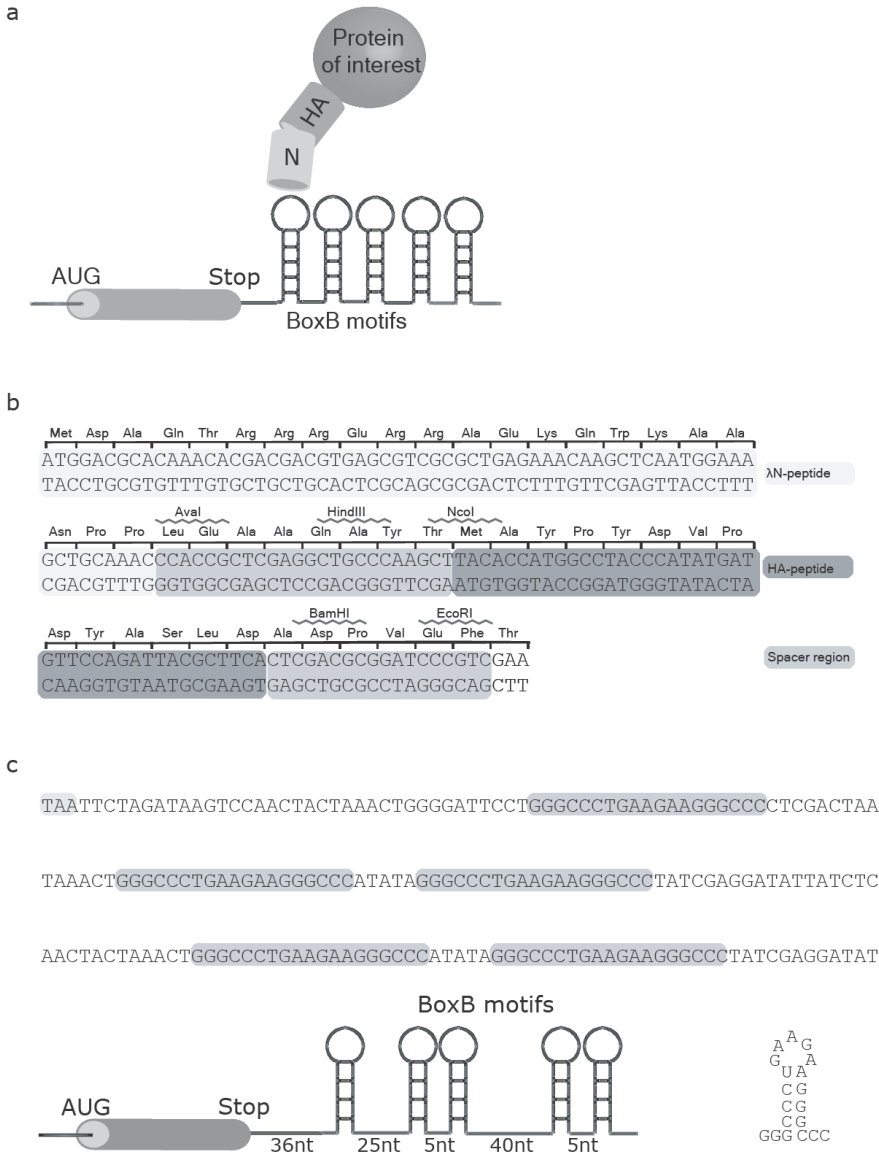


Fig. 1
The Argonaute tethering experiment.

(**a**) A schematic representation of the Renilla luciferase reporter mRNA with 5BoxB hairpins inserted into the 3'- untranslated region. The N-HA-tagged Argonaute protein is depicted as binding the BoxB hairpin; several of the five hairpins might be bound by the protein in the actual experiment.

(**b**) The coding sequence for the N-HA peptides in the pCIneo-N-HA vector (Pillai *et al.* 2004) and the EcoRI site used for cloning the Argonaute proteins is shown.

(**c**) Sequence context of the five BoxB hairpins (highlighted) in the 3'-UTR of the RL-5BoxB mRNA. The stop codon (TAA) of *Renilla* luciferase is indicated.

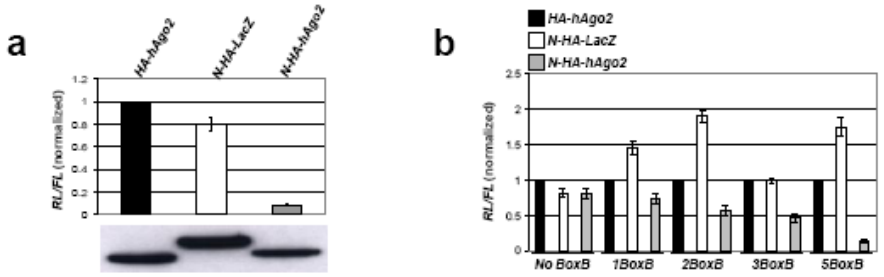


Fig. 2

(**a**) RL activity detected in extracts from HeLa cells expressing the indicated fusion proteins. Cells were co-transfected with constructs expressing the RL-5BoxB reporter, fire fly (FL) reference, and indicated fusion proteins. Histograms represent normalized mean values (±SD) of RL/FL activities from a minimum of three experiments. RL activity values seen in the presence of HA-hAgo2 were set as 1. Expression levels of fusion proteins, as determined by Western analysis using anti-HA antibody, are shown *below* the histogram. The aliquot of the N-HA-LacZ-expressing extract applied to the gel was 10 times smaller than aliquots of other extracts. Generally, the N-HA-LacZ protein is expressed at a ~10-fold higher level than N-HA-hAgo2 or HA-hAgo2. Increased RL activity in extracts expressing N-HA-LacZ is likely due to the effect of the protein on the stability of mRNA reporters containing BoxB sequences.

(**b**) Activity of reporter RL mRNAs containing different numbers of BoxB hairpins.
(Reproduced from **ref. 11** with permission from Cold Spring Harbour Laboratory Press.)

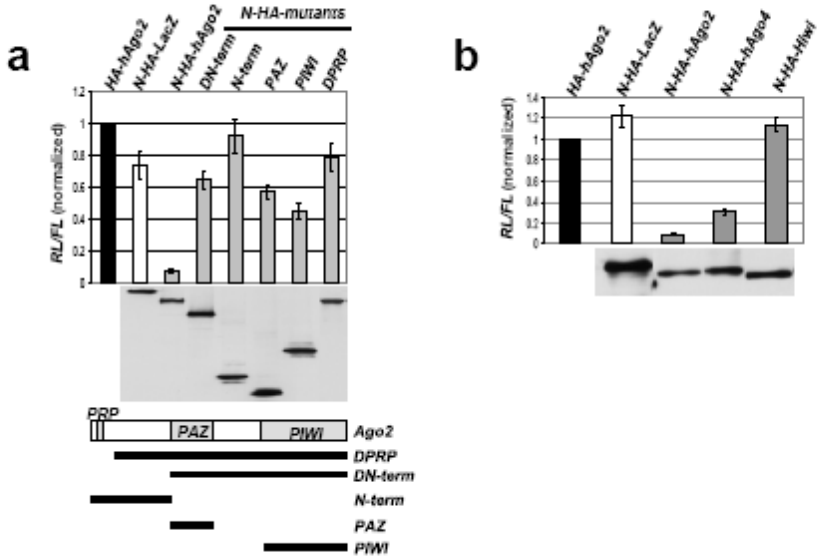


Fig. 3

Tethered hAgo2 and hAgo4, but not hAgo2 mutants and Hiwi, induce the repression.

(**a**) RL activity in extracts from HeLa cells cotransfected with plasmids expressing RL-5BoxB reporter and N-HA-tagged mutant hAgo2 fusions. Western analysis of fusion protein expression is shown *below* the histograms. Schematic representation of hAgo2 and its deletion mutants is shown in *lower* panel.

(**b**) Tethering of hAgo4 but not the Piwi protein Hiwi represses translation. (Reproduced from **ref. 11** with permission from Cold Spring Harbour Laboratory Press.)

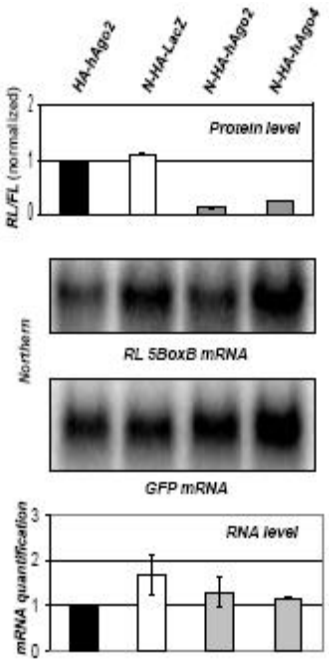


Fig. 4
Repression by N-HA-hAgo2 and N-HA-hAgo4 occurs without changes in reporter mRNA level. Northern analysis (*middle* panels) was performed with total RNA isolated from transfected cells, using probes specific for RL-5BoxB mRNA and green fluorescence protein (GFP) mRNA, expressed from the co-transfected plasmid. The RL activity in extracts from the same transfected cells is shown in the *upper* panel. Storage phosphor screen scan quantification of the RL-5BoxB mRNA, normalized to GFP mRNA, is shown in the *lower* panel; values are means (±SD) from three independent experiments. The RL-5BoxB mRNA level in cells co-transfected with HA-hAgo2 is set to 1.
(Reproduced from **ref. 11** with permission from Cold Spring Harbour Laboratory Press.)

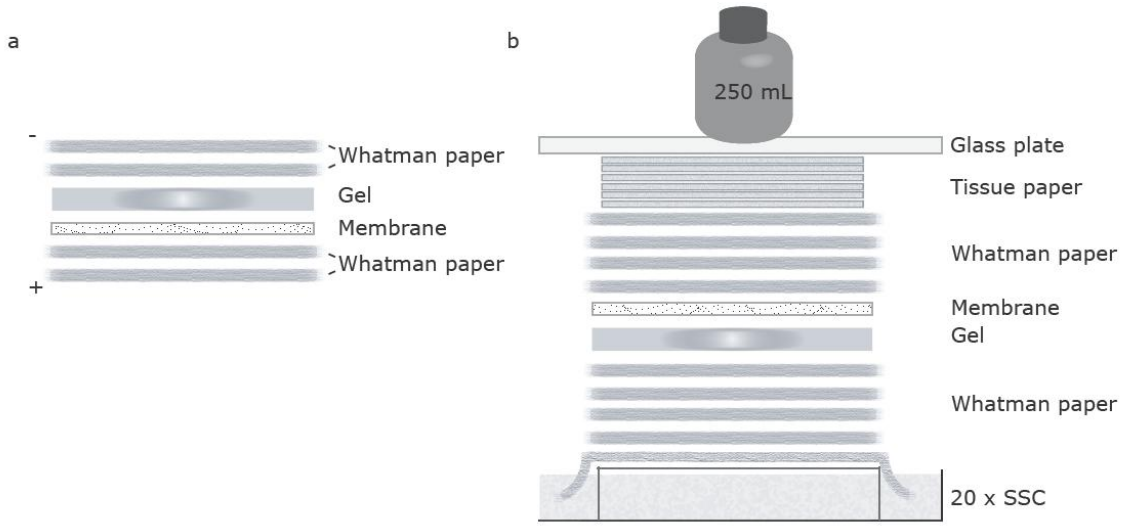


Fig. 5
(a) Assembly setup for Western blot transfer.
(b) Assembly setup for Northern blot transfer. Please see sections 3.4 and 3.5 for details.

5.2

Please cite this article in press as: Simon et al., Recognition of 2′-O-Methylated 3′-End of piRNA by the PAZ Domain of a Piwi Protein, Structure (2011), doi:10.1016/j.str.2010.11.015

Structure
Article

Recognition of 2′-*O*-Methylated 3′-End of piRNA by the PAZ Domain of a Piwi Protein

Bernd Simon,[1] John P. Kirkpatrick,[1,5] Stephanie Eckhardt,[2] Michael Reuter,[2] Elsa A. Rocha,[2] Miguel A. Andrade-Navarro,[3] Peter Sehr,[4] Ramesh S. Pillai,[2,*] and Teresa Carlomagno[1,*]

[1]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany
[2]European Molecular Biology Laboratory, Grenoble Outstation, 6 Rue Jules Horowitz, BP 181, 38042 Grenoble, France
[3]Max Delbrück Center for Molecular Medicine, Computational Biology and Data Mining, Robert-Rössle Strasse 10, D-13125 Berlin, Germany
[4]Chemical Biology Core Facility, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany
[5]Present address: Department of Structural and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK
*Correspondence: ramesh.pillai@embl.fr (R.S.P.), teresa.carlomagno@embl.de (T.C.)
DOI 10.1016/j.str.2010.11.015

SUMMARY

Piwi proteins are germline-specific Argonautes that associate with small RNAs called Piwi-interacting RNAs (piRNAs), and together with these RNAs are implicated in transposon silencing. The PAZ domain of Argonaute proteins recognizes the 3′-end of the RNA, which in the case of piRNAs is invariably modified with a 2′-*O*-methyl group. Here, we present the solution structure of the PAZ domain from the mouse Piwi protein, MIWI, in complex with an 8-mer piRNA mimic. The methyl group is positioned in a hydrophobic cavity made of conserved amino acids from strand β7 and helix α3, where it is contacted by the side chain of methionine-382. Our structure is similar to that of Ago-PAZ, but subtle differences illustrate how the PAZ domain has evolved to accommodate distinct 3′ ends from a variety of RNA substrates.

INTRODUCTION

In most eukaryotes, the conserved family of Argonaute proteins associate with small RNAs to participate in diverse gene silencing mechanisms (Ghildiyal and Zamore, 2009). The ubiquitously expressed members of the Ago subfamily associate with ~21 nucleotide (nt) microRNAs (miRNAs) and small interfering RNAs (siRNAs) to form the central component of the RNA-induced silencing complex (RISC). Guided by the small RNA, the RISC mediates posttranscriptional silencing by targeted mRNA degradation or translational repression (Filipowicz et al., 2008). In fission yeast, the single Argonaute protein incorporates siRNAs to form a related RNA-induced transcriptional silencing complex (RITS) that functions by inducing formation of heterochromatin (Verdel et al., 2004). Some Argonautes are RNA-guided endonucleases (Slicers) that catalyze cleavage of target nucleic acids. Thus, at least some of the RISC functions can be directly attributed to the enzymatic activity of its Argonaute component (Parker et al., 2006; Patel et al., 2006a; Tolia and Joshua-Tor, 2007). In prokaryotes, the role of the Argonaute proteins remains obscure, but RNA cleavage has been demonstrated in vitro with pAgo/DNA/RNA ternary complexes from *A. aeolicus* and *T. thermophilus* (Wang et al., 2009; Yuan et al., 2005).

Animal germlines express additional Argonautes that belong to the Piwi subfamily. They associate with a class of 25–31 nt small RNAs called the Piwi-interacting RNAs (piRNAs), which are implicated in transposon silencing (Aravin et al., 2007). One common feature of piRNAs in all systems analyzed so far is that they derive from repetitive transposon-rich regions of the genome. Considered together with the fact that Slicer activity has been demonstrated for fly Piwi proteins, it is likely that Piwi proteins execute part of their functionality by posttranscriptional cleavage of transposon transcripts (Brennecke et al., 2007; Gunawardane et al., 2007). In mammals, repeat-derived piRNAs are also believed to guide Piwi proteins to mediate transcriptional silencing by inducing DNA methylation of promoter elements of transposon sequences (Aravin et al., 2008; Kuramochi-Miyagawa et al., 2008).

Small RNAs interacting with members of the Ago and Piwi subfamilies display distinct mechanisms of biogenesis. siRNAs and miRNAs derive from long double-stranded RNA (dsRNA) and stem-loop structures, respectively. These precursors are processed by the RNaseIII enzyme DICER, which produces duplex intermediates with characteristic 2 nt 3′-overhangs. The 3′ overhangs are important features recognized by the RNAi machinery during small RNA biogenesis (Jinek and Doudna, 2009; Patel et al., 2006b). Subsequent processing involves strand separation in Ago-containing complexes, resulting in the mature single-stranded small RNA being loaded onto the Ago protein. On the other hand, biogenesis of piRNAs is believed not to include dsRNA precursors, a hypothesis supported by the observation that piRNA production is unaffected in *dicer* mutants (Houwing et al., 2007; Vagin et al., 2006). Long, single-stranded RNA (ssRNA) precursors are the likely sources of piRNAs (Aravin et al., 2007), but the exact mechanisms of piRNA biogenesis are presently unknown.

All mature small RNAs are characterized by 5′-monophosphate and 3′-hydroxyl (OH) groups, features that are important for their recognition by the RNAi machinery. Some Ago-bound small RNAs also carry an additional 2′-*O*-methyl marker on the 3′-terminal nucleotide, as has been shown for miRNAs and siRNAs in *Arabidopsis* (Yu et al., 2005). In *Drosophila*, miRNAs are mainly loaded into DmAGO1, while siRNAs and a few miRNAs associate with DmAGO2 (Tomari et al., 2007). Only small RNAs associating with DmAGO2 are modified with the

Please cite this article in press as: Simon et al., Recognition of 2′-O-Methylated 3′-End of piRNA by the PAZ Domain of a Piwi Protein, Structure (2011), doi:10.1016/j.str.2010.11.015

methyl marker (Ameres et al., 2010; Horwich et al., 2007; Yu et al., 2005). In contrast, the 2′-*O*-methyl group on the 3′-terminal nucleotide is recognized as a defining feature of piRNAs in all animals studied to date (Ghildiyal and Zamore, 2009).

In plants (which lack piRNAs), the 3′-end methyl marker is added by the RNA methyltransferase enzyme HEN1, and is known to enhance RNA stability by providing protection from 3′-end uridylation activity (Li et al., 2005). In *Tetrahymena*, deletion of the *hen1* gene produces unstable, shortened scanRNAs (equivalent to piRNAs), causing defects in programmed DNA elimination (Kurth and Mochizuki, 2009). In flies, the HEN1 homolog is responsible for modifying both piRNAs and DmAGO2-associated small RNAs. Loss of fly *hen1* results in reduced piRNA and siRNA stability, which in the case of siRNAs is shown to be due to 3′-uridylation and 3′-exonuclease activities (Ameres et al., 2010; Horwich et al., 2007; Saito et al., 2007). Importantly, in all these systems, in the absence of HEN1, nonmethylated small RNAs are still loaded properly into their respective Argonaute proteins. Taken together, these observations indicate a stabilizing role for the methyl marker on small RNAs, subsequent to their loading into Argonaute proteins.

The Argonaute proteins recognize RNA through their two protein domains PAZ and PIWI, which bind the 3′- and 5′-ends, respectively (Jinek and Doudna, 2009). Structures of full-length Argonautes in isolation or in complex with RNA substrates have been solved only for prokaryotic thermophilic organisms such as *P. furiosus* (Song et al., 2004), *A. aeolicus* (Rashid et al., 2007; Yuan et al., 2005, 2006), and *T. thermophilus* (Wang et al., 2009). The PIWI domain is structurally specialized into a region with an RNaseH-like fold harboring the Slicer activity, and the MID domain that contains a highly conserved binding pocket for the 5′-monophosphate of the small RNA (Frank et al., 2010; Ma et al., 2005; Parker et al., 2005; Wang et al., 2009). The PAZ domain also has nucleic acid-binding properties, and structures of Ago-PAZ domains from *Drosophila* and human have revealed the presence of an oligosaccharide/oligonucleotide-binding (OB)-fold. When analyzed in complex with RNA substrates, these structures also reveal how the 3′-terminal dinucleotide overhangs of siRNA duplexes or single-stranded (ss)RNAs are inserted into a preformed hydrophobic pocket constructed from conserved aromatic residues (Lingel et al., 2004; Ma et al., 2004; Song et al., 2003; Yan et al., 2003).

RNA-binding studies with the PAZ domain of the human Ago subfamily member, HsAGO1, indicate a preference for RNA substrates lacking the 2′-*O*-methyl modification (Ma et al., 2004). Since piRNAs invariably carry such a modification (Ghildiyal and Zamore, 2009), the PAZ domain of Piwi proteins must accommodate an additional methyl group in its putative 3′-end binding pocket. In order to understand this recognition process, we determined the solution structure of the PAZ domain from mouse Piwi-like protein 1 (PIWIL1; MIWI) (Girard et al., 2006) bound to an 8-mer RNA, which was 2′-*O*-methylated at its 3′-end. The methyl group is inserted into a hydrophobic cavity, within which it is contacted by the long side chain of a methionine residue. However, this methionine is not conserved across the PAZ domains of the Piwi family, indicating that the methionine contact is not essential for recognition of the piRNA methylated 3′-end. Conserved differences of residues between the central β barrel and the α helix/β hairpin motif between the Piwi, Ago and Dicer families translate into subtle structural differences. The α helix/β hairpin motif is slightly displaced in the MIWI-PAZ relative to Ago-PAZ, leading to a small widening of the RNA binding pocket. In addition to representing the first structural information for a ribonucleoprotein (RNP) complex involved in the piRNA pathway, our structure reveals the mechanisms of 3′-end methyl recognition and underlines the adaptability of PAZ domains to bind RNAs with different sequences, secondary structure, and modifications.

RESULTS

piRNAs Associating with All Mouse Piwi Proteins Are Modified at the 3′-End

Mouse piRNAs in adult testis total RNA samples have been demonstrated to show resistance to sodium periodate treatment, suggesting that they are 3′-terminally modified; this modification was identified to be a 2′-*O*-methyl group (Kirino and Mourelatos, 2007; Ohara et al., 2007). However, of the three mouse Piwi proteins (see Figure S1 available online), only MIWI and MILI are expressed in adult testes. The third Piwi member, MIWI2, is coexpressed with MILI during embryonic development (Aravin et al., 2008; Kuramochi-Miyagawa et al., 2008). To determine the methylation status of piRNAs interacting with all Piwi proteins, we immunopurified the three Piwi proteins from adult and embryonic mouse testes extracts and performed a β-elimination assay (Vagin et al., 2006). In brief, treatment of RNAs carrying 2′- and 3′-hydroxyl groups with sodium periodate ($NaIO_4$) is followed by β-elimination, and loss of the terminal nucleotide. In contrast, modification of either of the hydroxyl groups prevents reaction with $NaIO_4$. As shown in Figure 1A, a synthetic nonmethylated control RNA migrates faster upon treatment, while piRNAs associated with all three mouse Piwi proteins show unaltered mobility, demonstrating that all mouse Piwi proteins are associated with 3′-end methylated RNA in vivo.

RNA Binding

The binding of the MIWI-PAZ domain to methylated RNA was monitored by NMR spectroscopy. The 2D ^{15}N-^{1}H spectra of the MIWI-PAZ domain upon titration of a 6-mer ssRNA of sequence $^{5'}$CGACUU$^{3'}$ carrying a 2′-*O*-methyl group at the 3′-end are shown in Figure 1B. The NMR resonances of the bound and free protein forms are in fast-to-intermediate exchange, indicating a dissociation constant for the complex in the μM range. This number was verified using surface plasmon resonance and found to be 0.9 μM (Table 1, vide infra). ^{1}H-^{15}N chemical shift changes of MIWI-PAZ in the presence of 6-, 8-, or 15-mer RNAs with the same 3′-end sequence are virtually identical, indicating that only the last 6 nucleotides contribute to the binding. As in other PAZ-RNA complexes, the RNA binding site maps to the region of the cleft formed by the central β barrel and the α helix/β hairpin motif (Figure 2) (Lingel et al., 2004; Ma et al., 2004). As a comparison, we repeated the titration using a dsRNA with a 3′-end 2 nt overhang. Substantial chemical shift perturbations are shown by the same protein residues as in Figure 1B (Figure S2A), indicating that MIWI-PAZ recognizes ssRNA and dsRNA with a 2 nt overhang using the same surface.

Please cite this article in press as: Simon et al., Recognition of 2′-O-Methylated 3′-End of piRNA by the PAZ Domain of a Piwi Protein, Structure (2011), doi:10.1016/j.str.2010.11.015

Structure

Recognition of piRNA 2′-O-Methyl Mark by MIWI PAZ

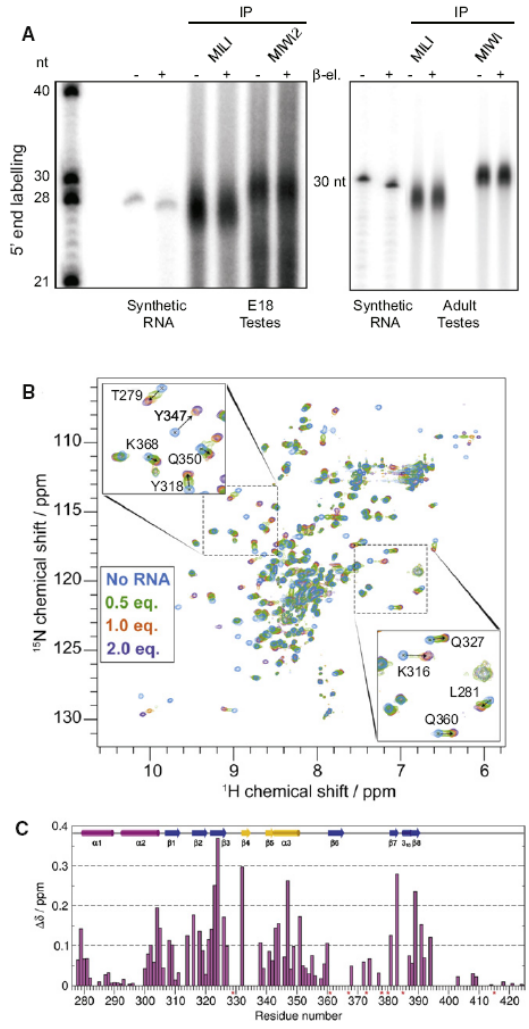


Figure 1. All Mouse Piwi Proteins Bind piRNAs with Modified 3′-Ends

(A) Piwi proteins were immunoprecipitated (IP) from adult or embryonic day 18 (E18) mouse testes and subjected to periodate oxidation and β-elimination (β-el). RNAs were visualized by 5′-end labeling. Nonmethylated synthetic RNAs were shortened by one nucleotide and migrate faster in a 15% urea-PAGE gel, while methylated piRNAs did not show any change in mobility. RNA markers with length in nucleotides (nt) are indicated. −, without β-el; +, with β-el.

(B) NMR 2D ^{15}N-^{1}H correlation spectra showing the chemical shift changes of the ^{1}H-^{15}N resonances of the MIWI-PAZ domain upon addition of a 6-mer ssRNA ($^{5'}$CGACUU$^{3'}$-2′-OMe).

(C) Chemical shift perturbation of ^{1}H-^{15}N resonances upon addition of two equivalents of the 6-mer RNA. Differences were calculated as $\sqrt{[(\Delta\delta_H)^2 + 0.2*(\Delta\delta_N)^2]}$. Prolines are indicated with asterisks. Most of the ^{1}H-^{15}N resonances of loop β3-β4, β4-β5, and β6-β7 could not be assigned either in the free or the bound protein. Additional missing data points are due to missing assignments in the bound form. A similar plot reporting on the chemical shift changes of the ^{1}H-^{15}N resonances of MIWI-PAZ upon addition of a 15-mer ssRNA and a 30-mer dsRNA with a 2 nt overhang is shown in Figure S2A. See also Figure S1.

We then used surface plasmon resonance (SPR) to determine the affinity of the wild-type MIWI-PAZ domain for RNAs with different secondary structures and modifications. Six RNA constructs were used, with 10, 2. and 1 nt overhangs, all with and without a 3′-end 2′-O-methyl group (Table 1). The MIWI-PAZ domain has a strong preference for long 3′-end overhangs, which mimic ssRNA for the purpose of binding. The affinity of MIWI-PAZ for ssRNA is 30 times stronger than for dsRNA with a 2 nt overhang, and it drops another 4-fold when comparing RNAs with 1 nt overhangs to those with 2 nt overhangs. Consistent with the fact that fly Piwi proteins are able to bind nonmethylated piRNAs in vivo (Horwich et al., 2007; Saito et al., 2007), the MIWI-PAZ domain shows only a modest preference for methylated RNA (2-fold), which is independent of the length of the overhangs.

Structure of the Mouse MIWI-PAZ Domain in Complex with 2′-O-Methylated ssRNA

The solution structure of the MIWI-PAZ domain in complex with an 8-mer RNA ($^{5'}$ACCGACUU$^{3'}$) 2′-O-methylated on the 3′-end uridine was determined by NMR (Table 2; see Experimental Procedures) (PDB accession code 2xfm). The MIWI-PAZ domain has the same fold as the PAZ domains of the Dicer and Ago proteins (Figures 2A and 2B; Figure S1). The core of the domain is formed by a short β barrel. It is flanked by two α helices on one side and an α helix and β hairpin on the other (Figure 2). The RNA binds in the cleft formed by the central β barrel and the α helix/β hairpin motif (Figures 2B and 2D). Overall, the position of the 2′-O-methyl group together with its sugar moiety is well defined by a total of 22 observed distances to the protein (Table 2; Table S1). Two additional hydrogen bonds between the 3′-hydroxyl and the 2′-methoxy groups of the terminal nucleotide to the carbonyl and amide groups of M382 were added in the final refinement step, as these groups were consistently found to be in close proximity in structures calculated without the hydrogen bond restraints. The remaining 12 RNA-protein and 44 RNA-RNA distances define the backbone trace of the RNA less accurately. There are no intermolecular NOEs involving the first two 5′-end nucleotides, in agreement with the proposal that a 6-mer RNA is sufficient to cover the binding site. Some of the protein side-chain resonances at the binding site (in particular, the three highly conserved, F333, Y346, and Y347 residues in the α helix/β hairpin motif) and the RNA resonances are exchange-broadened, which results in a poor definition of RNA nucleotides 3-7 in the complex (Table 2). The broadening of the RNA and protein resonances at the binding site can be either due to the low affinity of the RNA for the PAZ domain ($K_D = 0.9\ \mu M$), which can cause line broadening because of the exchange between the free and bound forms, or due to structural plasticity at the interface.

The 2′-O-methyl group at the 3′-end of the RNA is deeply inserted into the binding pocket and interacts with F333 on the β hairpin and M382 and A381 on strand β7 of the core β sheet (Figures 2B and 3A; Figure S3A). The long side chain of M382 bends from the back of strand β7 to contact the RNA methyl group. In agreement with this interaction, an M382A mutant

Please cite this article in press as: Simon et al., Recognition of 2′-O-Methylated 3′-End of piRNA by the PAZ Domain of a Piwi Protein, Structure (2011), doi:10.1016/j.str.2010.11.015

Table 1. Binding Affinities of MIWI-PAZ Domain and Different RNAsMIWI

PAZ Domain	10 nt Overhang	10 nt Overhang 2′-O-Methyl	2 nt Overhang	2 nt Overhang 2′-O-Methyl	1 nt Overhang	1 nt Overhang 2′-O-Methyl
WT MIWI-PAZ	2.0 ± 0.2	0.9 ± 0.1	55 ± 2	27 ± 1	>100	>100
M382A MIWI-PAZ	9.7 ± 0.3	2.0 ± 0.1	ND	ND	ND	ND
F343A MIWI-PAZ	15 ± 1	15 ± 1	ND	ND	ND	ND

SPR-derived dissociation constants of wild-type (WT) and mutant MIWI-PAZ domains for RNAs with overhangs of different lengths with and without 3′-end modification. The exact sequence of the RNAs is given in Experimental Procedures. K_Ds are given in μM. ND, not determined.

has a 2-fold reduced affinity for 3′-end methylated RNA (Table 1). However, a methionine at this position is not conserved across the Piwi proteins (Figure S1), indicating that this contact is not essential for methyl group recognition. Additional interactions are observed between the sugar protons of the terminal nucleotide and F333, A381, and L383, and between U8-H5′/H5″ and A381-Hβ (Figure 3A; Table S1). For the remaining five nucleotides of the RNA, we observe a number of contacts between RNA ribose rings and strand β2 of the central β barrel (T317, Y318, and R319) and the C terminus (L393) (Figures 2B and 3B; Figure S3B).

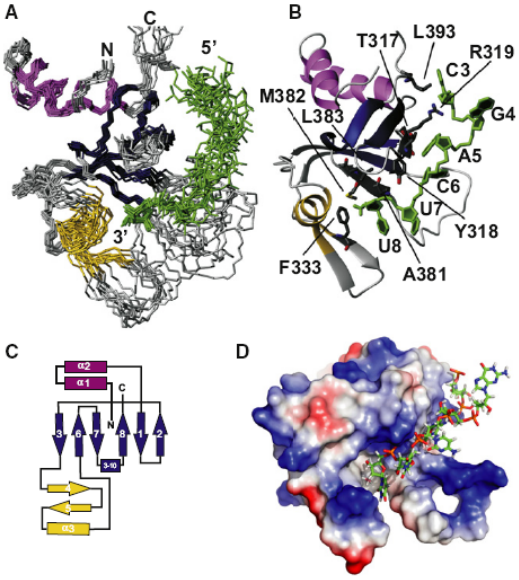


Figure 2. Overview of the Structure of the Complex between the MIWI-PAZ Domain and the 3′-End 2′-O-Methylated RNA $^{5'}$ACCGACUU$^{3'}$

(A) Overlap of the ten lowest energy NMR structures of the MIWI-PAZ-piRNA complex (PDB access code 2xfm). The PAZ domain is colored by structural motifs: α1, α2, magenta; central β-barrel β1, β2, β3, β6, β7, β8 blue; α helix/β hairpin motif α3, β4, β5, yellow. The color code for the protein secondary structure elements corresponds to that used in (C). The RNA is in green. The structural statistics are given in Table 2.

(B) Lowest energy structure of the ensemble shown in (A). Protein residues having NOEs with the RNA (Table S1) are shown in sticks.

(C) Schematic representation of the PAZ domain topology with numbering of the secondary structure elements.

(D) Surface representation of the MIWI-PAZ domain colored by electrostatic charge: white, neutral; blue, positive; red, negative. The RNA is colored by atom type: C green, N blue, O red, P orange, H white.

The second-last nucleotide shows no intermolecular and only very few intramolecular NOEs. The intermolecular contacts involving nucleotides 3–6, however, induce a stretched conformation in the phosphate backbone between U8 and U7. In comparison to the HsAGO1-PAZ-siRNA complex, the RNA appears to have slid up by approximately 1 nucleotide, with the second-last nucleotide in the MIWI-PAZ-piRNA complex approximately occupying the position of the third-last nucleotide in the HsAGO1-PAZ-siRNA structure. The protein surface covered by the 2 nt overhang in the HsAGO1-PAZ-siRNA complex is covered only by the last 3′-terminal nucleotide in the MIWI-PAZ-piRNA complex.

The NMR titration of the MIWI-PAZ domain with a siRNA mimic (dsRNA with a 2 nt overhang) indicates that the same protein surface is involved in recognition of the dsRNA (Figure S2A), but does not provide detailed information on the binding mode of MIWI-PAZ to dsRNA. The 30-fold higher affinity of MIWI-PAZ for ssRNA with respect to dsRNA with a 2 nt overhang confirms that this domain has evolved to optimize binding to ssRNA (Table 1). In contrast, HsAGO1 has been shown to bind dsRNA with a 2 nt overhang with higher affinity than longer overhangs (Ma et al., 2004). In agreement with this finding, trial structure calculations performed with a forced helical geometry for nucleotides 1–6 demonstrated that the intermolecular NOEs observed for MIWI-PAZ in complex with the 8-mer ssRNA are not compatible with dsRNA structure.

The conformational features of the piRNA on the MIWI-PAZ binding surface raise the question as to whether, when methylated at the 3′-end, RNAs with 1 nt overhangs can bind Piwi-PAZ domains with similar affinity as RNAs with 2 nt overhangs. To test this hypothesis, we used surface plasmon resonance (SPR) to determine the affinity of MIWI-PAZ for methylated and nonmethylated RNAs with 10 nt overhangs (mimicking ssRNA for the purpose of binding), and 2 and 1 nt overhangs (Table 1). The affinity of MIWI-PAZ for RNAs with 1 nt overhangs is poor and less than that for RNAs with 2 nt overhangs, irrespective of the methylation state.

DISCUSSION

The Overall Binding Topology of RNA to Piwi- and Ago-PAZ Domains Is Similar

The overall features of the 3′-end RNA recognition in the HsAGO1-PAZ-siRNA (or DmAGO2-PAZ-RNA) and in the

Please cite this article in press as: Simon et al., Recognition of 2′-O-Methylated 3′-End of piRNA by the PAZ Domain of a Piwi Protein, Structure (2011), doi:10.1016/j.str.2010.11.015

Structure

Recognition of piRNA 2′-O-Methyl Mark by MIWI PAZ

Table 2. NMR and Refinement Statistics for the MIWI-PAZ-piRNA Complex Structure

Distance constraints (protein/RNA/inter)	
Total NOE	2118 (2039/44/34)
Intraresidue	917/36
Interresidue	
Sequential ($\|i-j\|=1$)	380/8
Medium-range ($\|i-j\|<4$)	271
Long-range ($\|i-j\|>5$)	471
Hydrogen bonds (protein/inter)	14(12/2)
Total dihedral angle restraints	
ϕ	80
ψ	80
RNA (sugar pucker)	3x3
RNA (χ)	5
RNA (backbone)	45
Structure Statistics	
Deviations from idealized geometry	
Bond lengths (Å)	0.0038 ± 0.0001
Bond angles (°)	0.58 ± 0.01
Impropers (°)	1.38 ± 0.11
Violations (mean and SD)	
Distance constraints (Å)	0.022 ± 0.002
Dihedral angle constraints (°)	0.75 ± 0.08
Average pairwise RMS deviation** (Å)	
Heavy (complex)	1.71 ± 0.19
Backbone (complex)	1.07 ± 0.22
Heavy (protein + U8)	1.35 ± 0.17
Backbone (protein + U8)	0.81 ± 0.21
Heavy (protein only)	1.35 ± 0.16
Backbone (protein only)	0.80 ± 0.20
Heavy (RNA only)	2.66 ± 0.31
Backbone (RNA only)	1.57 ± 0.36
Ramachandran plot[a]	
Most favored	87.5% ± 1.4%
Additionally allowed	10.3% ± 1.5%
Generously allowed	1.6% ± 1.2%
Disallowed	0.7% ± 0.6%

[a] Average rmsd to the mean structure and Ramachandran plot were calculated for residues 2–90 and 105–118 of the protein and residues 3–8 of the RNA for the ten lowest energy structures, excluding unstructured termini and the long flexible β6-β7 loop (protein residues 91–104). The structures were refined in a shell of water molecules.

MIWI-PAZ-piRNA complexes are similar. In the HsAGO1-PAZ-siRNA structure, the 3′-end ribose displays two hydrogen bonds involving the RNA 3′-OH and 2′-OH and the CO and NH groups of Y336, respectively. The same hydrogen bonds can be formed in our structure between the RNA 3′-OH and 2′-O-methyl groups and the carbonyl and amide groups of M382. In both complexes, the last ribose of the RNA shows hydrophobic contacts with residues of the β barrel and the α helix/β hairpin insert (MIWI-PAZ-A381, L383, F333 corresponding to HsAGO1-T335, L337, F292) (Figure 3A; Figure S3A).

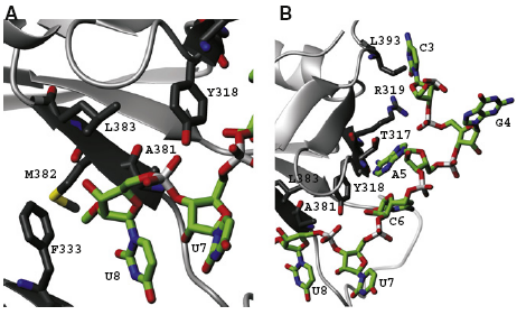


Figure 3. Molecular Contacts between MIWI-PAZ and RNA

(A) Binding cleft of the 3′-end 2′-O-methyl group. The methyl group is recognized by a hydrophobic cleft consisting of F333, A381, and M382.

(B) Hydrophobic and electrostatic contacts of the 5′-end of the piRNA with MIWI-PAZ. L393 (C-terminal tail) contacts the ribose of C3. R319 of sheet β2 is in place to form a H-bond with the phosphate of G4. A stereo-view of the two panels is given in Figure S3.

The backbone traces of the RNA on the central β barrel of the Piwi- and Ago-PAZ-RNA complexes are also similar (Figure S4). Two basic residues on β2 are highly conserved among all Piwi- and Ago-PAZ domains (HsAGO1-R275 and R278, MIWI-K316 and R319) (Figure S1). In MIWI, the side-chain protons of R319 display a number of NOEs to the C3-ribose; the arginine side chain is in most structures of the ensemble appropriately placed to form hydrogen bonds to the backbone phosphate of G4 (Figure 3B; Figure S3B). Correspondingly, the mutations DmAGO1-R355A (Yan et al., 2003) and HsAGO1-R278A (Ma et al., 2004) have been shown to reduce the RNA binding affinity by more than 10-fold.

In HsAGO1-PAZ, the C-terminal Q348, which is highly conserved across the Ago-PAZ family, establishes a hydrogen bond to the RNA phosphate backbone, whereas in MIWI-PAZ, a hydrophobic patch comprising L393 is found at this position. The side chain of L393 shows numerous contacts to the ribose ring of C3.

Comparison of PAZ Domain Structures

When superimposing MIWI- and HsAGO1-PAZ on the central β barrel, we observe an upward shift of the α helix/β hairpin insert in MIWI-PAZ by approximately one turn of helix α3 (Figure 4A). The observed shift of the insert can be rationalized by sequence analysis of the eukaryotic PAZ domain families. All Piwi-PAZ domains have an aromatic residue (Phe or Tyr) at the position of MIWI-F343 and a long aliphatic residue (Ile, Leu, or Met) at the position of MIWI-I384. By contrast, Ago-PAZ domains mostly feature a valine (HsAGO1-V306) at the position of the aromatic amino acid in α3 (MIWI-F343) and a conserved proline (HsAGO1-P338) at the position of the long aliphatic chain in β7 (MIWI-I384) (Figures 4A and 4B; Figures S1 and S5). The presence of a bulky aromatic side chain at position 343 would not be compatible with the relative arrangement of the α helix/β hairpin insert and the core β barrel in HsAGO1-PAZ. In MIWI, favorable packing is achieved between the two structural elements in a different fashion, involving contacts between

Please cite this article in press as: Simon et al., Recognition of 2′-O-Methylated 3′-End of piRNA by the PAZ Domain of a Piwi Protein, Structure (2011), doi:10.1016/j.str.2010.11.015

Structure

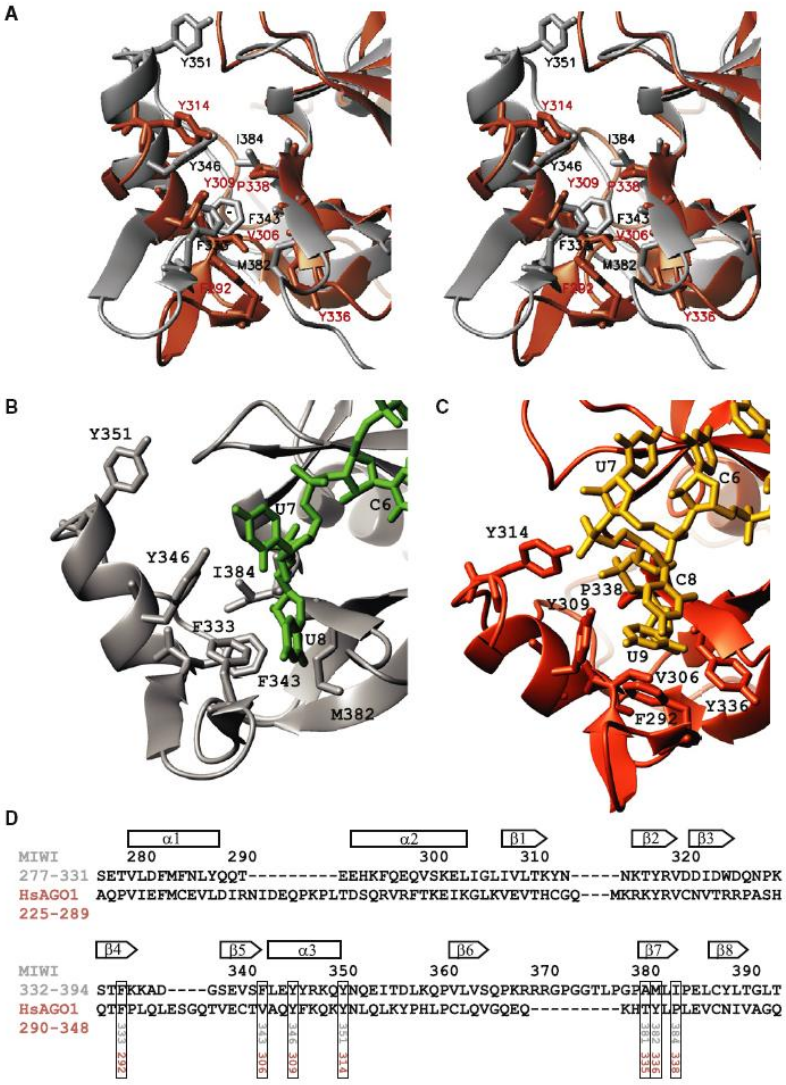


Figure 4. Comparison of RNA Recognition in MIWI and HsAGO1

The structures are superimposed on the Cα atoms of the core β barrel (sheets β1–3, 6–8) with a coordinate root-mean-square deviation of 0.67 Å.

(A) Stereo-view of the RNA binding pocket with important residues (see text) highlighted. Helix α3 moves out and up by about one helix turn in MIWI (protein gray, RNA green) compared with HsAGO1 (protein red, RNA gold).

(B and C) RNA 3′-end in the PAZ binding pocket of MIWI-PAZ (B) and HsAGO1 (C). In the MIWI-PAZ-piRNA complex (B), the RNA backbone between the ultimate and penultimate nucleotide is extended; the RNA register is shifted by approximately 1 nucleotide along the binding surface of the protein compared to the HsAGO1-PAZ-siRNA complex (C).

(D) Sequence alignment of MIWI and human AGO1-PAZ domains indicating structural motifs and key amino acid residues forming the 3′-end binding cleft.

See also Figures S4–S6.

F343 and I384 (Figure 4A; Figure S5). The MIWI-F343A mutant shows a 6-fold reduced affinity for 2′-O-methylated RNAs compared with wild-type MIWI indicating that the tight packing of the α helix/β hairpin insert and β7 of the core β barrel is important to ensure optimal coordination of the methyl group (Table 1).

Please cite this article in press as: Simon et al., Recognition of 2′-*O*-Methylated 3′-End of piRNA by the PAZ Domain of a Piwi Protein, Structure (2011), doi:10.1016/j.str.2010.11.015

Structure

Recognition of piRNA 2′-*O*-Methyl Mark by MIWI PAZ

In DmAGO2, I678 occupies the position of HsAGO1-V306, while the residue at the position of HsAGO1-P338 is conserved. The longer side chain in the α helix/β hairpin insert is accommodated by a slight sideways shift of this element with respect to the β barrel core (Figure S6). In contrast, in Dicer-PAZ domains (Macrae et al., 2006), which feature an aromatic residue on α3 at the position of MIWI-F343 or HsAGO1-V306, but share the proline on β7 with the Ago-PAZ sequence (HsAGO1-P338, MIWI-I384), the α3 helix is positioned in between those of hsAGO1 and MIWI with respect to the β barrel core (Figure S5).

Consequences of Structural Differences between PAZ Domains for RNA Binding

As a result of the shift of the α helix/β hairpin insert with respect to the β barrel core in MIWI, the base of the RNA binding pocket is slightly wider than in HsAGO1-PAZ, where F292 closes the pocket from the bottom; the corresponding side chain (F333) in MIWI moves up with respect to the central β barrel and contacts the RNA from the side. MIWI-Y346 takes the place of HsAGO1-Y314 instead of HsAGO1-Y309, as would be inferred from sequence alignment (Figure 4A; Figures S1 and S5). The widening of the RNA binding cleft in MIWI nicely accommodates the 3′-end 2′-*O*-methyl group, which would partly clash with F292 in HsAGO1 (Figure S6). These structural features are reflected in the strong preference of hs-AGO1 for nonmethylated RNA (Ma et al., 2004), while MIWI shows a modest preference for 3′-end methylated RNA (Table 1). In DmAGO2, which binds methylated RNA in vivo (Ameres et al., 2010), the replacement of V306 by I678 (Figure S1) and the resultant movement of the α helix/β hairpin insert away from the core β barrel (Figure S6) creates more space for an additional bulky methyl-group on the 2′-OH of the terminal sugar. It is interesting to note that other Ago-PAZ domains with a longer aliphatic side chain in the same position are those from organisms with 3′-end methylated si/miRNAs, including those from *Arabidopsis thaliana*.

Overall, the structure and analysis presented here reveal a delicate dependence of the structural details of PAZ domains belonging to different families on conserved sequence elements. These subtle structural changes can be correlated to the ability of PAZ domains to adapt to RNA substrates with different secondary structure, sequence and chemical modifications.

EXPERIMENTAL PROCEDURES

Sample Preparation

The PAZ domain (residues 276–425) of MIWI (Mouse Piwi-like protein 1, UniProt code Q9JMB7) was cloned into the pETM-11 vector (EMBL collection) between restriction sites NcoI and NotI, yielding an expression construct with an N-terminal His-tag separated from the PAZ sequence by a TEV protease cleavage site. This vector was transformed into *Escherichia coli* BL21 Rosetta 2 cells (Novagen) for protein expression.

Cell culture was performed using M9 minimal medium supplemented with 30 mg/ml^{-1} chloramphenicol and 50 mg/ml^{-1} kanamycin, and containing ^{15}N-labeled ammonium chloride and ^{13}C-labeled glucose as appropriate. After initial growth at 37°C to an optical density of 0.6, protein expression was induced by addition of 1 mM IPTG, and the temperature reduced to 20°C. Incubation was continued for ~20 hr before harvesting the cells by centrifugation.

Cells were resuspended in buffer A (20 mM Tris-HCl, 10 mM imidazole, 300 mM NaCl [pH 7.5]) containing 0.2 mg/ml lysozyme, 2 U ml^{-1} RNase-free DNase (Promega), 10 mM β-mercaptoethanol and EDTA-free protease inhibitors (Roche), and lysed by French press. After centrifugation, the cell lysate was loaded onto a His-TRAP FF column (GE Healthcare) and washed extensively in buffer A, buffer B (20 mM Tris-HCl, 40 mM imidazole, 300 mM NaCl [pH 7.5]), and buffer C (20 mM Tris-HCl, 10 mM imidazole, 1 M NaCl [pH 7.5]) prior to elution in buffer D (20 mM Tris-HCl, 250 mM imidazole, 300 mM NaCl [pH 7.5]). After exchange into buffer A, the eluate was reloaded onto the column, and the washing and elution steps repeated. The N-terminal His-tag was cleaved by overnight incubation at 4°C in buffer E (50 mM Na_2HPO_4, 10 mM imidazole, 150 mM NaCl [pH 7.4]) containing TEV protease. The reaction mix was exchanged into buffer A and uncleaved protein removed by reverse purification with the His-TRAP FF column. The purified PAZ protein was exchanged into buffer F (50 mM Na_2HPO_4, 150 mM NaCl [pH 7.4]) for NMR studies.

The 8-mer RNA ($^{5'}$ACCGACUU$^{3'}$) with 2′-*O*-methylation of the 3′-end uridine was obtained from IBA (Goettingen, Germany). For preparation of the PAZ-RNA complex, the RNA was dissolved in a small volume of buffer F, heated for 5 min at 70°C and after slowly cooling to room temperature added to the PAZ protein. Superase (Ambion) was added to the sample (1 U/ml) to inhibit residual RNase activity. Samples were prepared with a 2-fold molar ratio of RNA to protein. Protein concentrations for double-labeled samples were 0.20–0.40 mM.

NMR Experiments

NMR data were collected on Bruker Avance III 600 MHz and 800 MHz spectrometers, equipped with HCN triple-resonance cryo-probes. Spectra were acquired at temperatures of 300 K and 310 K. Protein backbone assignments for free PAZ and the PAZ-RNA complex were obtained from a combination of HNCO, HNCA (Grzesiek and Bax, 1992; Schleucher et al., 1993; Yamazaki et al., 1994), HNCACB (Muhandiram and Kay, 1994; Wittekind and Mueller, 1993), and HNCOCACB (Yamazaki et al., 1994) out-and-back triple-resonance experiments. Protein side-chain resonances in both free PAZ and the PAZ-RNA complex were assigned from (H)CCH-TOCSY and H(C)CH-TOCSY (Kay et al., 1993) spectra, and amide-detected (H)CC(CO)NH-TOCSY and H(CCCO)NH-TOCSY (Grzesiek et al., 1993; Montelione et al., 1992) experiments. For free PAZ, (HB)CB(CD)HD and (HB)CB(CDCE)HE spectra (Yamazaki et al., 1993) were acquired to aid assignment of aromatic protein resonances. RNA resonances in the PAZ-RNA complex were assigned from double-^{12}C,^{14}N-filtered 2D NOESY and TOCSY spectra (Zwahlen et al., 1997).

Secondary structure elements were identified by analysis of the chemical shift index (Wishart and Sykes, 1994) and NOE cross peak pattern. Distance constraints were collected from 3D ^{15}N-NOESY (τ_m = 100 ms), 3D ^{13}C-edited NOESY (τ_m = 100 ms), 3D ^{13}C-edited/filtered NOESY (τ_m = 150 ms), and 2D ^{12}C- filtered NOESY spectra (Zwahlen et al., 1997).

Data were processed with NMRPipe (Delaglio et al., 1995) and analyzed using NMRView (Johnson and Blevins, 1994).

Structure Calculation and Refinement

The experimentally determined distance and dihedral angle restraints were applied in a simulated annealing protocol with CNS/ARIA1.2 (Linge et al., 2003a). All NOEs were manually assigned. The CNS E_{repel} function was used to simulate van der Waals interactions with an energy constant of 25.0 kcal mol^{-1} Å^{-4} using "PROLSQ" van der Waals radii (Linge and Nilges, 1999). Distance restraints were employed with a soft square-well potential using an energy constant of 50 kcal mol^{-1} Å^{-2}. For hydrogen bonds, distance restraints with bounds of 1.8–2.3 Å (H-O), and 2.8–3.3 Å (N-O) were imposed for slowly exchanging amide protons. Dihedral angle restraints derived from TALOS (Cornilescu et al., 1999) were applied to φ and ψ backbone angles using energy constants of 200 kcal mol^{-1} rad^{-2}. On the RNA, dihedral angle restraints were applied for 5 χ angles (A1, C3, G4, A5, C6), as derived from the analysis of the NOEs. For the three nucleotides where the χ angle assumes the *syn* conformation (A1, G4, A5), the ribose was additionally restrained to the *C3′-endo* conformation. All RNA backbone dihedrals were restrained to a broad range of allowed conformations (α, ζ ∈ [25°,335°], β ∈ [50°,290°], γ ∈ [20°,335°], δ ∈ [55°,175°], ε ∈ [155°,310°]). For γ, a conformational energy term was added to the force-field, which restricts the dihedral angle to ranges around three minima, according to E_{dih} ~(1 + cos(3*γ+18)). The final ensemble of NMR structures was refined in a shell of water molecules (Linge et al., 2003b; Nozinovic et al., 2010). The statistics for an ensemble of 10 lowest-energy solution structures (out of 100 calculated) is reported in Table 2. Structural quality was analyzed using PROCHECK (Laskowski et al., 1993).

Please cite this article in press as: Simon et al., Recognition of 2′-O-Methylated 3′-End of piRNA by the PAZ Domain of a Piwi Protein, Structure (2011), doi:10.1016/j.str.2010.11.015

Structure

Recognition of piRNA 2′-O-Methyl Mark by MIWI PAZ

Immunoprecipitations and β-Eliminations

Piwi immunoprecipitations and β-eliminations were performed as previously described (Kirino and Mourelatos, 2007; Reuter et al., 2009).

Surface Plasmon Resonance

Surface plasmon resonance (SPR) experiments were performed at 25°C on a Biacore T100 machine (Biacore AB, Uppsala, Sweden). The RNA constructs were designed to consist of a duplex with biotin attached to the 3′-end of one strand and 3′-extensions of different length on the complementary strand leading to 1, 2, or 10 nt overhangs, which were either nonmethylated or 2′-O-methylated at the ribose of the 3′-end. The RNA sequences are (the nucleotides in bold indicate the overhangs): 10 nt overhang, 5′ UGACAUGAACACA GG UGC UC**AGAUAGCUUU** 3′; 2 nt overhang, 5′ UGACAUGAACACAGGUG CGC**UU** 3′; 1 nt overhang, 5′ UGACAUGAACACAGG UGCGC **U** 3′. Each 25 μM biotinylated RNA and complementary strands with different overhangs were incubated for 2 min at 90°C and cooled to room temperature over 45 min in 20 mM Tris (pH 7.5), 100 mM NaCl, 2 mM EDTA to form duplexes. The annealed dsRNA was subsequently diluted to 25 nM in Biacore running buffer (PBS [pH 7.4], 0.05% Tween 20, 1 mM DTT) and immobilized on a strepavidin-coated biosensor chip. The nonmodified sensor surface served as reference.

Two-fold serial dilutions of wild-type and mutant PAZ proteins ranging from 100 to 0.098 μM were injected as analytes at a flow rate of 30 μl/min and the equilibrium dissociation constants determined by steady-state affinity evaluation.

Alignment of PAZ Domains

We collected 40 eukaryotic protein sequences predicted or known to contain the PAZ domain according to the annotations of their records in the NCBI's Entrez database (Sayers et al., 2010) and domain predictions in the PFAM (Finn et al., 2010) and CDD databases (Marchler-Bauer et al., 2009) from *Homo sapiens* (9 sequences), *Mus musculus* (8), *Drosophila melanogaster* (7), *Schizosaccharomyces pombe* (1), *Arabidopsis thaliana* (14), and *Giardia lamblia* (1). The sequences of their PAZ domains were taken according to the positions given by the match to the CDD database, aligned, and then extended or shortened according to the alignment (done with ClustalW; Thompson et al., 1994)) in an iterative process. The alignment was supported by data from known 3D structures and secondary structure predictions (using JPRED3; Cole et al., 2008). The resulting alignment was used to build a phylogenetic tree in ClustalW.

ACCESSION NUMBERS

Coordinates and NMR data have been deposited in the Protein Data Bank with the accession code 2xfm.

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and one table and can be found with this article online at doi:10.1016/j.str.2010.11.015.

ACKNOWLEDGMENTS

We thank Maartje Luteijn and Jerome Bourdin for plasmids. We thank the EMBL Protein Expression and Purification Core Facility, and Dr. Magdalena Rakwalska for support in the production of proteins for the SPR experiments. This work was supported by the European Molecular Biology Laboratory and funding from Agence National de la Recherche (ANR; "piRmachines"), France to R.S.P. The authors declare that they have no competing financial interests.

Received: July 29, 2010
Revised: November 19, 2010
Accepted: November 21, 2010
Published online: January 13, 2011

REFERENCES

Ameres, S.L., Horwich, M.D., Hung, J.-H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010). Target RNA–directed trimming and tailing of small silencing RNAs. Science *328*, 1534–1539.

Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science *318*, 761–764.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol. Cell *31*, 785–799.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089–1103.

Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Res. *36*, W197–W201.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR *13*, 289–302.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPIPE—a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. *9*, 102–114.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., et al. (2010). The Pfam protein families database. Nucleic Acids Res. *38*, D211–D222.

Frank, F., Sonenberg, N., and Nagar, B. (2010). Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. Nature *465*, 818–822.

Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nature Rev. *10*, 94–108.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199–202.

Grzesiek, S., and Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31-kDa protein. J. Magn. Reson. *96*, 432–440.

Grzesiek, S., Anglister, J., and Bax, A. (1993). Correlation of backbone amides and aliphatic side-chain resonances in C-13/N-15-enriched proteins by isotropic mixing of C-13 magnetization. J. Magn. Reson. B. *101*, 114–119.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. Science *315*, 1587–1590.

Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr. Biol. *17*, 1265–1272.

Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B., et al. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. Cell *129*, 69–82.

Jinek, M., and Doudna, J.A. (2009). A three-dimensional view of the molecular machinery of RNA interference. Nature *457*, 405–412.

Johnson, B.A., and Blevins, R.A. (1994). NMR VIEW—a computer program for the visualization and analysis of NMR data. J. Biomol. NMR *4*, 603–614.

Kay, L.E., Xu, G.Y., Singer, A.U., Muhandiram, D.R., and Formankay, J.D. (1993). A gradient-enhanced HCCH-TOCSY experiment for recording side-chain H-1 and C-13 correlations in H2O samples of proteins. J. Magn. Reson. B. *101*, 333–337.

Kirino, Y., and Mourelatos, Z. (2007). Mouse Piwi-interacting RNAs are 2′-O-methylated at their 3′ termini. Nat. Struct. Mol. Biol. *14*, 347–348.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., 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.

Kurth, H.M., and Mochizuki, K. (2009). 2′-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in Tetrahymena. RNA *15*, 675–685.

Please cite this article in press as: Simon et al., Recognition of 2′-O-Methylated 3′-End of piRNA by the PAZ Domain of a Piwi Protein, Structure (2011), doi:10.1016/j.str.2010.11.015

Structure

Recognition of piRNA 2′-O-Methyl Mark by MIWI PAZ

Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK—a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in Arabidopsis. Curr. Biol. *15*, 1501–1507.

Linge, J.P., and Nilges, M. (1999). Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation. J. Biomol. NMR *13*, 51–59.

Linge, J.P., Habeck, M., Rieping, W., and Nilges, M. (2003a). ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics *19*, 315–316.

Linge, J.P., Williams, M.A., Spronk, C., Bonvin, A., and Nilges, M. (2003b). Refinement of protein structures in explicit solvent. Proteins: Struct. Funct. Bioinf. *50*, 496–506.

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2004). Nucleic acid 3′-end recognition by the Argonaute2 PAZ domain. Nat. Struct. Mol. Biol. *11*, 576–577.

Ma, J.B., Ye, K., and Patel, D.J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature *429*, 318–322.

Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T., and Patel, D.J. (2005). Structural basis for 5′-end-specific recognition of guide RNA by the A-fulgidus Piwi protein. Nature *434*, 666–670.

Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double-stranded RNA processing by Dicer. Science *311*, 195–198.

Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., et al. (2009). CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Res. *37*, D205–D210.

Montelione, G.T., Lyons, B.A., Emerson, S.D., and Tashiro, M. (1992). An efficient triple resonance experiment using C-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically enriched proteins. J. Am. Chem. Soc. *114*, 10974–10975.

Muhandiram, D.R., and Kay, L.E. (1994). Gradient-enhanced triple-resonance 3-dimensional NMR experiments with improved sensitivity. J. Magn. Reson. B. *103*, 203–216.

Nozinovic, S., Furtig, B., Jonker, H.R.A., Richter, C., and Schwalbe, H. (2010). High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. Nucleic Acids Res. *38*, 683–694.

Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H., Miyauchi, K., and Suzuki, T. (2007). The 3′ termini of mouse Piwi-interacting RNAs are 2′-O-methylated. Nat. Struct. Mol. Biol. *14*, 349–350.

Parker, J.S., Roe, S.M., and Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature *434*, 663–666.

Parker, J.S., Roe, S.M., and Barford, D. (2006). Molecular mechanism of target RNA transcript recognition by Argonaute-guide complexes. Cold Spring Harb. Symp. Quant. Biol. *71*, 45–50.

Patel, D.J., Ma, J.B., Yuan, Y.R., Ye, K., Pei, Y., Kuryavyi, V., Malinina, L., Meister, G., and Tuschl, T. (2006a). Structural biology of RNA silencing and its functional implications. Cold Spring Harb. Symp. Quant. Biol. *71*, 81–93.

Patel, D.J., Ma, J.B., Yuan, Y.R., Ye, K., Pei, Y., Kuryavyi, V., Malinina, L., Meister, G., and Tuschl, T. (2006b). Structural biology of RNA silencing and its functional implications. Cold Spring Harb. Symp. Quant. Biol. *71*, 81–93.

Rashid, U.J., Paterok, D., Koglin, A., Gohlke, H., Piehler, J., and Chen, J.C.H. (2007). Structure of Aquifex aeolicus Argonaute highlights conformational flexibility of the PAZ domain as a potential regulator of RNA-induced silencing complex function. J. Biol. Chem. *282*, 13824–13832.

Reuter, M., Chuma, S., Tanaka, T., Franz, T., Stark, A., and Pillai, R.S. (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., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2′-O-methylation of Piwi- interacting RNAs at their 3′ ends. Genes Dev. *21*, 1603–1608.

Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Federhen, S., et al. (2010). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. *38*, D5–D16.

Schleucher, J., Sattler, M., and Griesinger, C. (1993). Coherence selection by gradients without signal attenuation—application to the 3-dimensional HNCO experiment. Angew. Chem. *32*, 1489–1491.

Song, J.J., Liu, J., Tolia, N.H., Schneiderman, J., Smith, S.K., Martienssen, R.A., Hannon, G.J., and 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.

Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science *305*, 1434–1437.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

Tolia, N.H., and Joshua-Tor, L. (2007). Slicer and the argonautes. Nat. Chem. Biol. *3*, 36–43.

Tomari, Y., Du, T., and Zamore, P.D. (2007). Sorting of Drosophila small silencing RNAs. Cell *130*, 299–308.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320–324.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672–676.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., and Patel, D.J. (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature *461*, 754–761.

Wishart, D.S., and Sykes, B.D. (1994). The C-13 chemical shift index—a simple method for the identification of protein secondary structure using C-13 chemical shift data. J. Biomol. NMR *4*, 171–180.

Wittekind, M., and Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins. J. Magn. Reson. B. *101*, 201–205.

Yamazaki, T., Forman-Kay, J.D., and Kay, L.E. (1993). 2-dimensional NMR experiments for correlating C-13-beta and H-1-delta/epsilon chemical shifts of aromatic residues in C-13-labeled proteins via scalar couplings. J. Am. Chem. Soc. *115*, 11054–11055.

Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R., and Kay, L.E. (1994). A suite of triple-resonance NMR experiments for the backbone assignment of N-15, C-13, H-2 labeled proteins with high sensitivity. J. Am. Chem. Soc. *116*, 11655–11666.

Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M.M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature *426*, 468–474.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. Science *307*, 932–935.

Yuan, Y.R., Pei, Y., Chen, H.Y., Tuschl, T., and Patel, D.J. (2006). A potential protein-RNA recognition event along the RISC-loading pathway from the structure of A. aeolicus Argonaute with externally bound siRNA. Structure *14*, 1557–1565.

Yuan, Y.R., Pei, Y., Ma, J.B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.Y., Dauter, Z., Tuschl, T., and Patel, D.J. (2005). Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Mol. Cell *19*, 405–419.

Zwahlen, C., Legault, P., Vincent, S.J.F., Greenblatt, J., Konrat, R., and Kay, L.E. (1997). Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. J. Am. Chem. Soc. *119*, 6711–6721.

References

Abdel-Ghany, S.E. & Pilon, M. (2008) MicroRNA-mediated systemic down-regulation of copper protein expression in response to low copper availability in Arabidopsis. *J Biol Chem* **283**, 15932-15945.

Abdellah (2004) Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945.

Akbergenov, R., Si-Ammour, A., Blevins, T., *et al.* (2006) Molecular characterization of geminivirus-derived small RNAs in different plant species. *Nucleic acids research* **34**, 462-471.

Allen, E., Xie, Z., Gustafson, A.M. & Carrington, J.C. (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**, 207-221.

Ameres, S.L., Horwich, M.D., Hung, J.H., *et al.* (2010) Target RNA-directed trimming and tailing of small silencing RNAs. *Science* **328**, 1534-1539.

Ameres, S.L., Martinez, J. & Schroeder, R. (2007) Molecular basis for target RNA recognition and cleavage by human RISC. *Cell* **130**, 101-112.

Anantharaman, V., Koonin, E.V. & Aravind, L. (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic acids research* **30**, 1427-1464.

Ancelin, K., Lange, U.C., Hajkova, P., *et al.* (2006) Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nat Cell Biol* **8**, 623-630.

Anne, J. & Mechler, B.M. (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**, 2167-2177.

Aoki, Y., Nagao, I., Saito, D., Ebe, Y., Kinjo, M. & Tanaka, M. (2008) Temporal and spatial localization of three germline-specific proteins in medaka. *Dev Dyn* **237**, 800-807.

Aravin, A., Gaidatzis, D., Pfeffer, S., *et al.* (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203-207.

Aravin, A.A. & Bourc'his, D. (2008) Small RNA guides for de novo DNA methylation in mammalian germ cells. *Genes Dev* **22**, 970-975.

Aravin, A.A., Hannon, G.J. & Brennecke, J. (2007a) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**, 761-764.

Aravin, A.A., Klenov, M.S., Vagin, V.V., Bantignies, F., Cavalli, G. & Gvozdev, V.A. (2004) Dissection of a natural RNA silencing process in the Drosophila melanogaster germ line. *Mol Cell Biol* **24**, 6742-6750.

Aravin, A.A., Lagos-Quintana, M., Yalcin, A., *et al.* (2003) The small RNA profile during Drosophila melanogaster development. *Dev Cell* **5**, 337-350.

Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M. & Gvozdev, V.A. (2001) Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Curr Biol* **11**, 1017-1027.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., *et al.* (2008) A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31**, 785-799.

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G.J. (2007b) Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**, 744-747.

Aravin, A.A., van der Heijden, G.W., Castaneda, J., Vagin, V.V., Hannon, G.J. & Bortvin, A. (2009) Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. *PLoS Genet* **5**, e1000764.

Arkov, A.L., Wang, J.Y., Ramos, A. & Lehmann, R. (2006) The role of Tudor domains in germline development and polar granule architecture. *Development* **133**, 4053-4062.

Aza-Blanc, P., Cooper, C.L., Wagner, K., Batalov, S., Deveraux, Q.L. & Cooke, M.P. (2003) Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. *Mol Cell* **12**, 627-637.

Azuma-Mukai, A., Oguri, H., Mituyama, T., *et al.* (2008) Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Proc Natl Acad Sci U S A* **105**, 7964-7969.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P. & Blelloch, R. (2008) Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev* **22**, 2773-2785.

Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P. & Bartel, D.P. (2008) The impact of microRNAs on protein output. *Nature* **455**, 64-71.

Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-233.

Baumberger, N. & Baulcombe, D.C. (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci U S A* **102**, 11928-11933.

Becalska, A.N. & Gavis, E.R. (2009) Lighting up mRNA localization in Drosophila oogenesis. *Development* **136**, 2493-2503.

Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P. & Izaurralde, E. (2006) mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev* **20**, 1885-1898.

Berezikov, E., Chung, W.J., Willis, J., Cuppen, E. & Lai, E.C. (2007) Mammalian mirtron genes. *Mol Cell* **28**, 328-336.

Berger, I., Fitzgerald, D.J. & Richmond, T.J. (2004) Baculovirus expression system for heterologous multiprotein complexes. *Nature biotechnology* **22**, 1583-1587.

Bernstein, E., Caudy, A.A., Hammond, S.M. & Hannon, G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363-366.

Biemont, C. & Vieira, C. (2006) Genetics: junk DNA as an evolutionary force. *Nature* **443**, 521-524.

Bieniossek, C., Richmond, T.J. & Berger, I. (2008) MultiBac: multigene baculovirus-based eukaryotic protein complex production. *Current protocols in protein science / editorial board, John E. Coligan ... [et al* **Chapter 5**, Unit 5 20.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M. & Benning, C. (1998) AGO1 defines a novel locus of Arabidopsis controlling leaf development. *Embo J* **17**, 170-180.

Bohnsack, M.T., Czaplinski, K. & Gorlich, D. (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *Rna* **10**, 185-191.

Boland, A., Tritschler, F., Heimstadt, S., Izaurralde, E. & Weichenrieder, O. (2010) Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO Rep* **11**, 522-527.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R. & Zhu, J.K. (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* **123**, 1279-1291.

Boswell, R.E. & Mahowald, A.P. (1985) tudor, a gene required for assembly of the germ plasm in Drosophila melanogaster. *Cell* **43**, 97-104.

Bourc'his, D. & Bestor, T.H. (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96-99.

Boutet, S., Vazquez, F., Liu, J., *et al.* (2003) Arabidopsis HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. *Curr Biol* **13**, 843-848.

Bozzetti, M.P., Massari, S., Finelli, P., *et al.* (1995) The Ste locus, a component of the parasitic cry-Ste system of Drosophila melanogaster, encodes a protein that forms crystals in primary spermatocytes and mimics properties of the beta subunit of casein kinase 2. *Proc Natl Acad Sci U S A* **92**, 6067-6071.

Brahms, H., Raymackers, J., Union, A., de Keyser, F., Meheus, L. & Luhrmann, R. (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**, 17122-17129.

Brennecke, J., Aravin, A.A., Stark, A., *et al.* (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**, 1089-1103.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B. & Cohen, S.M. (2003) bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**, 25-36.

Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A. & Hannon, G.J. (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**, 1387-1392.

Brennecke, J., Stark, A., Russell, R.B. & Cohen, S.M. (2005) Principles of microRNA-target recognition. *PLoS biology* **3**, e85.

Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., *et al.* (2008) Widespread translational inhibition by plant miRNAs and siRNAs. *Science* **320**, 1185-1190.

Brower-Toland, B., Findley, S.D., Jiang, L., *et al.* (2007) Drosophila PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev* **21**, 2300-2311.

Buhler, D., Raker, V., Luhrmann, R. & Fischer, U. (1999) Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy. *Hum Mol Genet* **8**, 2351-2357.

Buhler, M. & Moazed, D. (2007) Transcription and RNAi in heterochromatic gene silencing. *Nat Struct Mol Biol* **14**, 1041-1048.

Buhler, M., Verdel, A. & Moazed, D. (2006) Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**, 873-886.

Buhtz, A., Pieritz, J., Springer, F. & Kehr, J. (2010) Phloem small RNAs, nutrient stress responses, and systemic mobility. *BMC plant biology* **10**, 64.

Bulgakov, O.V., Eggenschwiler, J.T., Hong, D.H., Anderson, K.V. & Li, T. (2004) FKBP8 is a negative regulator of mouse sonic hedgehog signaling in neural tissues. *Development* **131**, 2149-2159.

Cai, X., Hagedorn, C.H. & Cullen, B.R. (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna* **10**, 1957-1966.

Carmell, M.A., Girard, A., van de Kant, H.J., *et al.* (2007) MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* **12**, 503-514.

Catalanotto, C., Azzalin, G., Macino, G. & Cogoni, C. (2002) Involvement of small RNAs and role of the qde genes in the gene silencing pathway in Neurospora. *Genes Dev* **16**, 790-795.

Cavey, M., Hijal, S., Zhang, X. & Suter, B. (2005) Drosophila valois encodes a divergent WD protein that is required for Vasa localization and Oskar protein accumulation. *Development* **132**, 459-468.

Chan, C.M., Zhou, C. & Huang, R.H. (2009) Reconstituting bacterial RNA repair and modification in vitro. *Science* **326**, 247.

Chan, S.W. (2008) Inputs and outputs for chromatin-targeted RNAi. *Trends in plant science* **13**, 383-389.

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C. & Jacobsen, S.E. (2004) RNA silencing genes control de novo DNA methylation. *Science* **303**, 1336.

Cheloufi, S., Dos Santos, C.O., Chong, M.M. & Hannon, G.J. (2010) A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465**, 584-589.

Chen, C., Jin, J., James, D.A., *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**, 20336-20341.

Chen, X. (2005) MicroRNA biogenesis and function in plants. *FEBS letters* **579**, 5923-5931.

Chen, X. (2007) A marked end. *Nat Struct Mol Biol* **14**, 259-260.

Chen, X., Liu, J., Cheng, Y. & Jia, D. (2002) HEN1 functions pleiotropically in Arabidopsis development and acts in C function in the flower. *Development* **129**, 1085-1094.

Chendrimada, T.P., Finn, K.J., Ji, X., *et al.* (2007) MicroRNA silencing through RISC recruitment of eIF6. *Nature* **447**, 823-828.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., *et al.* (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740-744.

Chenna, R., Sugawara, H., Koike, T., *et al.* (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic acids research* **31**, 3497-3500.

Choudhuri, S. (2009) Lesser known relatives of miRNA. *Biochem Biophys Res Commun* **388**, 177-180.

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.

Chuma, S., Hosokawa, M., Kitamura, K., *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**, 15894-15899.

Chung, W.J., Okamura, K., Martin, R. & Lai, E.C. (2008) Endogenous RNA interference provides a somatic defense against Drosophila transposons. *Curr Biol* **18**, 795-802.

Cifuentes, D., Xue, H., Taylor, D.W., *et al.* (2010) A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science* **328**, 1694-1698.

Cole, C., Sobala, A., Lu, C., *et al.* (2009) Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *Rna* **15**, 2147-2160.

Cook, H.A., Koppetsch, B.S., Wu, J. & Theurkauf, W.E. (2004) The Drosophila SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. *Cell* **116**, 817-829.

Costa, Y., Speed, R.M., Gautier, P., *et al.* (2006) Mouse MAELSTROM: the link between meiotic silencing of unsynapsed chromatin and microRNA pathway? *Human molecular genetics* **15**, 2324-2334.

Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D. & Lin, H. (1998) A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev* **12**, 3715-3727.

Crackower, M.A., Kolas, N.K., Noguchi, J., *et al.* (2003) Essential role of Fkbp6 in male fertility and homologous chromosome pairing in meiosis. *Science* **300**, 1291-1295.

Czech, B., Malone, C.D., Zhou, R., *et al.* (2008) An endogenous small interfering RNA pathway in Drosophila. *Nature* **453**, 798-802.

Czech, B., Zhou, R., Erlich, Y., *et al.* (2009) Hierarchical rules for Argonaute loading in Drosophila. *Mol Cell* **36**, 445-456.

Davis, E., Caiment, F., Tordoir, X., *et al.* (2005) RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. *Curr Biol* **15**, 743-749.

Davis, T.L., Trasler, J.M., Moss, S.B., Yang, G.J. & Bartolomei, M.S. (1999) Acquisition of the H19 methylation imprint occurs differentially on the parental alleles during spermatogenesis. *Genomics* **58**, 18-28.

Davis, T.L., Yang, G.J., McCarrey, J.R. & Bartolomei, M.S. (2000) The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. *Hum Mol Genet* **9**, 2885-2894.

Deng, W. & Lin, H. (2002) miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev Cell* **2**, 819-830.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F. & Hannon, G.J. (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231-235.

Ding, S.W. & Voinnet, O. (2007) Antiviral immunity directed by small RNAs. *Cell* **130**, 413-426.

Doench, J.G., Petersen, C.P. & Sharp, P.A. (2003) siRNAs can function as miRNAs. *Genes Dev* **17**, 438-442.

Dorer, D.R. & Henikoff, S. (1994) Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. *Cell* **77**, 993-1002.

Dorer, D.R. & Henikoff, S. (1997) Transgene repeat arrays interact with distant heterochromatin and cause silencing in cis and trans. *Genetics* **147**, 1181-1190.

Ebhardt, H.A., Thi, E.P., Wang, M.B. & Unrau, P.J. (2005) Extensive 3' modification of plant small RNAs is modulated by helper component-proteinase expression. *Proc Natl Acad Sci U S A* **102**, 13398-13403.

El-Shami, M., Pontier, D., Lahmy, S., *et al.* (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev* **21**, 2539-2544.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. & Tuschl, T. (2001a) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494-498.

Elbashir, S.M., Lendeckel, W. & Tuschl, T. (2001b) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* **15**, 188-200.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. (2001c) Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *Embo J* **20**, 6877-6888.

Ender, C., Krek, A., Friedlander, M.R., *et al.* (2008) A human snoRNA with microRNA-like functions. *Mol Cell* **32**, 519-528.

Esquela-Kerscher, A. & Slack, F.J. (2006) Oncomirs - microRNAs with a role in cancer. *Nature reviews* **6**, 259-269.

Eulalio, A., Huntzinger, E. & Izaurralde, E. (2008) GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat Struct Mol Biol* **15**, 346-353.

Fang, Y. & Spector, D.L. (2007) Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants. *Curr Biol* **17**, 818-823.

Fanti, L., Dorer, D.R., Berloco, M., Henikoff, S. & Pimpinelli, S. (1998) Heterochromatin protein 1 binds transgene arrays. *Chromosoma* **107**, 286-292.

Filipowicz, W. (2005) RNAi: the nuts and bolts of the RISC machine. *Cell* **122**, 17-20.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. & Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811.

Fitzgerald, D.J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T.J. & Berger, I. (2006) Protein complex expression by using multigene baculoviral vectors. *Nature methods* **3**, 1021-1032.

Flynt, A.S., Greimann, J.C., Chung, W.J., Lima, C.D. & Lai, E.C. (2010) MicroRNA biogenesis via splicing and exosome-mediated trimming in Drosophila. *Mol Cell* **38**, 900-907.

Forstemann, K., Horwich, M.D., Wee, L., Tomari, Y. & Zamore, P.D. (2007) Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* **130**, 287-297.

Forstemann, K., Tomari, Y., Du, T., *et al.* (2005) Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS biology* **3**, e236.

Frank, F., Sonenberg, N. & Nagar, B. (2010) Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**, 818-822.

Frost, R.J., Hamra, F.K., Richardson, J.A., Qi, X., Bassel-Duby, R. & Olson, E.N. (2010) MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. *Proc Natl Acad Sci U S A*.

Fujiwara, Y., Komiya, T., Kawabata, H., *et al.* (1994) Isolation of a DEAD-family protein gene that encodes a murine homolog of Drosophila vasa and its specific expression in germ cell lineage. *Proc Natl Acad Sci U S A* **91**, 12258-12262.

Gaidatzis, D., van Nimwegen, E., Hausser, J. & Zavolan, M. (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC bioinformatics* **8**, 69.

Ghildiyal, M., Seitz, H., Horwich, M.D., *et al.* (2008) Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science* **320**, 1077-1081.

Ghildiyal, M. & Zamore, P.D. (2009) Small silencing RNAs: an expanding universe. *Nat Rev Genet* **10**, 94-108.

Gillespie, C.F., Phifer, J., Bradley, B. & Ressler, K.J. (2009) Risk and resilience: genetic and environmental influences on development of the stress response. *Depression and anxiety* **26**, 984-992.

Giraldez, A.J., Cinalli, R.M., Glasner, M.E., *et al.* (2005) MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308**, 833-838.

Giraldez, A.J., Mishima, Y., Rihel, J., *et al.* (2006) Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75-79.

Girard, A. & Hannon, G.J. (2008) Conserved themes in small-RNA-mediated transposon control. *Trends in cell biology* **18**, 136-148.

Girard, A., Sachidanandam, R., Hannon, G.J. & Carmell, M.A. (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199-202.

Glazov, E.A., Cottee, P.A., Barris, W.C., Moore, R.J., Dalrymple, B.P. & Tizard, M.L. (2008) A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. *Genome Res* **18**, 957-964.

Gouet, P., Courcelle, E., Stuart, D.I. & Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics (Oxford, England)* **15**, 305-308.

Gregory, R.I., Yan, K.P., Amuthan, G., *et al.* (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235-240.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P. & Bartel, D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**, 91-105.

Grimson, A., Srivastava, M., Fahey, B., *et al.* (2008) Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**, 1193-1197.

Grishok, A., Pasquinelli, A.E., Conte, D., *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**, 23-34.

Grivna, S.T., Beyret, E., Wang, Z. & Lin, H. (2006a) A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev* **20**, 1709-1714.

Grivna, S.T., Pyhtila, B. & Lin, H. (2006b) MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. *Proc Natl Acad Sci U S A* **103**, 13415-13420.

Gross, C.T. & McGinnis, W. (1996) DEAF-1, a novel protein that binds an essential region in a Deformed response element. *Embo J* **15**, 1961-1970.

Gunawardane, L.S., Saito, K., Nishida, K.M., *et al.* (2007) A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science* **315**, 1587-1590.

Haase, A.D., Jaskiewicz, L., Zhang, H., *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**, 961-967.

Hajkova, P., Erhardt, S., Lane, N., *et al.* (2002) Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**, 15-23.

Haley, B. & Zamore, P.D. (2004) Kinetic analysis of the RNAi enzyme complex. *Nat Struct Mol Biol* **11**, 599-606.

Hammond, S.M., Bernstein, E., Beach, D. & Hannon, G.J. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**, 293-296.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H. & Kim, V.N. (2004a) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**, 3016-3027.

Han, J., Lee, Y., Yeom, K.H., *et al.* (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887-901.

Han, M.H., Goud, S., Song, L. & Fedoroff, N. (2004b) The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc Natl Acad Sci U S A* **101**, 1093-1098.
Hardy, R.W., Lindsley, D.L., Livak, K.J., *et al.* (1984) Cytogenetic analysis of a segment of the Y chromosome of Drosophila melanogaster. *Genetics* **107**, 591-610.
Harris, A.N. & Macdonald, P.M. (2001) Aubergine encodes a Drosophila polar granule component required for pole cell formation and related to eIF2C. *Development* **128**, 2823-2832.
Hartig, J.V., Esslinger, S., Bottcher, R., Saito, K. & Forstemann, K. (2009) Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. *Embo J* **28**, 2932-2944.
Hartig, J.V., Tomari, Y. & Forstemann, K. (2007) piRNAs--the ancient hunters of genome invaders. *Genes Dev* **21**, 1707-1713.
Hata, K., Kusumi, M., Yokomine, T., Li, E. & Sasaki, H. (2006) Meiotic and epigenetic aberrations in Dnmt3L-deficient male germ cells. *Molecular reproduction and development* **73**, 116-122.
Hay, B., Jan, L.Y. & Jan, Y.N. (1990) Localization of vasa, a component of Drosophila polar granules, in maternal-effect mutants that alter embryonic anteroposterior polarity. *Development* **109**, 425-433.
Herr, A.J., Jensen, M.B., Dalmay, T. & Baulcombe, D.C. (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**, 118-120.
Hock, J. & Meister, G. (2008) The Argonaute protein family. *Genome Biol* **9**, 210.
Horwich, M.D., Li, C., Matranga, C., *et al.* (2007) The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol* **17**, 1265-1272.
Hosokawa, M., Shoji, M., Kitamura, K., *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. *Developmental biology* **301**, 38-52.
Houwing, S., Berezikov, E. & Ketting, R.F. (2008) Zili is required for germ cell differentiation and meiosis in zebrafish. *Embo J* **27**, 2702-2711.
Houwing, S., Kamminga, L.M., Berezikov, E., *et al.* (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129**, 69-82.
Hu, H.Y., Yan, Z., Xu, Y., *et al.* (2009) Sequence features associated with microRNA strand selection in humans and flies. *BMC genomics* **10**, 413.
Huang, Y., Ji, L., Huang, Q., Vassylyev, D.G., Chen, X. & Ma, J.B. (2009) Structural insights into mechanisms of the small RNA methyltransferase HEN1. *Nature* **461**, 823-827.
Hunter, N. (2008) The RecQ DNA helicases: Jacks-of-all-trades or master-tradesmen? *Cell Res* **18**, 328-330.
Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T. & Zamore, P.D. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834-838.
Hutvagner, G. & Simard, M.J. (2008) Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* **9**, 22-32.
Hutvagner, G. & Zamore, P.D. (2002) A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056-2060.
Jackson, R.J. & Standart, N. (2007) How do microRNAs regulate gene expression? *Sci STKE* **2007**, re1.
Jain, R. & Shuman, S. (2010) Bacterial Hen1 is a 3' terminal RNA ribose 2'-O-methyltransferase component of a bacterial RNA repair cassette. *Rna* **16**, 316-323.
Jain, R. & Shuman, S. (2011) Active site mapping and substrate specificity of bacterial Hen1, a manganese-dependent 3' terminal RNA ribose 2'O-methyltransferase. *Rna*.
Jiang, F., Ye, X., Liu, X., Fincher, L., McKearin, D. & Liu, Q. (2005) Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. *Genes Dev* **19**, 1674-1679.
Jin, Y.J., Burakoff, S.J. & Bierer, B.E. (1992) Molecular cloning of a 25-kDa high affinity rapamycin binding protein, FKBP25. *J Biol Chem* **267**, 10942-10945.
Kalmykova, A.I., Dobritsa, A.A. & Gvozdev, V.A. (1998) Su(Ste) diverged tandem repeats in a Y chromosome of Drosophila melanogaster are transcribed and variously processed. *Genetics* **148**, 243-249.
Kamminga, L.M., Luteijn, M.J., den Broeder, M.J., *et al.* (2010) Hen1 is required for oocyte development and piRNA stability in zebrafish. *Embo J*.
Kaneda, M., Okano, M., Hata, K., *et al.* (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900-903.
Kanoh, J., Sadaie, M., Urano, T. & Ishikawa, F. (2005) Telomere binding protein Taz1 establishes Swi6 heterochromatin independently of RNAi at telomeres. *Curr Biol* **15**, 1808-1819.

Katiyar-Agarwal, S., Gao, S., Vivian-Smith, A. & Jin, H. (2007) A novel class of bacteria-induced small RNAs in Arabidopsis. *Genes & development* **21**, 3123-3134.

Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., *et al.* (2006) A pathogen-inducible endogenous siRNA in plant immunity. *Proc Natl Acad Sci U S A* **103**, 18002-18007.

Kato, Y., Kaneda, M., Hata, K., *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**, 2272-2280.

Kawamura, Y., Saito, K., Kin, T., *et al.* (2008) Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**, 793-797.

Kawaoka, S., Hayashi, N., Katsuma, S., *et al.* (2008) Bombyx small RNAs: genomic defense system against transposons in the silkworm, Bombyx mori. *Insect Biochem Mol Biol* **38**, 1058-1065.

Kawaoka, S., Hayashi, N., Suzuki, Y., *et al.* (2009) The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. *Rna* **15**, 1258-1264.

Kedde, M., Strasser, M.J., Boldajipour, B., *et al.* (2007) RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* **131**, 1273-1286.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J. & Plasterk, R.H. (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev* **15**, 2654-2659.

Khvorova, A., Reynolds, A. & Jayasena, S.D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209-216.

Kim, K., Lee, Y.S. & Carthew, R.W. (2007) Conversion of pre-RISC to holo-RISC by Ago2 during assembly of RNAi complexes. *Rna* **13**, 22-29.

Kim, V.N. (2006) Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes. *Genes Dev* **20**, 1993-1997.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B. & Schilling, T.F. (1995) Stages of embryonic development of the zebrafish. *Dev Dyn* **203**, 253-310.

Kirino, Y., Kim, N., de Planell-Saguer, M., *et al.* (2009) Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. *Nat Cell Biol* **11**, 652-658.

Kirino, Y. & Mourelatos, Z. (2007a) 2'-O-methyl modification in mouse piRNAs and its methylase. *Nucleic acids symposium series (2004)*, 417-418.

Kirino, Y. & Mourelatos, Z. (2007b) The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *Rna* **13**, 1397-1401.

Kirino, Y. & Mourelatos, Z. (2007c) Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nat Struct Mol Biol* **14**, 347-348.

Kirino, Y., Vourekas, A., Kim, N., *et al.* (2010a) Arginine methylation of vasa protein is conserved across phyla. *J Biol Chem*.

Kirino, Y., Vourekas, A., Sayed, N., *et al.* (2010b) Arginine methylation of Aubergine mediates Tudor binding and germ plasm localization. *Rna* **16**, 70-78.

Klattenhoff, C., Bratu, D.P., McGinnis-Schultz, N., Koppetsch, B.S., Cook, H.A. & Theurkauf, W.E. (2007) Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev Cell* **12**, 45-55.

Klattenhoff, C. & Theurkauf, W. (2008) Biogenesis and germline functions of piRNAs. *Development* **135**, 3-9.

Klattenhoff, C., Xi, H., Li, C., *et al.* (2009) The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* **138**, 1137-1149.

Knight, S.W. & Bass, B.L. (2001) A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. *Science* **293**, 2269-2271.

Kojima, K., Kuramochi-Miyagawa, S., Chuma, S., *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**, 1155-1165.

Kotaja, N., Bhattacharyya, S.N., Jaskiewicz, L., *et al.* (2006) The chromatoid body of male germ cells: similarity with processing bodies and presence of Dicer and microRNA pathway components. *Proc Natl Acad Sci U S A* **103**, 2647-2652.

Kotaja, N. & Sassone-Corsi, P. (2007) The chromatoid body: a germ-cell-specific RNA-processing centre. *Nat Rev Mol Cell Biol* **8**, 85-90.

Krek, A., Grun, D., Poy, M.N., *et al.* (2005) Combinatorial microRNA target predictions. *Nat Genet* **37**, 495-500.

Krutzfeldt, J., Rajewsky, N., Braich, R., *et al.* (2005) Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438**, 685-689.

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., *et al.* (2004) Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* **131**, 839-849.

Kuramochi-Miyagawa, S., Kimura, T., Yomogida, K., *et al.* (2001) Two mouse piwi-related genes: miwi and mili. *Mech Dev* **108**, 121-133.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., *et al.* (2010) MVH in piRNA processing and gene silencing of retrotransposons. *Genes Dev* **24**, 887-892.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., *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.

Kurihara, Y., Takashi, Y. & Watanabe, Y. (2006) The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *Rna* **12**, 206-212.

Kurihara, Y. & Watanabe, Y. (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci U S A* **101**, 12753-12758.

Kurth, H.M. & Mochizuki, K. (2009) 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in Tetrahymena. *Rna* **15**, 675-685.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853-858.

Lai, E.C. (2004) Predicting and validating microRNA targets. *Genome Biol* **5**, 115.

Lammers, J.H., Offenberg, H.H., van Aalderen, M., Vink, A.C., Dietrich, A.J. & Heyting, C. (1994) The gene encoding a major component of the lateral elements of synaptonemal complexes of the rat is related to X-linked lymphocyte-regulated genes. *Mol Cell Biol* **14**, 1137-1146.

Lander, E.S., Linton, L.M., Birren, B., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.

Landgraf, P., Rusu, M., Sheridan, R., *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401-1414.

Landthaler, M., Yalcin, A. & Tuschl, T. (2004) The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. *Curr Biol* **14**, 2162-2167.

Lasko, P.F. & Ashburner, M. (1990) Posterior localization of vasa protein correlates with, but is not sufficient for, pole cell development. *Genes Dev* **4**, 905-921.

Lau, N.C. (2010) Small RNAs in the animal gonad: guarding genomes and guiding development. *The international journal of biochemistry & cell biology* **42**, 1334-1347.

Lau, N.C., Lim, L.P., Weinstein, E.G. & Bartel, D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**, 858-862.

Lau, N.C., Seto, A.G., Kim, J., *et al.* (2006) Characterization of the piRNA complex from rat testes. *Science* **313**, 363-367.

Lee, R.C. & Ambros, V. (2001) An extensive class of small RNAs in Caenorhabditis elegans. *Science* **294**, 862-864.

Lee, R.C., Feinbaum, R.L. & Ambros, V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854.

Lee, Y., Ahn, C., Han, J., *et al.* (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-419.

Lee, Y., Jeon, K., Lee, J.T., Kim, S. & Kim, V.N. (2002) MicroRNA maturation: stepwise processing and subcellular localization. *Embo J* **21**, 4663-4670.

Lee, Y., Kim, M., Han, J., *et al.* (2004a) MicroRNA genes are transcribed by RNA polymerase II. *Embo J* **23**, 4051-4060.

Lee, Y.S., Nakahara, K., Pham, J.W., *et al.* (2004b) Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69-81.

Leuschner, P.J., Ameres, S.L., Kueng, S. & Martinez, J. (2006) Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep* **7**, 314-320.

Lewis, B.P., Burge, C.B. & Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. & Burge, C.B. (2003) Prediction of mammalian microRNA targets. *Cell* **115**, 787-798.

Li, C., Vagin, V.V., Lee, S., *et al.* (2009) Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137**, 509-521.

Li, J., Yang, Z., Yu, B., Liu, J. & Chen, X. (2005) Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr Biol* **15**, 1501-1507.

Li, J.Y., Lees-Murdock, D.J., Xu, G.L. & Walsh, C.P. (2004) Timing of establishment of paternal methylation imprints in the mouse. *Genomics* **84**, 952-960.

Liang, L., Diehl-Jones, W. & Lasko, P. (1994) Localization of vasa protein to the Drosophila pole plasm is independent of its RNA-binding and helicase activities. *Development* **120**, 1201-1211.

Lim, L.P., Lau, N.C., Garrett-Engele, P., *et al.* (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769-773.

Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. (2003) Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. *Nature* **426**, 465-469.

Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. (2004) Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol* **11**, 576-577.

Linsley, P.S., Schelter, J., Burchard, J., *et al.* (2007) Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol Cell Biol* **27**, 2240-2252.

Liu, H., Wang, J.Y., Huang, Y., *et al.* (2010) Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes Dev* **24**, 1876-1881.

Liu, J., Carmell, M.A., Rivas, F.V., *et al.* (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437-1441.

Liu, J., Rivas, F.V., Wohlschlegel, J., Yates, J.R., 3rd, Parker, R. & Hannon, G.J. (2005a) A role for the P-body component GW182 in microRNA function. *Nat Cell Biol* **7**, 1261-1266.

Liu, J., Valencia-Sanchez, M.A., Hannon, G.J. & Parker, R. (2005b) MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat Cell Biol* **7**, 719-723.

Liu, Q., Rand, T.A., Kalidas, S., *et al.* (2003) R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. *Science* **301**, 1921-1925.

Liu, Y., Chen, W., Gaudet, J., *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**, 483-497.

Livak, K.J. (1984) Organization and mapping of a sequence on the Drosophila melanogaster X and Y chromosomes that is transcribed during spermatogenesis. *Genetics* **107**, 611-634.

Livak, K.J. (1990) Detailed structure of the Drosophila melanogaster stellate genes and their transcripts. *Genetics* **124**, 303-316.

Llave, C., Kasschau, K.D., Rector, M.A. & Carrington, J.C. (2002) Endogenous and silencing-associated small RNAs in plants. *The Plant cell* **14**, 1605-1619.

Lovasco, L.A., Seymour, K.A., Zafra, K., O'Brien, C.W., Schorl, C. & Freiman, R.N. (2010) Accelerated ovarian aging in the absence of the transcription regulator TAF4B in mice. *Biol Reprod* **82**, 23-34.

Lu, J., Getz, G., Miska, E.A., *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-838.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E. & Kutay, U. (2004) Nuclear export of microRNA precursors. *Science* **303**, 95-98.

Lutterbach, B., Sun, D., Schuetz, J. & Hiebert, S.W. (1998) The MYND motif is required for repression of basal transcription from the multidrug resistance 1 promoter by the t(8;21) fusion protein. *Mol Cell Biol* **18**, 3604-3611.

Ma, J.B., Ye, K. & Patel, D.J. (2004) Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**, 318-322.

Ma, L., Buchold, G.M., Greenbaum, M.P., *et al.* (2009) GASZ is essential for male meiosis and suppression of retrotransposon expression in the male germline. *PLoS Genet* **5**, e1000635.

Maatouk, D.M., Kellam, L.D., Mann, M.R., *et al.* (2006) DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development* **133**, 3411-3418.

Malecova, B. & Morris, K.V. (2010) Transcriptional gene silencing through epigenetic changes mediated by non-coding RNAs. *Current opinion in molecular therapeutics* **12**, 214-222.

Malone, C.D., Brennecke, J., Dus, M., *et al.* (2009) Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* **137**, 522-535.

Mansfield, J.H., Harfe, B.D., Nissen, R., *et al.* (2004) MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nat Genet* **36**, 1079-1083.

Martin-Morris, L.E., Csink, A.K., Dorer, D.R., Talbert, P.B. & Henikoff, S. (1997) Heterochromatic trans-inactivation of Drosophila white transgenes. *Genetics* **147**, 671-677.

Martinez, J. & Tuschl, T. (2004) RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev* **18**, 975-980.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P. & Zamore, P.D. (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**, 607-620.

Mattick, J.S. (2009) The genetic signatures of noncoding RNAs. *PLoS Genet* **5**, e1000459.

Meikar, O., Da Ros, M., Liljenback, H., Toppari, J. & Kotaja, N. (2010) Accumulation of piRNAs in the chromatoid bodies purified by a novel isolation protocol. *Experimental cell research* **316**, 1567-1575.

Meister, G., Eggert, C. & Fischer, U. (2002) SMN-mediated assembly of RNPs: a complex story. *Trends in cell biology* **12**, 472-478.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G. & Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* **15**, 185-197.

Meister, G., Landthaler, M., Peters, L., *et al.* (2005) Identification of novel argonaute-associated proteins. *Curr Biol* **15**, 2149-2155.

Meng, X., Lu, X., Morris, C.A. & Keating, M.T. (1998) A novel human gene FKBP6 is deleted in Williams syndrome. *Genomics* **52**, 130-137.

Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A. & Matzke, A.J. (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *Embo J* **19**, 5194-5201.

Miyoshi, K., Miyoshi, T., Hartig, J.V., Siomi, H. & Siomi, M.C. (2010) Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in Drosophila. *Rna* **16**, 506-515.

Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H. & Siomi, M.C. (2005) Slicer function of Drosophila Argonautes and its involvement in RISC formation. *Genes Dev* **19**, 2837-2848.

Mochizuki, K., Fine, N.A., Fujisawa, T. & Gorovsky, M.A. (2002) Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell* **110**, 689-699.

Moens, P.B., Heyting, C., Dietrich, A.J., van Raamsdonk, W. & Chen, Q. (1987) Synaptonemal complex antigen location and conservation. *J Cell Biol* **105**, 93-103.

Mourelatos, Z., Dostie, J., Paushkin, S., *et al.* (2002) miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev* **16**, 720-728.

Mui Chan, C., Zhou, C., Brunzelle, J.S. & Huang, R.H. (2009) Structural and biochemical insights into 2'-O-methylation at the 3'-terminal nucleotide of RNA by Hen1. *Proc Natl Acad Sci U S A* **106**, 17699-17704.

Nair, S.C., Rimerman, R.A., Toran, E.J., *et al.* (1997) Molecular cloning of human FKBP51 and comparisons of immunophilin interactions with Hsp90 and progesterone receptor. *Mol Cell Biol* **17**, 594-603.

Nakamura, M., Masuda, H., Horii, J., *et al.* (1998) When overexpressed, a novel centrosomal protein, RanBPM, causes ectopic microtubule nucleation similar to gamma-tubulin. *J Cell Biol* **143**, 1041-1052.

Nishida, K.M., Okada, T.N., Kawamura, T., *et al.* (2009) Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in Drosophila germlines. *Embo J* **28**, 3820-3831.

Nishida, K.M., Saito, K., Mori, T., *et al.* (2007) Gene silencing mechanisms mediated by Aubergine piRNA complexes in Drosophila male gonad. *Rna* **13**, 1911-1922.

Nishitani, H., Hirose, E., Uchimura, Y., *et al.* (2001) Full-sized RanBPM cDNA encodes a protein possessing a long stretch of proline and glutamine within the N-terminal region, comprising a large protein complex. *Gene* **272**, 25-33.

Noce, T., Okamoto-Ito, S. & Tsunekawa, N. (2001) Vasa homolog genes in mammalian germ cell development. *Cell structure and function* **26**, 131-136.

Noguchi, J., Ozawa, M., Nakai, M., *et al.* (2008) Affected homologous chromosome pairing and phosphorylation of testis specific histone, H2AX, in male meiosis under FKBP6 deficiency. *The Journal of reproduction and development* **54**, 203-207.

Noguchi, J., Yoshida, M., Ikadai, H., Imamichi, T., Watanabe, G. & Taya, K. (1993) Age-related changes in blood concentrations of FSH, LH and testosterone and testicular morphology in a new rat sterile mutant with hereditary aspermia. *Journal of reproduction and fertility* **97**, 433-439.

Obbard, D.J., Jiggins, F.M., Halligan, D.L. & Little, T.J. (2006) Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr Biol* **16**, 580-585.

Ochocka, A.M., Kampanis, P., Nicol, S., *et al.* (2009) FKBP25, a novel regulator of the p53 pathway, induces the degradation of MDM2 and activation of p53. *FEBS letters* **583**, 621-626.

Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H. & Miyauchi, K. (2007a) The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated. *Nat Struct Mol Biol* **14**, 349-350.

Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H., Miyauchi, K. & Suzuki, T. (2007b) The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated. *Nat Struct Mol Biol* **14**, 349-350.

Okada, C., Yamashita, E., Lee, S.J., *et al.* (2009) A high-resolution structure of the pre-microRNA nuclear export machinery. *Science* **326**, 1275-1279.

Okamoto, T., Nishimura, Y., Ichimura, T., *et al.* (2006) Hepatitis C virus RNA replication is regulated by FKBP8 and Hsp90. *Embo J* **25**, 5015-5025.

Okamura, K., Balla, S., Martin, R., Liu, N. & Lai, E.C. (2008a) Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster. *Nat Struct Mol Biol* **15**, 998.

Okamura, K., Chung, W.J., Ruby, J.G., Guo, H., Bartel, D.P. & Lai, E.C. (2008b) The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* **453**, 803-806.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M. & Lai, E.C. (2007) The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell* **130**, 89-100.

Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M.C. (2004) Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* **18**, 1655-1666.

Okamura, K., Liu, N. & Lai, E.C. (2009) Distinct mechanisms for microRNA strand selection by Drosophila Argonautes. *Mol Cell* **36**, 431-444.

Okamura, K., Phillips, M.D., Tyler, D.M., Duan, H., Chou, Y.T. & Lai, E.C. (2008c) The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. *Nat Struct Mol Biol* **15**, 354-363.

Olivieri, D., Sykora, M.M., Sachidanandam, R., Mechtler, K. & Brennecke, J. (2010) An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. *Embo J*.

Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O. & Pikaard, C.S. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**, 613-622.

Pal-Bhadra, M., Bhadra, U. & Birchler, J.A. (2002) RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. *Mol Cell* **9**, 315-327.

Pal, S., Vishwanath, S.N., Erdjument-Bromage, H., Tempst, P. & Sif, S. (2004) Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes. *Mol Cell Biol* **24**, 9630-9645.

Palumbo, G., Bonaccorsi, S., Robbins, L.G. & Pimpinelli, S. (1994) Genetic analysis of Stellate elements of Drosophila melanogaster. *Genetics* **138**, 1181-1197.

Pan, J., Goodheart, M., Chuma, S., Nakatsuji, N., Page, D.C. & Wang, P.J. (2005) RNF17, a component of the mammalian germ cell nuage, is essential for spermiogenesis. *Development* **132**, 4029-4039.

Pane, A., Wehr, K. & Schupbach, T. (2007) zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. *Dev Cell* **12**, 851-862.

Papp, I., Mette, M.F., Aufsatz, W., *et al.* (2003) Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. *Plant physiology* **132**, 1382-1390.

Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H. & Poethig, R.S. (2005) Nuclear processing and export of microRNAs in Arabidopsis. *Proc Natl Acad Sci U S A* **102**, 3691-3696.

Park, W., Li, J., Song, R., Messing, J. & Chen, X. (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr Biol* **12**, 1484-1495.

Parker, J.S., Parizotto, E.A., Wang, M., Roe, S.M. & Barford, D. (2009) Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol Cell* **33**, 204-214.

Parker, J.S., Roe, S.M. & Barford, D. (2004) Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *Embo J* **23**, 4727-4737.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., *et al.* (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86-89.

Pedersen, K.M., Finsen, B., Celis, J.E. & Jensen, N.A. (1999) muFKBP38: a novel murine immunophilin homolog differentially expressed in Schwannoma cells and central nervous system neurons in vivo. *Electrophoresis* **20**, 249-255.

Pelisson, A., Sarot, E., Payen-Groschene, G. & Bucheton, A. (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**, 1951-1960.

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L. & Poethig, R.S. (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes Dev* **18**, 2368-2379.

Peters, L. & Meister, G. (2007) Argonaute proteins: mediators of RNA silencing. *Mol Cell* **26**, 611-623.

Pfeffer, S., Zavolan, M., Grasser, F.A., *et al.* (2004) Identification of virus-encoded microRNAs. *Science* **304**, 734-736.

Pillai, R.S., Artus, C.G. & Filipowicz, W. (2004) Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *Rna* **10**, 1518-1525.

Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., *et al.* (2005) Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **309**, 1573-1576.

Pivot-Pajot, C., Caron, C., Govin, J., Vion, A., Rousseaux, S. & Khochbin, S. (2003) Acetylation-dependent chromatin reorganization by BRDT, a testis-specific bromodomain-containing protein. *Mol Cell Biol* **23**, 5354-5365.

Pontier, D., Yahubyan, G., Vega, D., *et al.* (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev* **19**, 2030-2040.

Ponting, C.P. (1997) Tudor domains in proteins that interact with RNA. *Trends Biochem Sci* **22**, 51-52.

Poy, M.N., Eliasson, L., Krutzfeldt, J., *et al.* (2004) A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* **432**, 226-230.

Rand, T.A., Petersen, S., Du, F. & Wang, X. (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123**, 621-629.

Raz, E. (2000) The function and regulation of vasa-like genes in germ-cell development. *Genome Biol* **1**, REVIEWS1017.

Reinhart, B.J. & Bartel, D.P. (2002) Small RNAs correspond to centromere heterochromatic repeats. *Science* **297**, 1831.

Reinhart, B.J., Slack, F.J., Basson, M., *et al.* (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901-906.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B. & Bartel, D.P. (2002) MicroRNAs in plants. *Genes Dev* **16**, 1616-1626.

Reuter, M., Chuma, S., Tanaka, T., Franz, T., Stark, A. & Pillai, R.S. (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.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B. & Bartel, D.P. (2002) Prediction of plant microRNA targets. *Cell* **110**, 513-520.

Ro, S., Park, C., Song, R., *et al.* (2007) Cloning and expression profiling of testis-expressed piRNA-like RNAs. *Rna* **13**, 1693-1702.

Rosner, M., Hofer, K., Kubista, M. & Hengstschlager, M. (2003) Cell size regulation by the human TSC tumor suppressor proteins depends on PI3K and FKBP38. *Oncogene* **22**, 4786-4798.

Ruby, J.G., Jan, C., Player, C., *et al.* (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell* **127**, 1193-1207.

Ruby, J.G., Jan, C.H. & Bartel, D.P. (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**, 83-86.

Saito, K., Inagaki, S., Mituyama, T., *et al.* (2009) A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. *Nature* **461**, 1296-1299.

Saito, K., Ishizu, H., Komai, M., *et al.* (2010) Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. *Genes Dev* **24**, 2493-2498.

Saito, K., Ishizuka, A., Siomi, H. & Siomi, M.C. (2005) Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. *PLoS biology* **3**, e235.

Saito, K., Nishida, K.M., Mori, T., *et al.* (2006) Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes Dev* **20**, 2214-2222.

Saito, K., Sakaguchi, Y., Suzuki, T., Siomi, H. & Siomi, M.C. (2007a) Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes Dev* **21**, 1603-1608.

Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H. & Siomi, M.C. (2007b) Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes Dev* **21**, 1603-1608.

Saito, K. & Siomi, M.C. (2010) Small RNA-mediated quiescence of transposable elements in animals. *Dev Cell* **19**, 687-697.

Saraiya, A.A. & Wang, C.C. (2008) snoRNA, a novel precursor of microRNA in Giardia lamblia. *PLoS pathogens* **4**, e1000224.

Sarot, E., Payen-Groschene, G., Bucheton, A. & Pelisson, A. (2004) Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene. *Genetics* **166**, 1313-1321.

Sasaki, T., Shiohama, A., Minoshima, S. & Shimizu, N. (2003) Identification of eight members of the Argonaute family in the human genome small star, filled. *Genomics* **82**, 323-330.

Sasidharan, R. & Gerstein, M. (2008) Genomics: protein fossils live on as RNA. *Nature* **453**, 729-731.

Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A. & Gvozdev, V. (2006) Telomere elongation is under the control of the RNAi-based mechanism in the Drosophila germline. *Genes Dev* **20**, 345-354.

Schmidt, A., Palumbo, G., Bozzetti, M.P., Tritto, P., Pimpinelli, S. & Schafer, U. (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**, 749-760.

Schupbach, T. & Wieschaus, E. (1991) Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* **129**, 1119-1136.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N. & Zamore, P.D. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199-208.

Schwarz, D.S., Tomari, Y. & Zamore, P.D. (2004) The RNA-induced silencing complex is a Mg2+-dependent endonuclease. *Curr Biol* **14**, 787-791.

Seki, Y., Yamaji, M., Yabuta, Y., *et al.* (2007) Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* **134**, 2627-2638.

Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R. & Rajewsky, N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58-63.

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.

Sen, G.L. & Blau, H.M. (2005) Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nat Cell Biol* **7**, 633-636.

Seto, A.G., Kingston, R.E. & Lau, N.C. (2007) The coming of age for Piwi proteins. *Mol Cell* **26**, 603-609.

Shibata, N., Tsunekawa, N., Okamoto-Ito, S., Akasu, R., Tokumasu, A. & Noce, T. (2004) Mouse RanBPM is a partner gene to a germline specific RNA helicase, mouse vasa homolog protein. *Molecular reproduction and development* **67**, 1-7.

Shirane, M. & Nakayama, K.I. (2003) Inherent calcineurin inhibitor FKBP38 targets Bcl-2 to mitochondria and inhibits apoptosis. *Nat Cell Biol* **5**, 28-37.

Shoji, M., Tanaka, T., Hosokawa, M., *et al.* (2009) The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Dev Cell* **17**, 775-787.

Shpiz, S., Kwon, D., Uneva, A., *et al.* (2007) Characterization of Drosophila telomeric retroelement TAHRE: transcription, transpositions, and RNAi-based regulation of expression. *Mol Biol Evol* **24**, 2535-2545.

Shpiz, S.G. & Kalmykova, A.I. (2007) Structure of telomeric chromatin in Drosophila. *Biochemistry (Mosc)* **72**, 618-630.

Simon, B., Kirkpatrick, J.P., Eckhardt, S., *et al.* (2011) Recognition of 2'-O-Methylated 3'-End of piRNA by the PAZ Domain of a Piwi Protein. *Structure*.

Siomi, M.C., Mannen, T. & Siomi, H. (2010) How does the royal family of Tudor rule the PIWI-interacting RNA pathway? *Genes Dev* **24**, 636-646.

Song, J.J., Liu, J., Tolia, N.H., *et al.* (2003) The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Biol* **10**, 1026-1032.

Song, J.J., Smith, S.K., Hannon, G.J. & Joshua-Tor, L. (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**, 1434-1437.

Stark, A., Brennecke, J., Bushati, N., Russell, R.B. & Cohen, S.M. (2005) Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell* **123**, 1133-1146.

Strahl, B.D., Briggs, S.D., Brame, C.J., *et al.* (2001) Methylation of histone H4 at arginine 3 occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1. *Curr Biol* **11**, 996-1000.

Sullivan, C.S., Grundhoff, A.T., Tevethia, S., Pipas, J.M. & Ganem, D. (2005) SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature* **435**, 682-686.

Sun, H., Li, D., Chen, S., *et al.* (2010) Zili inhibits transforming growth factor-beta signaling by interacting with Smad4. *J Biol Chem* **285**, 4243-4250.

Szakmary, A., Reedy, M., Qi, H. & Lin, H. (2009) The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in Drosophila melanogaster. *J Cell Biol* **185**, 613-627.

Tabara, H., Sarkissian, M., Kelly, W.G., *et al.* (1999) The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell* **99**, 123-132.

Tabara, H., Yigit, E., Siomi, H. & Mello, C.C. (2002) The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans. *Cell* **109**, 861-871.

Talbot, K., Miguel-Aliaga, I., Mohaghegh, P., Ponting, C.P. & Davies, K.E. (1998) Characterization of a gene encoding survival motor neuron (SMN)-related protein, a constituent of the spliceosome complex. *Hum Mol Genet* **7**, 2149-2156.

Tam, O.H., Aravin, A.A., Stein, P., *et al.* (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**, 534-538.

Tanaka, S.S., Toyooka, Y., Akasu, R., *et al.* (2000) The mouse homolog of Drosophila Vasa is required for the development of male germ cells. *Genes Dev* **14**, 841-853.

Tang, G., Reinhart, B.J., Bartel, D.P. & Zamore, P.D. (2003) A biochemical framework for RNA silencing in plants. *Genes Dev* **17**, 49-63.

Thomson, T. & Lasko, P. (2004) Drosophila tudor is essential for polar granule assembly and pole cell specification, but not for posterior patterning. *Genesis* **40**, 164-170.

Tian, Y., Simanshu, D.K., Ma, J.B. & Patel, D.J. (2010) Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. *Proc Natl Acad Sci U S A*.

Tkaczuk, K.L., Obarska, A. & Bujnicki, J.M. (2006) Molecular phylogenetics and comparative modeling of HEN1, a methyltransferase involved in plant microRNA biogenesis. *BMC evolutionary biology* **6**, 6.

Tolia, N.H. & Joshua-Tor, L. (2007) Slicer and the argonautes. *Nature chemical biology* **3**, 36-43.

Tomari, Y., Du, T. & Zamore, P.D. (2007) Sorting of Drosophila small silencing RNAs. *Cell* **130**, 299-308.

Tomari, Y., Matranga, C., Haley, B., Martinez, N. & Zamore, P.D. (2004) A protein sensor for siRNA asymmetry. *Science* **306**, 1377-1380.

Tomari, Y. & Zamore, P.D. (2005) Perspective: machines for RNAi. *Genes Dev* **19**, 517-529.

Toyooka, Y., Tsunekawa, N., Takahashi, Y., Matsui, Y., Satoh, M. & Noce, T. (2000) Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. *Mech Dev* **93**, 139-149.

Tran, R.K., Henikoff, J.G., Zilberman, D., Ditt, R.F., Jacobsen, S.E. & Henikoff, S. (2005) DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes. *Curr Biol* **15**, 154-159.

Ueda, T., Abe, K., Miura, A., *et al.* (2000) The paternal methylation imprint of the mouse H19 locus is acquired in the gonocyte stage during foetal testis development. *Genes Cells* **5**, 649-659.

Unhavaithaya, Y., Hao, Y., Beyret, E., *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**, 6507-6519.

Vagin, V.V., Klenov, M.S., Kalmykova, A.I., Stolyarenko, A.D., Kotelnikov, R.N. & Gvozdev, V.A. (2004) The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of Drosophila melanogaster. *RNA Biol* **1**, 54-58.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V. & Zamore, P.D. (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320-324.

Vagin, V.V., Wohlschlegel, J., Qu, J., *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**, 1749-1762.

Vasileva, A., Tiedau, D., Firooznia, A., Muller-Reichert, T. & Jessberger, R. (2009) Tdrd6 is required for spermiogenesis, chromatoid body architecture, and regulation of miRNA expression. *Curr Biol* **19**, 630-639.

Vazquez, F. (2006) Arabidopsis endogenous small RNAs: highways and byways. *Trends in plant science* **11**, 460-468.

Vazquez, F., Gasciolli, V., Crete, P. & Vaucheret, H. (2004a) The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Curr Biol* **14**, 346-351.

Vazquez, F., Vaucheret, H., Rajagopalan, R., *et al.* (2004b) Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol Cell* **16**, 69-79.

Verdel, A., Jia, S., Gerber, S., *et al.* (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672-676.

Vilcek, J. (2006) Fifty years of interferon research: aiming at a moving target. *Immunity* **25**, 343-348.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I. & Martienssen, R.A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833-1837.

Wang, H., Huang, Z.Q., Xia, L., *et al.* (2001a) Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Science* **293**, 853-857.

Wang, J., Saxe, J.P., Tanaka, T., Chuma, S. & Lin, H. (2009a) Mili interacts with tudor domain-containing protein 1 in regulating spermatogenesis. *Curr Biol* **19**, 640-644.

Wang, J., Tong, W., Zhang, X., *et al.* (2006) Hepatitis C virus non-structural protein NS5A interacts with FKBP38 and inhibits apoptosis in Huh7 hepatoma cells. *FEBS letters* **580**, 4392-4400.

Wang, P.J., McCarrey, J.R., Yang, F. & Page, D.C. (2001b) An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* **27**, 422-426.

Wang, X., Han, Y., Dang, Y., *et al.* (2010) Moloney leukemia virus 10 (MOV10) protein inhibits retrovirus replication. *J Biol Chem* **285**, 14346-14355.

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T. & Patel, D.J. (2008a) Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921-926.

Wang, Y., Juranek, S., Li, H., *et al.* (2009b) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461**, 754-761.

Wang, Y., Medvid, R., Melton, C., Jaenisch, R. & Blelloch, R. (2007) DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* **39**, 380-385.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D.J. (2008b) Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**, 209-213.

Watanabe, T., Takeda, A., Tsukiyama, T., *et al.* (2006) Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev* **20**, 1732-1743.

Watanabe, T., Totoki, Y., Toyoda, A., *et al.* (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**, 539-543.

Webster, K.E., O'Bryan, M.K., Fletcher, S., *et al.* (2005) Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc Natl Acad Sci U S A* **102**, 4068-4073.

Wightman, B., Ha, I. & Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**, 855-862.

Williams, B.R. (1999) PKR; a sentinel kinase for cellular stress. *Oncogene* **18**, 6112-6120.

Williams, L., Carles, C.C., Osmont, K.S. & Fletcher, J.C. (2005) A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the Arabidopsis ARF2, ARF3, and ARF4 genes. *Proc Natl Acad Sci U S A* **102**, 9703-9708.

Wulczyn, F.G., Smirnova, L., Rybak, A., *et al.* (2007) Post-transcriptional regulation of the let-7 microRNA during neural cell specification. *Faseb J* **21**, 415-426.

Xie, X., Lu, J., Kulbokas, E.J., *et al.* (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**, 338-345.

Xie, Z., Johansen, L.K., Gustafson, A.M., *et al.* (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS biology* **2**, E104.

Xie, Z., Kasschau, K.D. & Carrington, J.C. (2003) Negative feedback regulation of Dicer-Like1 in Arabidopsis by microRNA-guided mRNA degradation. *Curr Biol* **13**, 784-789.

Xu, M., You, Y., Hunsicker, P., *et al.* (2008) Mice deficient for a small cluster of Piwi-interacting RNAs implicate Piwi-interacting RNAs in transposon control. *Biol Reprod* **79**, 51-57.

Xu, P., Vernooy, S.Y., Guo, M. & Hay, B.A. (2003) The Drosophila microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. *Curr Biol* **13**, 790-795.

Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L. & Zhou, M.M. (2003) Structure and conserved RNA binding of the PAZ domain. *Nature* **426**, 468-474.

Yang, L., Liu, Z., Lu, F., Dong, A. & Huang, H. (2006a) SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. *Plant J* **47**, 841-850.

Yang, N. & Kazazian, H.H., Jr. (2006) L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol* **13**, 763-771.

Yang, W.M., Yao, Y.L. & Seto, E. (2001) The FK506-binding protein 25 functionally associates with histone deacetylases and with transcription factor YY1. *Embo J* **20**, 4814-4825.

Yang, Z., Ebright, Y.W., Yu, B. & Chen, X. (2006b) HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic acids research* **34**, 667-675.

Yekta, S., Shih, I.H. & Bartel, D.P. (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**, 594-596.

Yeung, M.L., Bennasser, Y., Watashi, K., Le, S.Y., Houzet, L. & Jeang, K.T. (2009) Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid. *Nucleic acids research* **37**, 6575-6586.

Yi, R., Qin, Y., Macara, I.G. & Cullen, B.R. (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**, 3011-3016.

Yin, H. & Lin, H. (2007) An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. *Nature* **450**, 304-308.

Yoda, M., Kawamata, T., Paroo, Z., *et al.* (2010) ATP-dependent human RISC assembly pathways. *Nat Struct Mol Biol* **17**, 17-23.

Yoshikawa, M., Peragine, A., Park, M.Y. & Poethig, R.S. (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes Dev* **19**, 2164-2175.

Yu, B., Bi, L., Zheng, B., *et al.* (2008) The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *Proc Natl Acad Sci U S A* **105**, 10073-10078.

Yu, B., Yang, Z., Li, J., *et al.* (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932-935.

Yuan, Y.R., Pei, Y., Ma, J.B., *et al.* (2005) Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol Cell* **19**, 405-419.

Zamore, P.D., Tuschl, T., Sharp, P.A. & Bartel, D.P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**, 25-33.

Zeng, Y. & Cullen, B.R. (2004) Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic acids research* **32**, 4776-4785.

Zhang, J.F., Yuan, L.J., Shao, Y., Du, W., Yan, D.W. & Lu, Y.T. (2008) The disturbance of small RNA pathways enhanced abscisic acid response and multiple stress responses in Arabidopsis. *Plant, cell & environment* **31**, 562-574.

Zheng, K., Xiol, J., Reuter, M., *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*.

Zhou, X., Zuo, Z., Zhou, F., *et al.* (2010) Profiling Sex-specific piRNAs in Zebrafish. *Genetics*.

Zhu, Q.H., Spriggs, A., Matthew, L., *et al.* (2008) A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. *Genome Res* **18**, 1456-1465.

Zilberman, D., Cao, X. & Jacobsen, S.E. (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**, 716-719.