



HAL
open science

Analyses fonctionnelles de les ARN hélicases Vasa et TDRD9 dans la voie des piRNA

Pietro Spinelli

► **To cite this version:**

Pietro Spinelli. Analyses fonctionnelles de les ARN hélicases Vasa et TDRD9 dans la voie des piRNA. Sciences agricoles. Université Grenoble Alpes, 2015. Français. NNT : 2015GREAV047 . tel-01686348

HAL Id: tel-01686348

<https://theses.hal.science/tel-01686348>

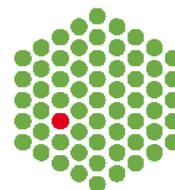
Submitted on 17 Jan 2018

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

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

UNIVERSITÉ DE GRENOBLE

EMBL



THESIS / THÈSE

To obtain the title of / Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Discipline/*Spécialité*: Biotechnologie

Arrêté ministériel : 7 août 2006

Presented by / *Présentée par*

Pietro Spinelli

Thesis supervisor / *Thèse dirigée par* **Dr. Ramesh Pillai**

Thesis prepared at / *Thèse préparée au sein de*

European Molecular Biology Laboratory (EMBL), Grenoble

In/*Dans* l'École Doctorale de Chimie et Science du Vivant

Functional studies on the RNA helicases Vasa and TDRD9 in the piRNA pathway

Public defense on / Thèse soutenue publiquement le 01/12/2015

Jury members / devant le jury composé de:

Prof. Winfried Weissenhorn

Président du jury

Dr. Alena Shkumatava

Rapporteur

Dr. Marie-Laure Baudet

Rapporteur

Dr. André Verdel

Examineur

Prof. Dónal O'Carroll

Examineur

Dr. Ramesh Pillai

Directeur de thèse



A Iacopo, Aldo e Luigi

Résumé

Les protéines PIWI sont exprimées dans les gonades (testicules et ovaires) des animaux où elles s'associent à des petits ARN nommés piRNAs (Piwi-interaction RNAs) et répriment les éléments transposables, défendant ainsi de l'intégrité du génome. En effet, les animaux *knock-out* pour les protéines PIWI montrent une perte de piRNAs et une activation des éléments transposables avec des conséquences catastrophiques: l'arrêt du développement des cellules germinales, due potentiellement à des dommages du génome qui entraîne l'infertilité. Le « Silencing » est réalisé soit par l'activité endonucléase guidée par les piRNAs de la protéine PIWI cytosolique ou par le recrutement de la machinerie de répression transcriptionnelle sur les *loci* génomiques des cibles par PIWI nucléaire. La biogenèse des piRNAs peut être divisée en deux voies, une primaire et une secondaire. Des longs ARN précurseurs simple-brin sont transformés en piRNAs matures de 25-30 nt qui sont chargés dans les protéines PIWI. Une fois chargée avec le piRNA, PIWI se lie aux transcrits de transposons ayant la séquence complémentaire et les clive en générant deux ARN, dont l'un peut être chargé dans une nouvelle protéine PIWI, produisant un piRNA secondaire. Des études génétiques ont identifié plusieurs facteurs protéiques essentiels à ce processus. Certains de ces facteurs sont des hélicases à ARN dont le rôle spécifique reste inconnu, principalement parce que ce sont des enzymes dynamiques et l'identification de leurs cibles et leurs partenaires protéiques avec des approches biochimiques standards est difficile.

Dans la première partie de cette thèse je décris comment l'introduction d'une mutation ponctuelle dans la boîte DEAD de l'hélicase à ARN Vasa (DEAD à DQAD) peut bloquer son activité *in vivo* et figer le complexe transitoire de biogenèse qui contient VASA et les deux protéines PIWI responsables de la biogenèse secondaire.

La résolution de la structure de VASA^{DQAD} en complexe avec l'ATP a révélé les détails moléculaires de cette inhibition et a expliqué le phénotype observé *in vivo*. VASA^{DQAD} a un taux d'hydrolyse de l'ATP réduit, car après hydrolyse, le phosphate libre est bloqué à l'intérieur du site actif en raison d'une liaison hydrogène supplémentaire formée avec la Gln (Q) mutée. La réduction de l'hydrolyse de l'ATP se traduit par une faible liaison à l'ARN mesurée par des expériences biophysiques. L'introduction de la même mutation chez

Résumé

l'homologue murin de VASA (MVH) produit un phénotype de dominant négatif où MVH^{DQAD} est agglutinée sur le complexe ribonucléoprotéine (RNP) contenant les protéines PIWI et les piRNAs.

Dans la deuxième partie de cette thèse, j'ai introduit la même mutation dans TDRD9, une autre hélicase à ARN impliquée dans la voie des piRNAs mais dont la fonction est inconnue. Nous avons d'abord exprimé, purifié TDRD9 et montré que la mutation dans son domaine hélicase DEVH à DQVH abolit complètement son activité ATPase sans impacter sur sa stabilité. Par la suite, nous avons généré une souris *Knock-in* et analysé son phénotype. Les souris *Knock-in* mâles sont stériles et présentent un blocage au début de la spermatogenèse qui est probablement une conséquence des dommages de l'ADN générés par l'activation des éléments transposables. Ces éléments, comme LINE-1, présentent un défaut de méthylation à leur *loci* génomiques, mais qui ne semble pas être contrôlé par la voie piRNA dans le mutant, étant donné que les protéines PIWIs sont correctement chargées avec les piRNAs dérivés de LINE-1. Fait intéressant, une baisse des piRNAs issus des éléments SINE est observée chez ce mutant, ce qui reflète probablement un rôle de TDRD9 dans le tri des transcrits primaires vers MILI et MIWI2 au cours de la méthylation *de novo* de l'ADN.

Dans l'ensemble, j'ai étudié le rôle moléculaire de deux hélicases à ARN dans la voie des piRNAs, j'ai élucidé le rôle de VASA et montré que l'activité ATPase de TDRD9 est essentielle pour la régulation des transposons au cours de la spermatogenèse de la souris.

Summary

PIWI proteins are expressed in the gonads (testis and ovary) of animals, where they associate with PIWI-interacting RNAs (piRNAs) and silence transposable elements, thus defending the integrity of the genome. Indeed, animal knock-outs of PIWI proteins display a loss of piRNAs and activation of transposon sequences with catastrophic consequences: potentially due to genome damage, germ cell development comes to a halt, resulting in infertility. Silencing is achieved either by piRNA-guided endonuclease activity of cytosolic PIWI protein or by recruitment of the transcriptional repression machinery on target genomic loci by nuclear PIWI. Biogenesis of piRNAs can be divided into a primary and a secondary pathway. The primary pathway describes how single-stranded RNA precursors are processed into mature 25-30 nt piRNAs and loaded into PIWI proteins. PIWI-loaded piRNAs bind and cleave complementary transposon transcripts generating two RNA products, one of which can be loaded into a new PIWI protein, generating a secondary piRNA. Different protein factors are essential in this process, as identified by genetic studies. Few of these factors are putative RNA helicases, their specific role however is unknown, mainly because RNA helicases are dynamic enzymes and identification of their targets and protein partners with standard biochemical approaches is challenging.

In the first part of this thesis I describe how the introduction of a point mutation in the DEAD box of the RNA helicase Vasa (DEAD to DQAD) can stall its activity *in vivo* and freeze a transient biogenesis complex that contains Vasa and the two PIWI proteins which are responsible for secondary biogenesis.

The crystal structure of Vasa^{DQAD} in complex with ATP revealed the molecular details of this inhibition and explains the phenotype observed *in vivo*. Vasa^{DQAD} shows a reduced ATP hydrolysis rate because after hydrolysis the free phosphate is blocked inside the active site due to an additional hydrogen bond formed with the mutated Gln (Q). Reduction in ATP hydrolysis is mirrored by an impaired RNA binding activity as measured by biophysical experiments. Introduction of the same mutation in the mouse homologue of Vasa (MVH) has a dominant-negative phenotype *in vivo* since heterozygote knock-in mice are male-infertile. Furthermore, the resulting MVH^{DQAD} protein is stalled on an RNP complex.

Summary

In the second part of this thesis I introduced the same mutation in TDRD9, another RNA helicase involved in the piRNA pathway, with an unknown function. First I expressed and purified TDRD9 and showed that a DEVH to DQVH mutation in its helicase domain completely abolishes its ATPase activity but does not affect its stability. Next, I created a knock-in mutation in the mouse genomic locus for *Tdrd9* and analysed the resulting phenotype in the mutant. Knock-in mice are male-sterile with an early arrest in spermatogenesis, which is probably a consequence of uncontrolled DNA damages generated by de-repressed transposon elements. These elements, like LINE-1, fail to be correctly methylated at their genomic loci in the *Tdrd9* mutant. Although *Tdrd9* is important for LINE-1 transposon silencing, it is likely not via a role in piRNA biogenesis since PIWI proteins are correctly loaded with LINE-1 derived piRNAs.

Overall, I investigated the molecular role of two RNA helicases in the piRNA pathway, elucidating the role of Vasa and showing that the ATPase activity of TDRD9 is essential for transposon regulation in mouse spermatogenesis.

Abbreviations

3' UTR	Three prime untranslated region
Ab	Antibody
A-bias	Adenine bias
Ago	Argonaute
ATP	Adenosine triphosphate
Aub	Aubergine
Bm	<i>Bombyx mori</i>
BoYb	Brother of Yb
BSA	Bovine serum albumin
casiRNA	Cis-acting RNA
cDNA	Complementary DNA
cenRNAs	Centromeric RNA
CNTR	Control
CTD	C-terminal domain
Dcr	Dicer
DGCR8	DiGeorge critical region 8
Dm	<i>Drosophila melanogaster</i>
DNMT1	DNA methyltransferase 1
dpc	Days post coitum
DPS	Dithiobis (succimidyl propionate)
dsRNA	Double-stranded RNA
DTT	Dithiothreitol
<i>E.coli</i>	<i>Escherichia coli</i>
E16.5	Embryonic stage 16.5
EDTA	Ethylenediaminetetraacetic acid
eIF4G	Eukaryotic translation initiation factor 4 gamma
EJC	Exon junction complex
Endo-siRNA	Endogenous siRNA
ES cells	Embryonic stem cells
ESRF	European Synchrotron Radiation Facility
Exo-siRNA	Exogenous siRNA
FL	Full-length
FLP	Flippase protein
FT	Flow through
GKT	Glycine-lysine-threonine
Gln	Glutamine
Glu	Glutamic acid
H3K9me2-3	Dimethylated and trimethylated histone H3 on lysine 9
HA	Influenza hemagglutinin epitope

Abbreviations

hAgo2	Human Ago2
H-bound	Hydrogen bound
HD	Helicase domain
HP1	Heterochromatin protein 1
HTX	High-throughput crystallization facility
IAP	intracisternal A-particle
IF	Immunofluorescence
IP	Immunoprecipitation
Kb	Kilobase
KI	Knock-in
KO	Knock-out
L1	LINE1
LINE	Long Interspersed Nuclear Elements
LOQS	Loquacious
LTR	Long terminal repeats
MID	Middle domain
miRNA	Micro RNA
Mov10	Moloney leukemia virus 10
MOV10L1	Mov10Like-1
MVH	Mouse vasa homologue protein
ncRNA	Non coding RNA
NLS	Nuclear localization signal
nt	Nucleotides
NTD	N-terminal domain
OCT	Optimal cutting temperature compound
OSC	Ovarian somatic cell line
<i>osk</i>	<i>Oskar</i>
P14	14 days after birth
PAZ domain	Piwi/Argonaute/Zwille domain
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PGCs	Primordial germ cells
P _i	Inorganic Phosphate
piRNA	Piwi-interacting RNA
PIWI	P-element induced Wimpy Testis
PMSF	phenylmethylsulfonyl fluoride
PolII	Polymerase II
pri-miRNA	Primary miRNA
PTGS	Post-transcriptional gene silencing
PTS	piRNA trigger sequence
RDC	Rhino-Deadlock-Cutoff complex

Abbreviations

RdRC	RNA-directed RNA polymerase complex
RecA	Recombinase A
RISC	RNA-induced silencing complex
RITS	RNA-induced transcriptional silencing complex
RLC	RISC loading complex
RNAi	RNA interference
RNP	Ribonucleoprotein complex
RT	Room temperature
RT-PCR	Reverse Transcriptase-PCR
sDMAs	Symmetrically dimethylated arginines
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SF1	Super family 1
Sf21	Spodoptera frugiperda 21
SINE	short interspersed nuclear elements
siRNA	Small interfering RNA
SN	Supernatant
SoYb	Sister of Yb
<i>spn-E</i>	<i>Spindle-E</i>
SPR	Surface plasmon resonance
SSC	Saline Sodium Citrate
ssRNA	Single-stranded RNA
Strep	Streptavidin
tasiRNA	Trans-acting RNAs
Tdrd1	Tudor domain-containing protein 1
TE	Transposon elements
TEV	Tobacco Etch Virus protease
TGS	Transcriptional gene silencing
TLC	Thin Layer Chromatography
U-bias	Uridine bias
UTR	Untranslated region
UV	Ultraviolet
Vret	Vreteno
W _a	Catalytic water
WB	Western blot
Zuc	Zucchini
μL	Microliter
μM	Micromolar
3' UTR	Three prime untranslated region
Ago	Argonaute
Aub	Aubergine
Bm	<i>Bombyx mori</i>
BoYb	Brother of Yb

Abbreviations

casRNA	Cis-acting RNA
cDNA	Complementary DNA
cenRNAs	Centromeric RNA
CTD	C-terminal domain
Dcr	Dicer
DGCR8	DiGeorge critical region 8
Dm	<i>Drosophila melanogaster</i>
DNMT1	DNA methyltransferase 1
dpc	Days post coitum
DPS	Dithiobis (succinimidyl propionate)
dsRNA	Double-stranded RNA
eIF4G	Eukaryotic translation initiation factor 4 gamma
EJC	Exon junction complex
Endo-siRNA	Endogenous siRNA
ES cells	Embryonic stem cells
ESRF	European Synchrotron Radiation Facility
Exo-siRNA	Exogenous siRNA
GKT	Glycine-lysine-threonine
Gln	Glutamine
Glu	Glutamic acid
H3K9me2-3	Dimethylated and trimethylated histone H3 on lysine 9
HA	Influenza hemagglutinin epitope
HTX	High-throughput crystallization facility
IP	Immunoprecipitation
LTR	Long terminal repeat
MID	Middle domain
miRNA	Micro RNA
Mov10	Moloney leukemia virus 10
MOV10L1	Mov10Like-1
MVH	Mouse vasa homologue protein
nt	Nucleotides
NTD	N-terminal domain
OCT	Optimal cutting temperature compound
<i>osk</i>	<i>Oskar</i>
P14	14 days after birth
PAZ domain	Piwi/Argonaute/Zwille domain
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PGCs	Primordial germ cells
piRNA	Piwi-interacting RNA
PIWI	P-element induced Wimpy Testis
pri-miRNA	Primary miRNA
PTGS	Post-transcriptional gene silencing

Abbreviations

RecA	Recombinase A
RISC	RNA-induced silencing complex
RITS	RNA-induced transcriptional silencing complex
RNAi	RNA interference
RT-PCR	Reverse Transcriptase-PCR
sDMAs	Symmetrically dimethylated arginines
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SF1	Superfamily 1
Sf21	<i>Spodoptera frugiperda</i> 21
siRNA	Small interfering RNA
SoYb	Sister of Yb
<i>spn-E</i>	<i>Spindle-E</i>
SPR	Surface Plasmon resonance
ssRNA	Single-stranded RNA
tasiRNA	Trans-acting RNAs
Tdrd1	Tudor domain-containing protein 1
TE	Transposon elements
TEV	Tobacco Etch Virus protease
TGS	Transcriptional gene silencing
Vret	Vreteno
W _a	Catalytic water
WB	Western blot
Zuc	Zucchini
μL	Microliter
μM	Micromolar

Table of contents

Résumé	3
Summary.....	5
Abbreviations.....	7
Table of contents	12
1 - INTRODUCTION	16
1.1 Small non-coding RNA.....	17
1.2 The Argonaute protein family	18
1.3 The siRNA pathway.....	20
1.3.1 Exo-siRNA: biogenesis and function.....	20
1.3.3 siRNAs in transcriptional gene silencing.....	22
1.4 The micro RNA biogenesis pathway.....	24
1.4.1 Target regulation of miRNA	25
1.5 The piRNA pathway	28
1.5.1 Mouse PIWI proteins and early events in spermatogenesis.....	31
1.5.2 The nature and origin of piRNA sequences	34
1.5.3 piRNA biogenesis	35
1.5.4 Subcellular localisation of piRNA pathway and additional factors required.....	42
1.6 RNA helicases	43
1.6.1 DEAD-box RNA helicases	44
1.6.2 DExH box RNA helicases and other families.....	46

Table of contents

1.7 Aim of the thesis.....	49
2 - MATERIALS AND METHODS.....	50
2.1 Clones and plasmids	51
2.2 Cell lines.....	53
2.3 <i>In vivo</i> studies.....	54
2.3.1 Mouse strains and genotyping.....	54
2.3.2 Antibodies	61
2.3.4 Western blots	62
2.3.5 Immunoprecipitations	63
2.3.6 RNA extraction and Small RNA library preparation.....	65
2.3.7 Northern blot.....	66
2.3.8 Southern blot.....	67
2.3.9 Histology and Immunofluorescence of mouse testes sections.....	68
2.4 Expression of recombinant proteins and crystallization.....	69
2.4.1 Recombinant protein production in insect cells.....	69
2.4.2 Recombinant expression in <i>E.coli</i> and Crystallization of BmVasa ^{DQAD}	71
2.5 Biochemical <i>in vitro</i> assays.....	74
2.5.1 ATPase assay	74
2.5.2 UV cross-linking assays.....	74
2.5.3 <i>In vitro</i> footprinting assay	75
2.5.4 Kinetic studies with Surface plasmon resonance (SPR)	75

3-THE ROLE OF VASA IN THE piRNA PATHWAY	78
Résumé.....	79
3.1 Introduction	80
3.1.1 The role of Vasa in embryogenesis: an old story.....	80
3.1.2 A piRNA amplifier complex is assembled on the DEAD box RNA helicase Vasa82	
3.2 Results	84
3.2.1 Protein purification and biochemical studies	84
3.2.2 Vasa ^{DQ} is an ATPase product release trap mutant	89
3.2.3 Insights into the ATP hydrolysis mechanism of DEAD-box helicases	93
3.2.4 Blocking phosphate release impairs RNA binding kinetics of Vasa	95
3.2.5 <i>In vivo</i> studies	98
3.2.6 Purification of the recombinant Amplifier complex (Vasa-Siwi-Ago3).....	99
3.2.7 DQAD mutation in mouse Vasa homologue (MVH) blocks a pachytene piRNA biogenesis complex.....	103
3.3 Discussions and perspectives	106
4 - FUNCTIONAL CHARACTERIZATION OF TDRD9 IN THE piRNA PATHWAY	109
Résumé.....	110
4.1 Introduction	111
4.2 Results	114
4.2.1 Generation and validation of antibodies against mouse TDRD9.....	114
4.2.2 TDRD9 localisation in mouse testes.....	116
4.2.3 Immunoprecipitation of TDRD9 and search for interactors	119
4.2.4 Mouse TDRD9 is an active ATPase	121

Table of contents

4.2.5 Loss of TDRD9 ATPase activity results in male-specific infertility in mice	126
4.2.6 Tdrd9 catalytic-dead mutant mice fails to suppress LINE1 retrotransposons	129
4.2.7 TDRD9 mutant protein is not detectable in adult mouse testes.	130
4.2.8 Impact on piRNA biogenesis in the <i>Tdrd9</i> catalytic-dead mutant mice	132
4.3 Discussions and perspectives	137
Concluding remarks.....	140
List of references.....	142
Acknowledgments.....	159
Appendix.....	160

1 - INTRODUCTION

1.1 Small non-coding RNA

The majority of our genome is transcribed into non-coding RNAs, many of which are processed into small regulatory molecules of 20 to 40 nucleotides (nt). The first small non-coding RNA discovered was Lin-4, an RNA element that down-regulates the translation of its target messenger RNA (mRNA), Lin14, by base pairing to its three prime untranslated region (3' UTR) (Lee et al., 1993; Wightman et al., 1993). Lin-4 is now considered as the founder of the micro RNA family (miRNA), a class of endogenous regulators that is predicted to regulate more than half of our protein-coding genes. This process was called RNA interference (RNAi) and it was later found to operate in a wide variety of organisms, including insects, flatworms, plants and fungi (Cogoni and Macino, 1999; Kennerdell and Carthew, 1998; Sanchez Alvarado and Newmark, 1999; Waterhouse et al., 1998).

Small non-coding RNAs acquire their silencing potential when they are loaded into Argonaute proteins, with whom they form the RNA induced silencing complex (RISC). Argonaute is the core of the RISC complex and cleaves a complementary mRNA target or blocks its translation, conferring resistance to exogenous parasitic nucleic acids or regulating endogenous gene expression (Fire et al., 1998; Hammond et al., 2000; Zamore et al., 2000). Small non-coding RNAs are involved in a wide variety of biological processes such as antiviral defence, development and maintenance of genomic stability (Ghildiyal and Zamore, 2009). RNAi can also be triggered experimentally by introduction of a double-strand RNA (dsRNA) into a cell and it has proved to be an essential tool to analyse loss-of-function phenotypes of specific genes. RNAi has great potential in medicine for the *in vivo* inactivation of gene products linked to human disease progression and pathology.

Three main families of small non-coding RNAs are currently identified, based on their differences in biogenesis and functions: small-interfering RNA (siRNA), microRNA (miRNA) and PIWI-interacting RNA (piRNA) (Ghildiyal and Zamore, 2009). piRNA is the latest discovery within the class of small RNAs whose biogenesis and function are still nebulous in many aspects. In this chapter, I will summarize the three major small RNA pathways with a special focus on piRNAs.

1.2 The Argonaute protein family

RNAi and all the related small RNA-mediated silencing pathways exert their function through an RNA induced silencing complex (RISC) with an Argonaute protein as its core component (Hammond et al., 2000). Argonaute directly interacts with the small RNA that provides the target specificity, whereas additional proteins in the complex can extend or modify the function of RISC.

Phylogenetic analyses identified three subclades of Argonaute proteins: AGO, WAGO and PIWI (Figure 1.1A). The AGO proteins, which are found in all organisms including archaea, bacteria and plants, are involved in gene regulation, anti-viral defence and genome organization (Joshua-Tor and Hannon, 2011). In animals, they are ubiquitously expressed and bind to miRNA and siRNA. The PIWI proteins are found in diverse eukaryotic lineages (although not in plants or fungi) and in sexually reproducing metazoan, their expression is restricted to the gonads (Grimson et al., 2008). They bind piRNAs generated from long single-strand RNA precursors and generally function in repressing transposon activity (Siomi et al., 2011; Vagin et al., 2006). Finally the WAGO proteins are found only in worms, with 18 different Wago proteins encoded in the *C. elegans* genome.

Argonaute proteins are composed of four well conserved domains : an N-terminal domain (N), a Piwi/Argonaute/Zwille domain (PAZ), a middle domain (MID) and a P-element induced Wimpy Testis domain (PIWI) organized in a bilobate architecture, with the MID and PIWI domains forming one lobe, and the N-terminal and PAZ domains constituting the other (Figure 1.1C) (Song et al., 2004).

1-Introduction

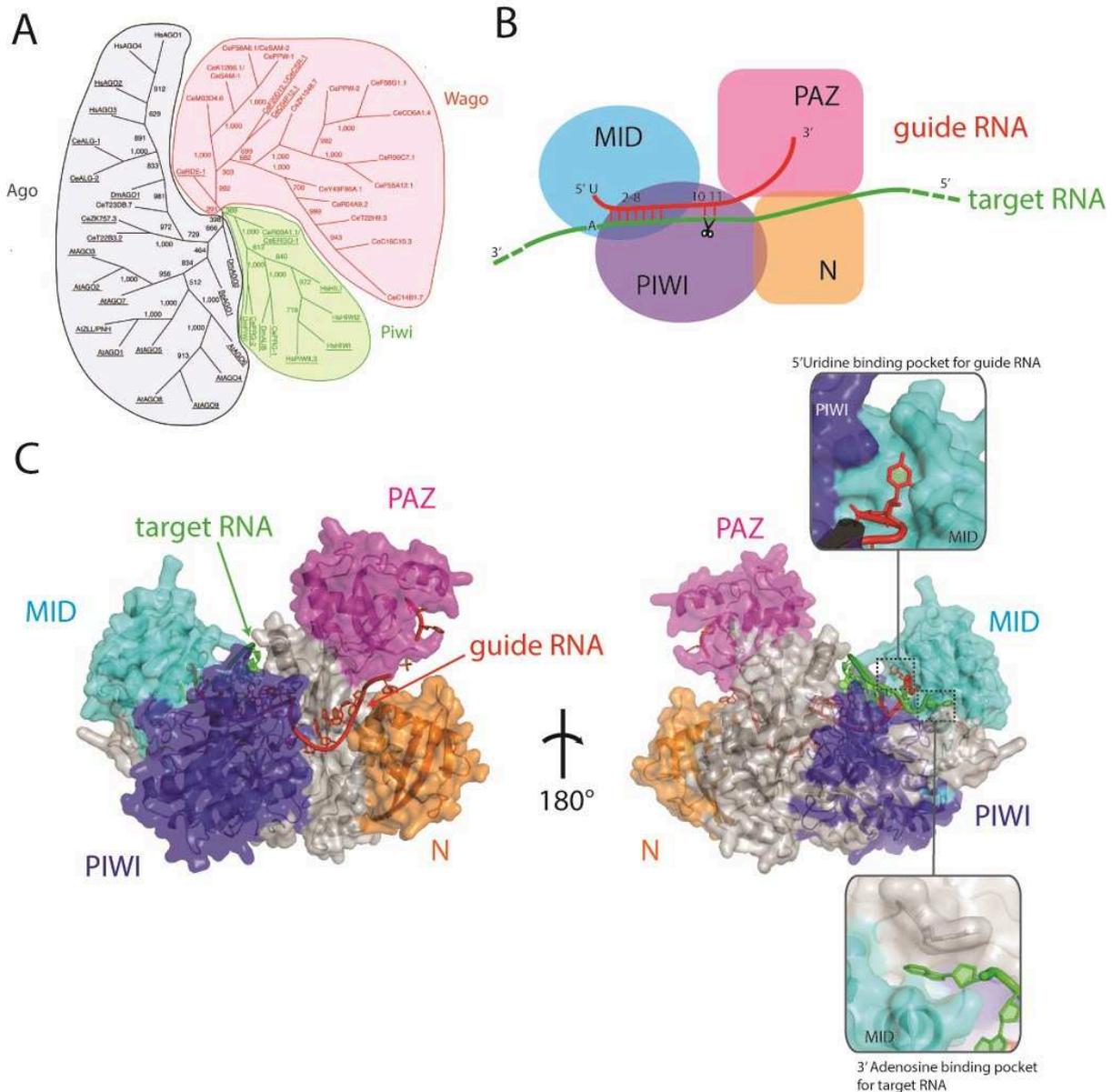


Figure 1.1 Argonaute proteins. (A) Phylogenetic tree of Argonaute protein family (adapted from Joshua-Tor and Hannon, 2011). (B) Domains organization of Argonaute proteins: the guide RNA (e.g. miRNA) binds its target through the seed region (nt 2-8), the PIWI domain slices between the nucleotides of the target complementary to nt 10-11 of the guide RNA. (C) Structure of human Ago2 in complex with a guide RNA (red) and an RNA target (green), the four domains are shown in different colours, linker regions between the main domains are depicted in grey. Details of the binding pocket of the 3' end of the target RNA (A-bias) and the 5' end of the guide RNA (U-bias) are shown (PDB: 4W5N); (Schirle and MacRae, 2012).

The PAZ domain binds the 3' end of the guide RNA (Song et al., 2004), whereas the MID and the PIWI domain form a binding pocket that accommodates its 5' end (Figure 1.1B). The binding affinity is higher if the first nucleotide at the 5' end of the guide RNA is an adenosine or a uridine, since these nucleotides fit better into the binding pocket (due to Hydrogen-bound interactions with specific residues) (Cora et al., 2014; Frank et al., 2010;

1-Introduction

Schirle and MacRae, 2012). This intrinsic propriety of Argonautes may explain why miRNAs have a strong A and U bias at their 5' end *in vivo* (Lewis et al., 2005). The rest of the guide RNA contacts the protein through sugar-phosphate backbone interactions, with nucleotides 2 to 8 (the seed region of the miRNA) exposed at the exterior of the protein, ready to pair with a complementary RNA target. Recently, the structure of hAgo2 in complex with a miRNA and an RNA target has been solved (Schirle et al., 2014). Interestingly, the complementary target strand accommodates its 3' end in a narrow pocket between the MID and the PIWI domain which usually shows a preference for adenosine nucleotides, explaining why miRNA targets are bias by 1A at their 3' end *in vivo* (Lewis et al., 2005) (Figure 1.1C).

The PIWI domain has an RNase H-like fold which confer to some components of the family the ability to slice a complementary RNA/DNA target. First structural studies of prokaryotic Ago proteins identified a conserved Asp-Asp-His (DDH) motif as the metal ion-coordinated catalytic triad which is responsible for slicer activity (Song et al., 2004). Mutations to alanine of any of these residues abolish the slicer activity of Ago2 *in vitro* and *in vivo* (Liu et al., 2004; Rivas et al., 2005). More recently, a fourth residue was found to be important for the catalytic activity of the protein, a Glu present in a loop that folds back into the active site of the Ago and complete the DEDD catalytic tetrad that is also found in RNase H (Nakanishi et al., 2012). This motif is quite conserved amongst the entire protein family but not every Argonaute has slicer activity. For example, within the four human Ago proteins (Ago1-4), only Ago2 has a proven slicer activity *in vitro* (Meister et al., 2004). The target is always sliced at the phosphodiester bond between the nucleotides that pair with nucleotide 10 and 11 of the guide RNA (Figure 1.1B). The ability to slice a target can be correlated to the way an Argonaute functions *in vivo*. For example fly Ago2 loads siRNA and silences mRNA transcripts through target slicing. Ago1 binds mainly miRNA and represses its mRNA targets by blocking mRNA translation, thus its slicer activity is dispensable (Jinek and Doudna, 2009).

1.3 The siRNA pathway

1.3.1 Exo-siRNA: biogenesis and function

dsRNAs can be introduced experimentally in cell cultures or in a living organism to knock-down the expression of specific genes. Once entered the cell, the dsRNA is recognized by the dsRNA specific ribonuclease Dicer and cleaved into a shorter 21nt long dsRNA with a

1-Introduction

2nt overhang at its 3' end (Bernstein et al., 2001) (Figure 1.2A). Following slicing, Dicer and Ago2 form the RISC-loading complex (RLC) with another RNA binding protein which is called R2D2 in *Drosophila* (Liu et al., 2003). The dsRNA is then loaded into Argonaute which keeps only one of the two strands, the guide RNA, whereas the other one, the passenger, is cleaved and degraded. The strand with more thermodynamic stability at its 5' end becomes the guide RNA and the other is destined to be the passenger (Schwarz et al., 2003). R2D2 senses this asymmetry and binds the siRNA end with the greatest double-strand character, orienting the RLC on the siRNA duplex (Tomari et al., 2004). The RLC complex recruits Ago2 to which it transfers the dsRNA. The guide RNA is further 2'-O-methylated at its 3' end by the S-adenosyl methionine-dependent methyltransferase Hen1, completing mature RISC assembly (Horwich et al., 2007). This modification is present on both miRNAs and siRNAs in plants as well as on piRNAs in animals and it protects the RNA from degradation by increasing its stability (Yu et al., 2005). Mature RISC can now bind a complementary RNA target and cleave it, inducing its degradation and leading to silencing.

In plants, an RNA dependent RNA polymerase (RdRP) can bind the 3' end of the cleaved mRNA and synthesize a complementary strand, generating a new dsRNA that can be processed into new siRNAs (Figure 1.2A). RdRP can also bind directly ssRNA which is very important for anti-viral defence in plants, where viral ssRNAs are converted into dsRNA by RdRPs and thus targeted to the siRNA pathway. Secondary siRNAs are particularly relevant in *C. elegans*, where primary siRNAs induce the synthesis of secondary siRNAs through the action of RdRP (Pak and Fire, 2007). The secondary pool of siRNAs can spread and lead to systemic silencing in the entire organism, both in *C. elegans* and plants (Baulcombe, 2004).

1.3.2 Endo-siRNA

Some organisms, like insects and plants, use RNAi as an anti-viral defence (Voinnet, 2008). In plants there are different classes of endogenous siRNA (endo-siRNAs) like the Cis-acting RNAs (casiRNAs), which originate from transposons and repetitive elements and promote heterochromatin formation by directing DNA methylation and histone modifications (Ghildiyal and Zamore, 2009). Endo-siRNA pathways are very complex in *C. elegans* (Ambros, 2003), which expresses as many as 27 Argonaute proteins. Secondary siRNAs are the most abundant ones and they are produced through the action of RdRPs.

1-Introduction

Endo-siRNAs are as well expressed in mice but so far they have been identified in only a limited number of tissues, e.g. in female germ cells. Mice with an oocyte-specific deletion of *Dicer* or *Ago2* are infertile due to defects during meiosis I (Murchison et al., 2007; Tam et al., 2008; Tang et al., 2007; Watanabe et al., 2008). These defects are most probably due to lack of siRNA and not miRNA, since the deletion of a miRNA biogenesis factor *Dgcr8* have no discernible phenotype (Suh et al., 2010). These endo-siRNAs arise from dsRNAs generated from interactions between gene and pseudogene transcripts, from overlapping transcription units and from transcripts that can form long hairpins (Czech and Hannon, 2011). Female-specific siRNAs are also involved in transposon defence, since depletion of *Dicer* induces overexpression of different transposon elements, like the Short Interspersed Nuclear Element (SINE) (Murchison et al., 2007; Watanabe et al., 2008).

1.3.3 siRNAs in transcriptional gene silencing

Usually siRNAs silence their targets by slicing the mRNA but they can also be loaded into non-slicer Ago protein and trigger silencing by blocking translation, employing a mechanism similar to miRNA's. These silencing mechanisms are referred to as post-transcriptional gene silencing (PTGS). siRNAs can also act in the nucleus where they can mediate chromatin-dependent gene silencing, e.g. transcriptional gene silencing (TGS). The role of siRNA in chromatin remodelling is studied particularly well in unicellular organisms like the budding yeast *S. pombe* (Moazed, 2009; Volpe et al., 2002) (Figure 1.2B). *S. pombe* contains single genes encoding for Argonaute (Ago1), Dicer (Dcr1) and RdRP proteins (Rdp1). The centromeric heterochromatin of *S. pombe* is characterized by enrichment of dimethylated and trimethylated histone H3 on lysine 9 (H3K9me2-3) that is opposed to euchromatic regions enriched with methylated histone H3 on lysine 4 (H3K4me). Non-coding RNAs, called centromeric transcripts (cenRNAs) are actively transcribed from this region even when the chromatin is highly compacted. An RNA-induced transcriptional silencing complex (RITS) that contains Ago1, loaded with centromeric siRNA, is recruited through RNA base-pairing to the nascent RNA transcript. The RITS also contains the chromodomain-binding protein Chp1 and the glycine and tryptophan (GW)-motif-containing protein Tas3 (Verdel et al., 2004). The RITS binds to H3K9me2-3 with Chp1 and recruits the RNA-dependent RNA polymerase complex (RDRC). Using the cenRNA as a substrate the polymerase contained within the complex synthesizes a dsRNA that is subsequently sliced by Dicer to generate new siRNAs (Sugiyama et al., 2005). These siRNAs can now enter a

1-Introduction

positive feedback loop by generating new RITS complexes. A methyltransferase Clr4 (a subunit of the CLRC complex) is also recruited to the site by the RDRC-Dicer complex and enhances the H3K9 methylation. Newly methylated histones can thus recruit the two HP1 proteins Swi6 and Chp2 that promote the associations of new RITS complexes (Moazed, 2009).

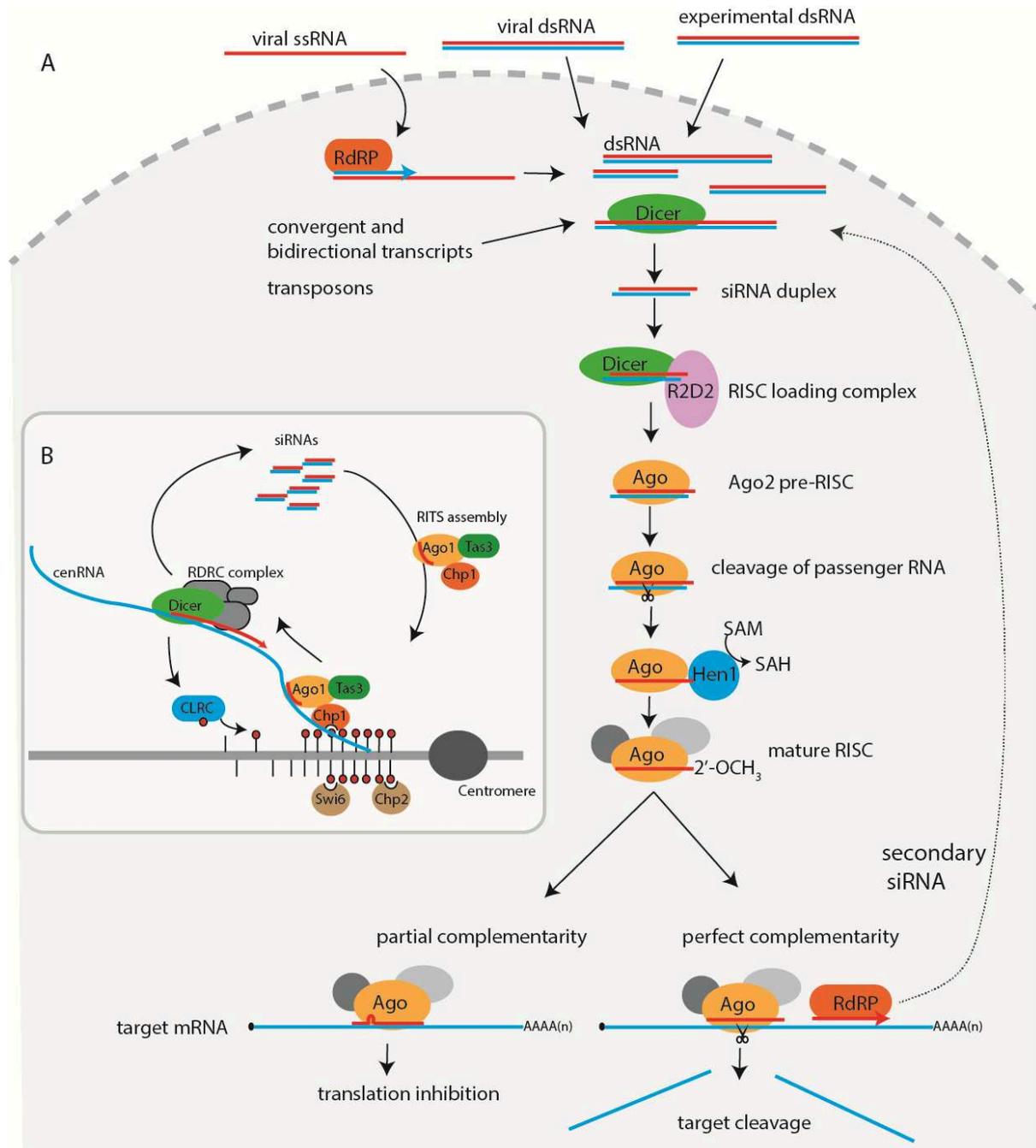


Figure 1.2: The siRNA biogenesis pathway. (A) Different sources of dsRNAs trigger the siRNA pathway: first Dicer produces 22nt dsRNA with a 2nt 3'end overhang. R2D2 (in fly) orients the RISC loading complex, so that the appropriate strand is loaded into Ago. Loaded Ago cleaves the passenger

1-Introduction

strand and forms the mature RISC complex that silences complementary RNA targets. **(B)** The nuclear siRNA pathway in *S. pombe* controls the epigenetic silencing of centromeric chromatin; red dots indicate heterochromatin marks (H3K9me₂₋₃). Ago1-loaded siRNAs bind nascent cenRNA transcripts, recruiting the RDRD complex, thus producing new dsRNAs that are later converted into siRNAs. The CLRC complex promotes methylation of H3K9 (the centromere is not in scale).

1.4 The micro RNA biogenesis pathway

miRNAs are endogenous ~22nt long RNAs that can regulate the expression of target mRNAs by binding to them with different degrees of strand complementarity. Animal miRNAs bind their target with partial complementarity, whereas plant miRNAs naturally recognize targets with precise base pairing (Ambros, 2003). The first miRNA, *lin-4*, was identified in *C. elegans* during a screen for genes required for embryonic development (Lee et al., 1993; Wightman et al., 1993). In the following decades hundreds of miRNAs were identified by small-RNA cloning in humans, flies and worms. miRBase, the registry that coordinates miRNA nomenclature, counts almost 36.000 mature miRNA products, found in 223 different species (release 21) (Kozomara and Griffiths-Jones, 2014), tracing miRNA origins at the early evolution of Metazoa (Grimson et al., 2008).

miRNAs are often transcribed from transcription units containing multiple miRNA genes, but they are also found as independent single genes. They are transcribed by RNA polymerase II into polyadenylated and capped pri-miRNA: a stem loop structure with long ssRNA branches (Cai et al., 2004; Lee et al., 2004a) (Figure 1.3). The pri-miRNA is first processed into a shorter 60-80nt long pre-miRNA by the RNase III endonuclease Drosha (Lee et al., 2004b) (fig.3). Drosha forms the Microprocessor complex with its protein partner –DiGeorge critical region 8 (DGCR8) (Gregory et al., 2004; Landthaler et al., 2004). The resulting pre-miRNA has a stem-loop structure, a 2nt overhang at the 3'-end and a 5'-phosphate group. Some miRNAs can skip Drosha processing since they are directly spliced out from pre-miRNA-like-introns, called mirtons. After 2'-5' debranching they are converted into mature pre-miRNAs that can re-enter the miRNA biogenesis pathway (Okamura et al., 2007; Ruby et al., 2007). The pre-miRNA is then exported to the cytoplasm by exportin5 (Yi et al., 2003) where it is processed by Dicer (Bernstein et al., 2001; Ketting et al., 2001). As for siRNA, Dicer needs a co-factor to process pre-miRNA: TRBP in mammals and Loquacius, LOQS, in flies (Chendrimada et al., 2005). Dicer cleavage generates a duplex, consisting of miRNA and miRNA*, which are similar to siRNA and the siRNA passenger, respectively. The choice of which of the two strands will become the mature miRNA follows

1-Introduction

similar thermodynamic criteria as for siRNA, but in some cases both strands can give rise to mature miRNAs. miRNAs are then loaded into Ago (Ago1-4 in humans and Ago1 in flies) and mature miRISCs are formed.

1.4.1 Target regulation of miRNA

miRNAs regulate gene expression by binding their mRNA targets and repressing their translation. miRNAs usually bind to the 3'-UTR of the mRNA with nucleotides 2-8, a region known as the 'seed' (Figure 1.1B). miRNAs and their targets are usually mismatched at positions 10-12, preventing endonucleolytic cleavage, and the pairing is reinforced by additional matches beyond the seed region (Bartel, 2004; Lewis et al., 2005). In plants, most of the miRNAs bind to the target with perfect complementarity, promoting its cleavage and degradation, with an siRNA-like mechanism. In contrast, most animal miRNAs have discontinuous pairing to their targets which are then silenced through translational repression or mRNA decay. It is still controversial whether these two mechanisms act in sequence, in parallel or whether they are connected at all (Filipowicz et al., 2008; Huntzinger and Izaurralde, 2011).

Numerous evidence supports the theory that animal miRNAs can trigger mRNA degradation, since altering miRNA levels induces changes in the abundance of many of their target mRNAs (Lim et al., 2005). An mRNA that is competent for translation, possesses a 5'-cap structure and a 3' poly(A) tail. mRNA degradation is achieved through removal of the poly(A)-tail followed by decapping and 5' to 3' exonucleolytic digestion (Behm-Ansmant et al., 2006) (Figure 1.3). This is mediated by the action of the GW182 protein, which associates to the PIWI domain of the Argonaute protein through glycine-tryptophan repeats in its N-terminus (Eulalio et al., 2009). GW182 interacts with the poly(A)-binding protein (PABP), reducing its binding levels to the poly(A)-tail, and recruiting the CAF1-CCR4-NOT1 deadenylase complex (Behm-Ansmant et al., 2006; Giraldez et al., 2006). Following deadenylation, the mRNA is decapped by the DCP complex and finally degraded by the 5'-to-3' exonuclease XRN1 (Braun et al., 2012).

It is also evident that miRNA can regulate mRNA by blocking translation without affecting its stability. PABP binds the poly(A) tail and interacts with the translation-initiation factor 4G (eIF4G) which is associated with the cap structure through interaction with the cap-binding protein eIF4E. These proteins form the eIF4F complex, a multiprotein complex

1-Introduction

required for the recruitment of the 43S pre-initiation complex. miRNA interferes with the function of the eIF4F complex, thus blocking translation most probably at the initiation step (Fukaya and Tomari, 2012; Humphreys et al., 2005; Mathonnet et al., 2007; Pillai et al., 2005; Thermann and Hentze, 2007).

Recent studies using time-course analysis, suggest that miRNA silencing is firstly established by obstructing translation initiation and subsequently reinforced by mRNA destabilization (Figure 1.3), but a final agreement has not been reached yet (Djuranovic et al., 2012; Meijer et al., 2013; Wilczynska and Bushell, 2015).

1-Introduction

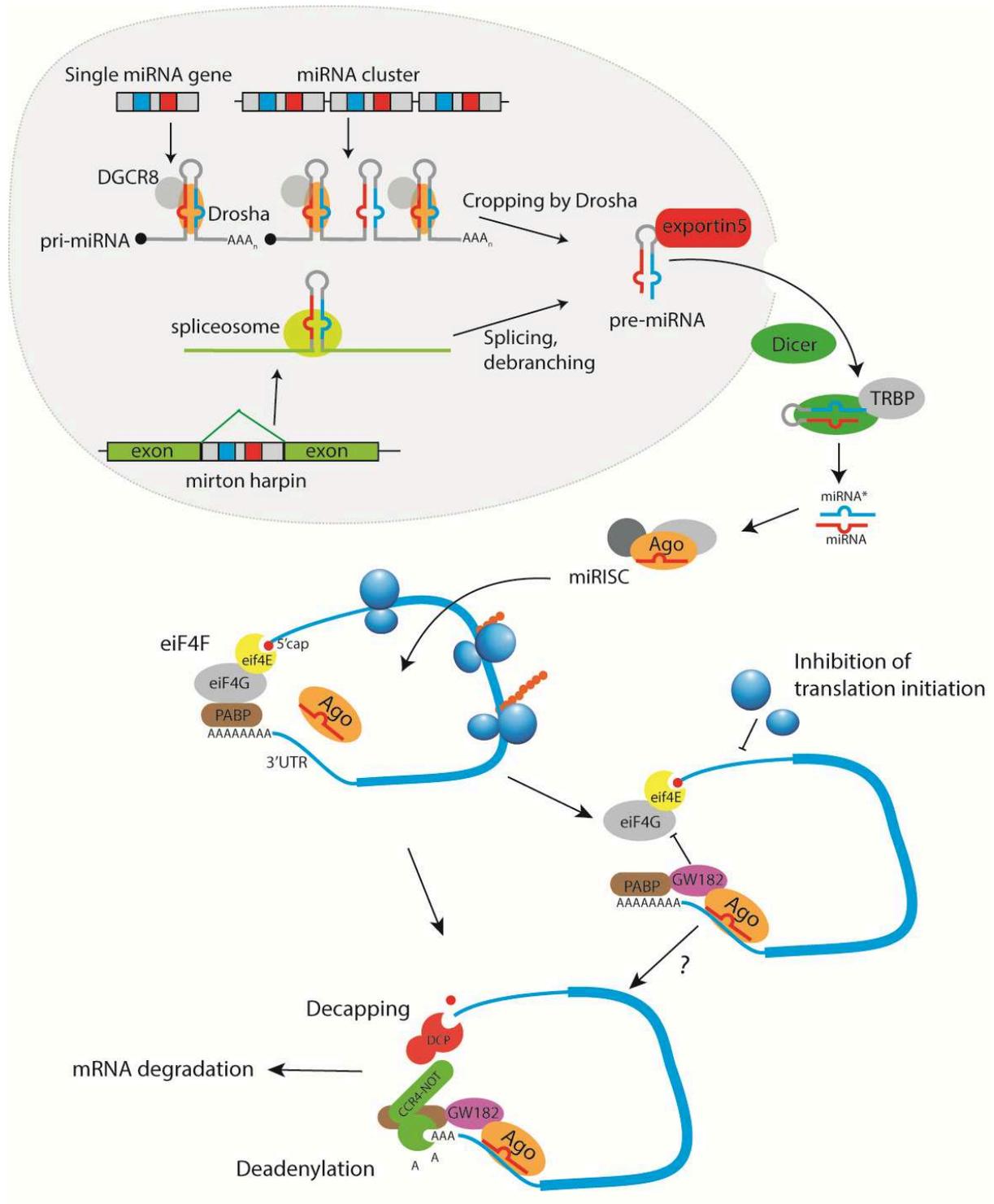


Figure 1.3 The miRNA pathway. miRNAs are produced from endogenous genes and exported into the cytoplasm after Drosha-processing. Here they are processed by Dicer into mature miRNAs that silence mRNA by translation inhibition (most probably at the initiation step) or by mRNA degradation. The interconnection between these events is not clear yet (question mark).

1.5 The piRNA pathway

piRNAs are a class of small non-coding RNAs that are loaded into the PIWI subclade of Argonaute proteins. PIWI proteins and piRNA are conserved in a very large set of eukaryotes, from sponges to humans (Grimson et al., 2008). The most accepted function of the piRNA pathway is to protect the genome stability of germ cells from transposon activity. These parasitic elements play essential roles in the evolution of species but their uncontrolled activity can cause mutations and translocations that threaten genome integrity (Slotkin and Martienssen, 2007). Loaded PIWI proteins form piRISC complexes that can bind their targets with sequence complementarity and silence transposons post-transcriptionally through direct cleavage of the transposon mRNA, or with TGS, by inducing epigenetic silencing on transposon genomic loci (Pillai and Chuma, 2012).

piRNAs were first described in *Drosophila* as repeat-associated small interfering RNA (rasiRNA) due to their sequence homology to repetitive elements (Aravin et al., 2001). When similar small RNAs were found associated with PIWI proteins in *Drosophila* ovaries and in mouse testes, they were re-defined as PIWI interacting RNAs (piRNA), since they are essentially different to the other small RNAs involved in RNAi (Aravin et al., 2006; Girard et al., 2006; Grivna et al., 2006; Saito et al., 2006; Vagin et al., 2006). piRNAs are slightly longer (24-31nt) than miRNA and siRNA, and they are processed from single-strand long RNAs in a Dicer-independent manner; furthermore they are tissue-specific, since in animals they can be found mainly in the gonads, whereas miRNA and siRNA are ubiquitously expressed in all tissues (Saito et al., 2006; Siomi et al., 2011; Vagin et al., 2006). As fly siRNAs, they accommodate a 2'-O-methyl modification at their 3' end, which is added by Hen1 (Horwich et al., 2007; Kirino and Mourelatos, 2007; Saito et al., 2007). In zebrafish this modification most likely protects piRNAs from uridylation and consequent degradation (Houwing et al., 2007; Kamminga et al., 2010)

The first studies of PIWI proteins were performed in *Drosophila*, where *Piwi* (P-element induced wimpy testes) was first identified as a gene implicated in germ stem cells self-renewal (Cox et al., 1998; Lin and Spradling, 1997). Adult *piwi* flies were found sterile with male and female gonads depleted of germline cells. Three PIWI proteins are expressed in *Drosophila* gonads: Piwi, Aub (Aubergine) and Ago3. Like *piwi*, mutation of *aub* also affects fertility of female and male flies, whereas mutations in *Ago3* affect fertility only in

1-Introduction

females (Cox et al., 1998; Li et al., 2009; Lin and Spradling, 1997). Since Ago3 is the only PIWI specifically involved in secondary biogenesis (see later), whereas Piwi and Aub are needed for primary biogenesis, it is possible that in testes only primary piRNA are produced. In the ovaries, all the three proteins are expressed in the nurse cells of the egg chamber and in the apical part of the testis. Piwi can also be found in the follicle cells, a thin layer of somatic cells that surround the egg chamber and helps to establish the anterior–posterior and dorsal–ventral axes (Cox et al., 1998) (Figure 1.4). Beside severe development defects in gametogenesis, depletion of PIWI proteins was found to upregulate certain transposons in the germ cells (Sarot et al., 2004; Savitsky et al., 2006), implying a role in transposon control.

Fly PIWI proteins have a specific cellular localisation in the egg chamber, with Piwi being the only nuclear protein found in the nurse cells of the egg chamber and in follicle somatic cells. Ago3 and Aub are cytoplasmic proteins, they both localise in a specific perinuclear granule called *nuage* in the cytoplasm of nurse cells (Lim and Kai, 2007). Aub and Ago3 mediate transposon silencing by direct cleavage of their targets, whereas Piwi silences its targets at transcriptional level, by recruiting enzymes that establish repressive heterochromatin at transposon genomic loci (Huang and Zhu, 2014; Le Thomas et al., 2013; Rozhkov et al., 2013)

1-Introduction

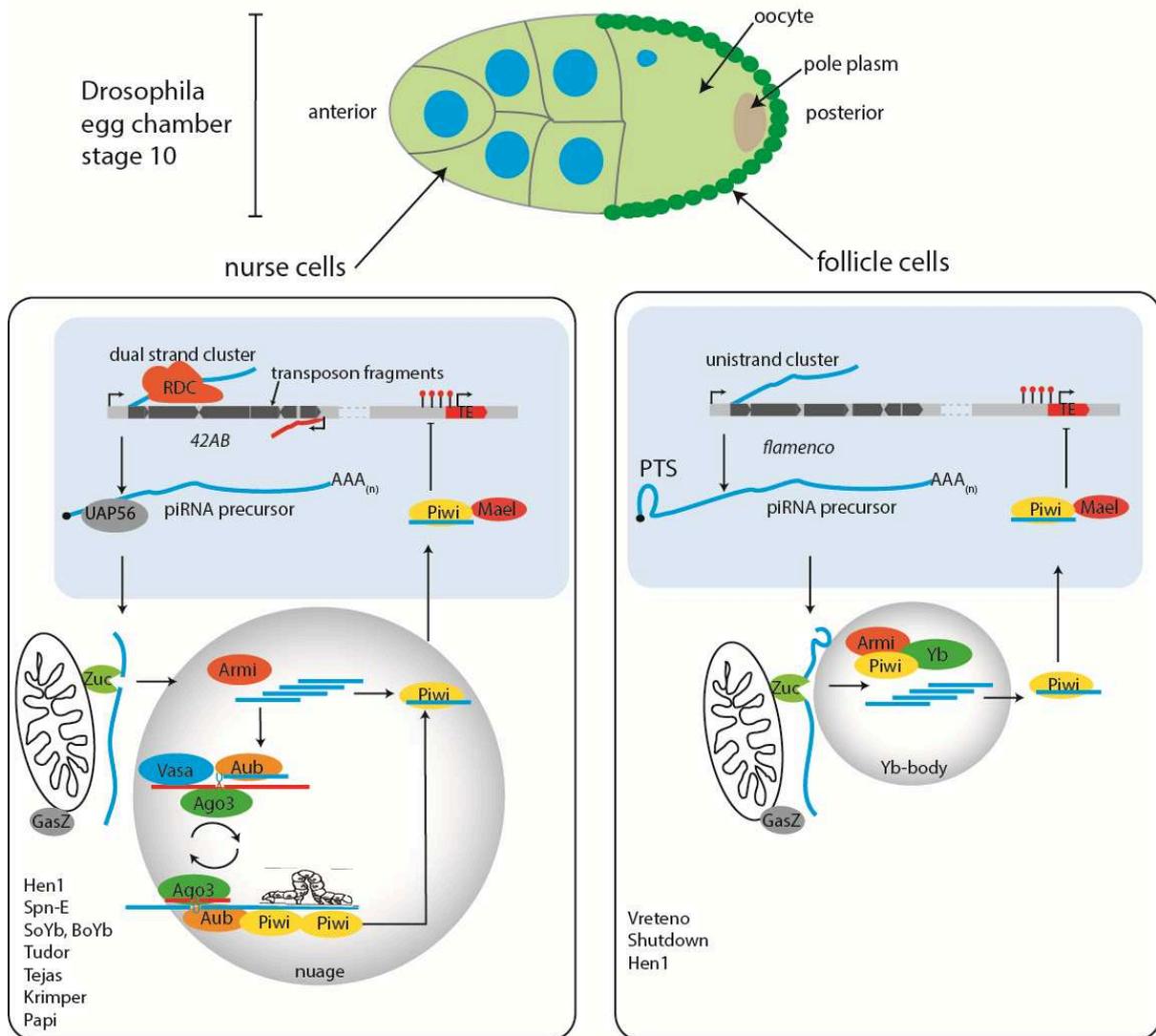


Figure 1.4 piRNA biogenesis pathway in *Drosophila*: Many egg chambers populate the germarium of *Drosophila*'s ovary. A single egg chamber at stage 10 is represented. Nurse cells express the primary and the secondary piRNA pathway. Piwi is loaded through the primary or the inchworm pathway (see Figure 1.6 for a better description) and silences TE in the nucleus. Follicle cells express only the primary pathway and Piwi is loaded into the Yb bodies. Additional factors involved are listed on the side.

1.5.1 Mouse PIWI proteins and early events in spermatogenesis

The mouse genome also encodes three PIWI-like genes: *Miwi*, *Mili* and *Miwi2* (also known as PIWI-like1, PIWIL-2 and PIWIL-4 respectively). *MILI* and *MIWI* are cytoplasmic proteins, localised in perinuclear granules, and silence transposons post-transcriptionally by cleaving transposon transcripts (De Fazio et al., 2011; Reuter et al., 2011). *MIWI2* is a nuclear protein and promotes *de novo* DNA methylation of long interspersed nuclear element and long terminal repeat (LTR) retrotransposons, like LINE1 and the IAP element, by directing CpG DNA methylation on their genomic loci (Kuramochi-Miyagawa et al., 2008). Depletion of *Mili*, *Miwi* and *Miwi2* leads to male-specific sterility, due to an arrest in spermatogenesis (Carmell et al., 2007; Deng and Lin, 2002; Kuramochi-Miyagawa et al., 2004). *Mili* is the only PIWI protein expressed in female gonads but its role in oogenesis is still not clear. Female fertility is not affected in a mouse *Mili* mutant, most likely because the endo-siRNA pathway that targets transposons in the oocyte renders the female piRNA pathway redundant (Tam et al., 2008; Watanabe et al., 2008).

Spermatogenesis is a tightly regulated cellular process as it involves cell proliferation, differentiation and morphogenesis. In mouse embryos, primordial germ cells (PGCs) start to be visible at 7.5dpc (days post coitum) when they start migrating and reach the genital ridges at 11.5dpc (Figure 1.5). Here, they arrest cell growth at 13.5 dpc, becoming fetal prospermatogonia and undergo *de novo* DNA methylation. This event is part of the epigenetic reprogramming that involves the removal of pre-existing bi-parental epigenetic information and the establishment of male-specific methylation patterns. The rationale is that the diploid spermatogonia gives rise to haploid sperms, thus the bi-parental marks must be erased and replaced with the “male-exclusive” patterns. Three days after birth (P3) spermatogonia restart proliferating and some cells duplicate into spermatogonial stem cells whereas some others enter a first wave of synchronous spermatogenesis at P10-P12. This cycle proceeds through meiosis, generating the first mature spermatozoa at P35 (Figure 1.5). In adult mouse testes, spermatogenesis is a cyclical pathway promoted by the continuous but asynchronous differentiation of spermatogonial stem cells into new spermatocytes. For this reason the first wave of spermatogenesis is particularly convenient to study the different phases of meiosis, particularly in the context of meiotic prophase I (P12-P19) and spermiogenesis (P19-P35).

1-Introduction

The three PIWI proteins are temporally and spatially regulated during spermatogenesis (Figure 1.5). MILI is expressed in two different stages on germ cells development: from primordial germ cells to post-natal Spermatogonial stem cells (from 12.5dpc to three days after birth, P3) and during meiosis (from zygotene stage until round spermatid) (Di Giacomo et al., 2013). MIWI2 is only expressed in embryonic stages starting at 14.5dpc and its expression levels drop completely at P3 (Carmell et al., 2007). MIWI expression commences only post-natal during meiotic pachytene phase, peaks in round spermatids and drops in post-meiotic cells (Deng and Lin, 2002). Depletion of *Mili* or *Miwi2* leads to spermatogenic stem cell arrest in prophase I of meiosis, whereas the arrest in *Miwi*-null mice takes place later, at the early spermatid stage (Carmell et al., 2007; Deng and Lin, 2002; Kuramochi-Miyagawa et al., 2004).

1-Introduction

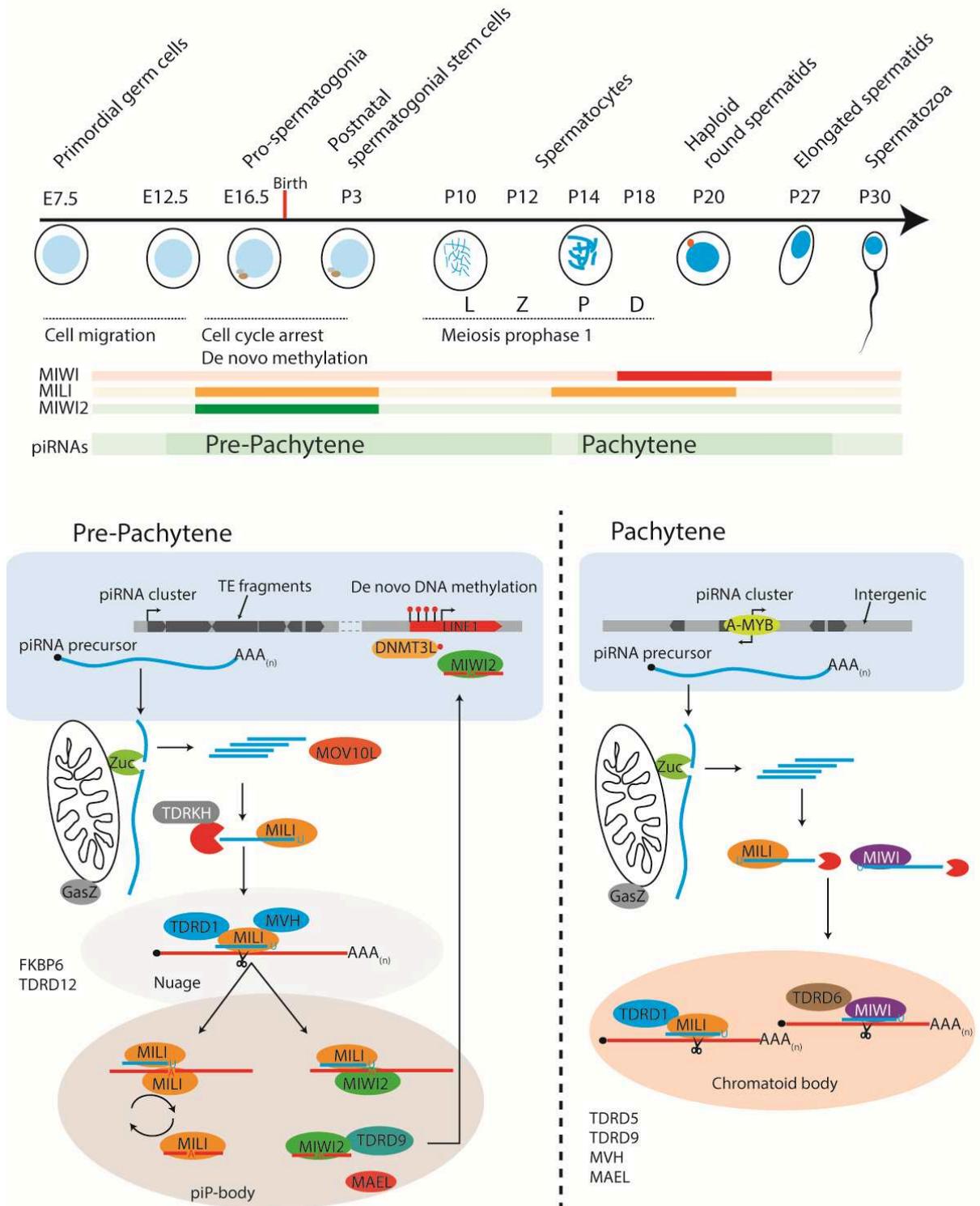


Figure 1.5 The mouse piRNA pathway. On top: The first synchronous cycle of spermatogenesis is shown, starting from germ cell migration in the gonads at embryonic stage 7.5 till mature spermatozoa 30 days post birth (P30). Prophase 1 of meiosis includes Leptotene, Zygotene, Pachytene and Diplotene phases. At the bottom: Pre-pachytene piRNAs are expressed from prenatal to P3. They are

1-Introduction

bound to MILI and MIWI2, which are involved in primary and secondary biogenesis. This ends with DNA *de novo* methylation of transposons driven by nuclear MIWI2. Primary biogenesis factors localise in the Nuage (intermitochondrial cement) and secondary factors in piP-bodies. Pachytene piRNAs are loaded into MILI and MIWI and localise in chromatoid bodies in late spermatocytes and round spermatids; besides reinforcing silencing of LINE1 by PTGS, their function is unknown. Additional factors implicated in the pathway are listed on the side.

1.5.2 The nature and origin of piRNA sequences

The large majority of *Drosophila* piRNAs map to transposon sequences (Brennecke et al., 2007). In mice however, the sequence composition of piRNAs changes during spermatogenesis. A first wave of piRNAs is expressed in prenatal testes and they are commonly called pre-pachytene piRNA. A second class of piRNAs is found during meiosis, at P14.5, and is called pachytene-piRNA (Figure 1.5). Only Pre-pachytene piRNAs are enriched in transposon sequences (approximately 80% in MIWI2) and they are actively involved in their silencing (Aravin et al., 2008). Pachytene piRNAs are predominantly derived from intergenic unannotated sequences (almost 80% in MILI), they are bound to MILI and MIWI and their role in spermatogenesis is yet to be fully understood (Aravin et al., 2008; Reuter et al., 2011).

piRNAs are produced in millions of copies during gametogenesis, compared to a few hundred copies of miRNAs, and they display a rather complex sequence composition. Mapping of piRNA libraries to the genome revealed that they are particularly enriched in discrete genomic loci: in mouse adult testes almost 8 million piRNA reads can be mapped to 214 genomic loci (Li et al., 2013). Interestingly in mammals, piRNAs do not show sequence conservation but their locations are syntenic (Aravin et al., 2006; Girard et al., 2006). These clusters can vary from a few kilobases (Kb) to over a 100 Kb in length and they are transcribed into long single-stranded RNA precursors from where silencing-competent piRNAs are produced. A strong genetic support of this idea is that the insertion of a P-element into the 5'-end of a fly somatic piRNA cluster called *flamenco* abolishes piRNA production of the whole 168kb cluster (Brennecke et al., 2007).

In flies, piRNA clusters are found predominantly in pericentromeric and telomeric heterochromatin. *Flamenco* is the dominant cluster in follicle cells and it is an example of a uni-strand cluster, where piRNAs are produced only from one strand of the genome (Figure

1-Introduction

1.4). This long transcript encodes piRNAs that are antisense to transposons and can therefore bind them through sequence complementarity. Other clusters are dual-stranded, producing piRNAs from both strands, like cluster *42AB* in *Drosophila* germ cells (Figure 1.4). Finally some clusters are bidirectional, having a divergent transcription from a central promoter, like a very prominent pachytene cluster in mouse chromosome 17 (Figure 1.5) (Aravin et al., 2006).

1.5.3 piRNA biogenesis

As mentioned above, unlike siRNA and miRNA, piRNA biogenesis does not require Drosha and Dicer (Houwing et al., 2007; Saito et al., 2010; Vagin et al., 2006) and despite the great efforts of the last 10 years, it still remains poorly understood in regard to certain aspects. For clarity it can be divided into primary and secondary biogenesis pathways. The processing of a piRNA precursor originating from a cluster transcript into an active piRISC complex is referred to as primary piRNA biogenesis. A loaded PIWI protein can then bind its target and slice it. This PIWI-mediated cleavage can trigger the generation of a secondary piRNA (Brennecke et al., 2007; Gunawardane et al., 2007), with a pathway referred as secondary piRNA biogenesis or ping-pong cycle amplification. These two pathways are strongly interconnected as it has recently been discovered that secondary piRNAs can trigger the generation of new waves of primary piRNAs (phased piRNAs) (Han et al., 2015; Mohn et al., 2015), within what it is also known as the inchworm pathway (Homolka et al., 2015).

Primary piRNA biogenesis

A long primary piRNA precursor that carries a 5'-cap and polyA tail (Li et al., 2013) is transcribed from a piRNA cluster by RNA PolII and exported to the cytoplasm, processed into intermediates and loaded into PIWIs. How the transcription machinery distinguishes piRNA clusters from other genes remains yet to be elucidated. A general model suggests that the cell recruits a transcription machinery to these genomic loci via specific chromatin marks and that transcripts are co-transcriptionally tagged by specific RNA-binding proteins which shuttle the precursors to the cytosol (Pillai and Chuma, 2012).

In *C. elegans*, piRNAs (also known as 21U RNAs) derive from individual gene-like loci with a conserved sequence motif immediately upstream, which may serve as a promoter

1-Introduction

to drive the expression of the locus. However, these motifs are not found in higher eukaryotes (Gu et al., 2012; Ruby et al., 2006).

In *Drosophila*, somatic follicle cells express only primary piRNAs, whereas ovarian germ cells employ both primary and secondary piRNAs (Malone et al., 2009). Recently, different transcription factors have been found to drive expression of specific clusters in both somatic and germ cells. In somatic cells, the *flamenco* cluster transcription requires the transcription factor cubitus interruptus (Goriaux et al., 2014). Recently, it has been observed that a small sequence of flamenco promoter (1-718nt of the promoter fused with exon1) is sufficient to induce piRNA production from a reporter gene in a *Drosophila* ovarian somatic cell line (OSC) (Homolka et al., 2015). This suggests that the primary transcript contains a piRNA trigger sequence (PTS) in the first exon, which is sufficient to recruit the primary processing machinery (Figure 1.4). In germ cells, the RDC protein complex (composed of Rhino, Deadlock and Cutoff) acts as the central licensing factor for dual-strand piRNA cluster transcription (Klattenhoff et al., 2009; Mohn et al., 2014). Rhino is a heterochromatin protein 1 (HP1) homologue that associates with methylated H3K9 found in heterochromatic dual strand clusters. Subsequent recruiting of Deadlock, Cutoff and the RNA helicase UAP56 block the splicing of the cluster transcript, promoting export of piRNA precursors to the cytoplasm where they are further processed (Zhang et al., 2012; Zhang et al., 2014b).

A common feature of primary piRNAs is a strong bias for a U at their 5' end (1U-bias) (Aravin et al., 2006; Brennecke et al., 2007; Girard et al., 2006). Structural studies and *in vitro* experiments suggest that the 1U-bias is an inherent property of PIWI proteins since certain PIWI proteins preferentially load piRNAs with a 1U-bias (Cora et al., 2014; Kawaoka et al., 2011). The U-bias could also be an output of the endonuclease which slices the long precursor. Zucchini (Zuc) is a putative nuclease protein, located on the surface of mitochondria and it is so far the best candidate to be the nuclease responsible for 5' end generation of piRNAs (Olivieri et al., 2010; Pane et al., 2007; Saito et al., 2010) since deletion of Zuc causes accumulation of unprocessed piRNA intermediates.

Once the 5' end of the piRNA is loaded into the PIWI protein, the 3' end of the intermediate is trimmed by an unknown Mg^{2+} -dependent 3' to 5' exonuclease, followed by 2'-O-methylation (Kawaoka et al., 2011). Tudor domain-containing (Tdrd) protein known as Tdrkh (or TDRD2) promotes the 3' trimming process (Honda et al., 2013; Saxe et al., 2013).

1-Introduction

Another important primary pathway component identified by genetic studies is Armi and its homolog in mouse, Mov10like-1 (Mov10L1) (Frost et al., 2010; Olivieri et al., 2010; Saito et al., 2010). piRNA biogenesis and transposon silencing are impaired in *armi/Mov10L1* mutants. MOV10L1 is an ATP-dependent RNA helicase that probably functions in funnel piRNA precursors into the endonuclease which catalyses the first cleavage step (Vourekas et al., 2015). In fly somatic follicle cells, Armi co-localises with Yb in perinuclear granules where primary piRNA loading takes place. These granules are known as Yb bodies. Yb is a tudor domain-containing protein and a putative RNA helicase, essential for primary piRNA biogenesis in somatic follicle cells. Its role in germ cells may be rendered redundant by the other two similar proteins, named sister of Yb (SoYb) and brother of Yb (BoYb) (Handler et al., 2011). Its homologue in mouse, TDRD12 is also essential for piRNA biogenesis (Pandey et al., 2013). Moreover, other components of Yb bodies including Vreteno (Vret) and Shutdown (Handler et al., 2011; Olivieri et al., 2010; Zamparini et al., 2011) and proteins associated with the mitochondria, like Minotaur and GasZ (Handler et al., 2013; Vagin et al., 2013), function in primary piRNA processing, but their specific contribution is unknown.

Once loaded with a primary piRNA, the piRISC follows different fates depending on the PIWI protein involved. In fly, Piwi and Aub are loaded with primary piRNAs. Piwi is imported into the nucleus to transcriptionally regulate target genes whereas Aub silences transposons post-transcriptionally, through direct cleavage of their mRNA target. In mice, primary loading generates MILI and MIWI piRISC complexes that can silence transposons post-transcriptionally in pachytene spermatocytes (Di Giacomo et al., 2013; Reuter et al., 2011).

Secondary piRNA biogenesis: ping-pong pathway

The first model of secondary piRNA processing was proposed in 2007, when two independent groups found three main differences between Ago3 and Aub-bound piRNAs in *Drosophila* germ cells. First Aub-bound piRNAs are typically antisense to TE while Ago3-bound piRNA are usually sense. Second, the first 10nt of Aub are often complementary to the first 10 nt of Ago3-bound piRNA (the ping-pong signature). Third, Aub-bound piRNAs have a 1U-bias whereas Ago3 piRNAs have no bias for the first nucleotide position but a strong bias for an adenosine at position 10 (A10-bias) (Brennecke et al., 2007; Gunawardane et al.,

1-Introduction

2007) (Figure 1.6). This led to the proposal of the ping-pong model where Aub binds its target through sequence complementarity and cleaves it between the 10th and 11th nucleotide of the guide piRNA. This generates two fragments, one containing a 3'OH and the other containing a 5'P. The 5'P is loaded into Ago3, forming a new piRNA complex, whereas the 3'OH-containing fragment is degraded (Figure 1.6). Once Ago3 is loaded with a new piRNA, it is able to cleave a new primary transcript and promote the formation of a new Aub-loaded piRNA. The new piRNA loaded into Aub has exactly the same sequence as the first one that initiated the cycle. The 5' ends of Aub- and Ago3-loaded piRNAs are generated by their reciprocal slicer activity. How the 3' end is processed however, remains so far unknown.

This feed-forward loop guarantees that every time Aub slices a target, a new piRNA with the same sequence is produced, preventing loss of information. Thinking of piRNA pathway as a cellular immune system against transposon, the ping-pong cycle can be referred to as an adaptive response against active elements (Siomi et al., 2011). This cell memory is also transmitted to the next generation, since piRNAs are maternally deposited in the embryo as Aub-piRISCs. This will boost piRNA production in the ovaries of the offspring by initiating the ping-pong cycle (Brennecke et al., 2008). Genetic screenings have implicated additional essential factors in the ping-pong cycle, including Spindle-E, Vasa and Krimper (Malone et al., 2009). In chapter 3, I will describe the role of the DEAD-box RNA helicase Vasa in this pathway, (Xiol et al., 2014) (Figure 1.6).

Secondary piRNA biogenesis in mouse

In mouse, piRNAs carrying a ping-pong signature can be found in all development stages, but in pachytene stage they represent a very minor proportion of the total pool (Berninger et al., 2011). Secondary piRNA production is more obvious in prenatal testis where MILI slicer products are loaded into MIWI2, generating a secondary piRNA complex (Aravin et al., 2008; De Fazio et al., 2011) Loaded MIWI2 is transported into the nucleus where it represses transposons by promoting DNA methylation and H3K9 methylation (Kuramochi-Miyagawa et al., 2008; Pezic et al., 2014). In the absence of MILI, MIWI2 fails to get loaded with piRNAs and fails to localise in the nucleus (Aravin et al., 2008). MILI and MIWI are active endonucleases and their slicer activity is necessary for their function, since mutant mice carrying a point mutation in the catalytic site (DDH to ADH) are infertile. On the contrary, *Miw2* slicer mutant is fertile, indicating that it does not need its slicer activity to

1-Introduction

exert its function (De Fazio et al., 2011; Reuter et al., 2011). This suggests that secondary biogenesis in mice is a linear process: MILI slicing generates MIWI2-secondary piRNAs that migrate into the nucleus. Nevertheless a ping-pong signature (10nt sequence overlap between the 5' end of two piRNAs) can also be observed within MILI piRNAs, and this is lost in the MILI slicer-mutant, suggesting that a homotypic ping-pong (MILI to MILI) may occur in mouse testes (De Fazio et al., 2011). Furthermore, it is possible that MIWI2 is also loaded with primary piRNAs, since in *Mili* catalytic mutant the piRNA population in MIWI2 is not completely lost, indicating that MIWI2 loading can be independent of MILI slicer activity (De Fazio et al., 2011)

A possible molecular link between MIWI2 and DNA methylation is not understood. DNA methylation is established and maintained during development by three distinctive DNA cytosine methyltransferases (DNMT1, DNMT3a and DNMT3b). Whereas DNMT1 has been involved in methylation maintenance (Li et al., 1992) and is required for stable inheritance of tissue-specific methylation patterns, DNMT3A and 3B are responsible for de-novo methylation (Okano et al., 1999). DNMT3L is a germ cell-specific, catalytically-inactive adaptor, which is also essential for *de novo* DNA methylation. Its deletion causes male-sterility, with a blockage of spermatogenesis due to a failure to establish methylation at retrotransposon elements (Bourc'his and Bestor, 2004) with a phenotype very similar to *Mili* and *Miwi2* mutants. *Miwi2* was shown to act upstream of *Dnmt3L*, since piRNAs are still produced in *Dnmt3L* mutant (Aravin et al., 2008; Kuramochi-Miyagawa et al., 2008).

As for *Drosophila*, additional factors implicated in secondary biogenesis have been characterized. These include Mouse Vasa Homologue (MVH), Hsp90 co-chaperone FKBP6, TDRD1, TDRD12 (Kuramochi-Miyagawa et al., 2010; Pandey et al., 2013; Reuter et al., 2009; Xiol et al., 2012) (Figure 1.4). Deletion of these proteins cause MIWI2 to be unloaded and mislocalised into the cytosol, suggesting that the role of these factors is upstream of MIWI2, probably facilitating ribonucleoprotein (RNP) remodelling, required for inter-Piwi exchange of piRNA intermediates. TDRD9 is a well conserved tudor domain containing protein, implicated in transposon silencing of mammals and insects. The *Tdrd9* mouse mutant is phenotypically similar to that of *Miwi2* (Carmell et al., 2007; Kuramochi-Miyagawa et al., 2008; Shoji et al., 2009) with regard to LINE-1 deregulation, cognate DNA demethylation and later spermatocyte defects. Investigation of the role of *Tdrd9* will be described in [chapter 4](#).

The Inchworm amplification loop (phased piRNAs)

An additional pathway that amplifies piRNA population has been recently described in flies and it involves both primary and secondary generated piRNAs (Han et al., 2015; Homolka et al., 2015; Mohn et al., 2015) (Figure 1.6B). As described before, in the feed-forward loop of the ping-pong cycle, Ago3 cleaves an RNA transcript and generates a new Aub-loaded piRNA (responder-piRNA). The sequence downstream of this piRNA is not degraded but it is used to produce new primary piRNAs covering all the sequence downstream (Mohn et al., 2015). These piRNAs have a strong 1U-bias and are loaded mainly into Piwi (85%) but also into Aub (15%) and are called phased piRNAs (Han et al., 2015). The 5' end of the first phased piRNA is immediately downstream of the responder-piRNA and the following ones are phased every 27nt. Mapping single phased piRNAs on the genome revealed that they have a +1U bias, meaning that the following nucleotide after the 3' end is often a U. This suggests that a single endonucleolytic event (most probably performed by Zuc) is sufficient to produce the 3' end of a phased piRNA and the 5' end of the following one (Figure 1.6B) (Han et al., 2015; Mohn et al., 2015).

These findings lead to a model where the primary biogenesis machinery moves along the transposon transcript, starting from the responder-piRNA binding site, stopping every 27nt to cut and produce the 5' end of a new piRNA that is immediately loaded into the MID domain of a PIWI. This 'crawling' of the primary biogenesis machinery along the transcript in a 5' to 3' direction in regular 27nt steps is reminiscent of an inchworm movement (inchworm model). This finding finally explains some observations made many years ago. For example, it was unclear why a considerable percentage of primary Piwi-loaded antisense piRNAs are lost in *Ago3* fly mutant (a secondary biogenesis factor): the generation of a large percentage of Piwi-loaded piRNAs is dependent Ago3 slicer activity.

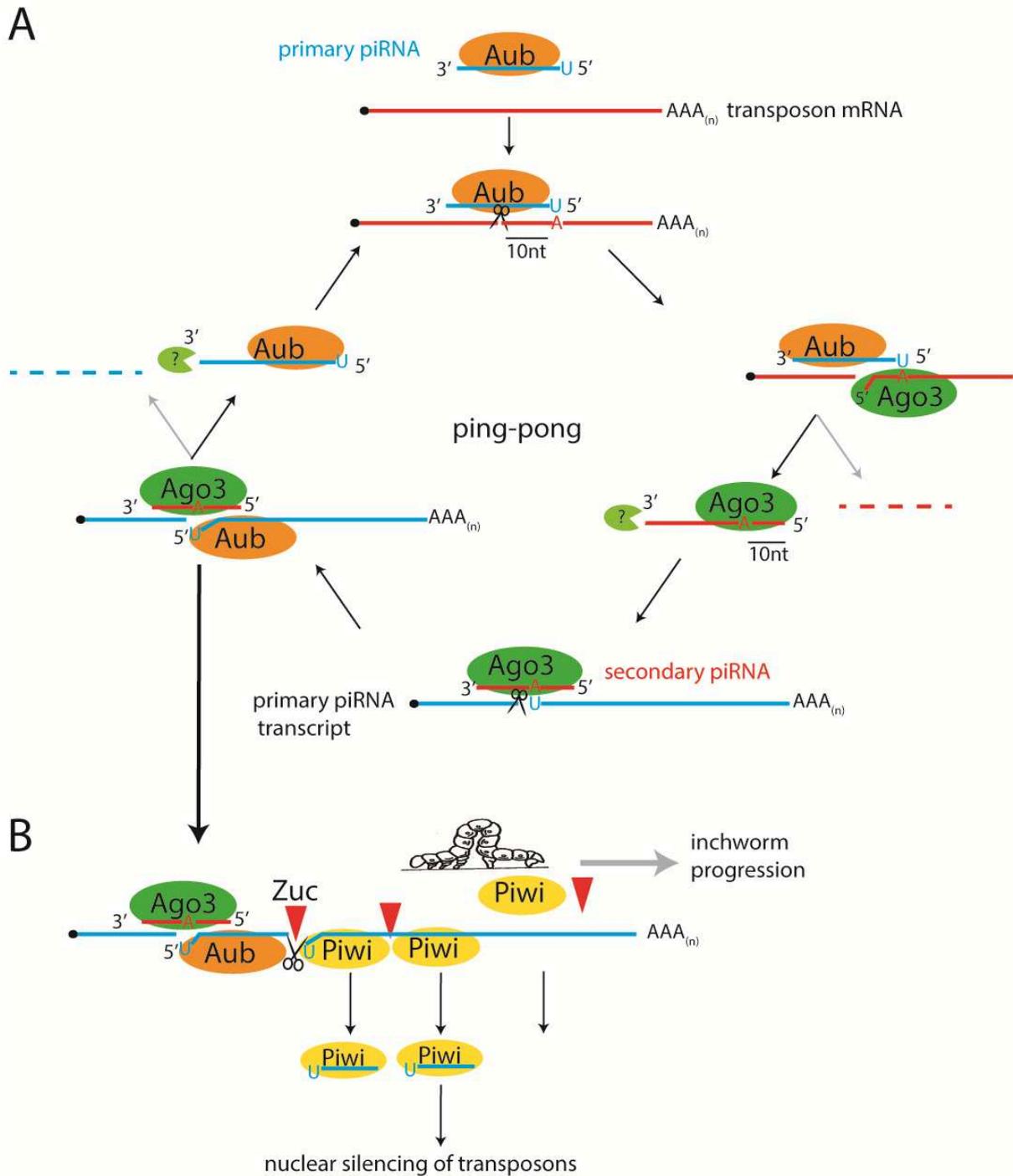


Figure 1.6 Ping-pong and Inchworm amplification loops in *Drosophila*. (A) 1U-bias primary antisense (blue) piRNAs loaded into Aub initiate the ping-pong cycle. The cleavage product with a free 5'P end is loaded into Ago3 generating a secondary (red) piRNA that will have a 10nt overlap with the primary piRNA. Loaded Ago3 initiates a feed-forward loop where it cuts a primary piRNA transcript and produces a new 5' end for Aub. (B) Ago3 slicer can also trigger the inchworm amplification: a secondary piRNA is loaded into Aub (responder-piRNA), its 3' end is generated by an endonuclease (most probably Zucchini). The 5' end resulting from this cleavage is loaded into Piwi, producing phase-piRNAs along the entire transcript in a 5' to 3' direction. Loaded Piwi migrates into the nucleus and silences transposon genomic loci.

1.5.4 Subcellular localisation of piRNA pathway and additional factors required

Most of the piRNA biogenesis in insect germ cells takes place in perinuclear granules that are referred to as *nuage* (“cloud” in French). This is an electron-dense particle that is surrounded by mitochondria and it is present in 50-100 copies per cell. The *nuage* contains cytoplasmic PIWI proteins and piRNA biogenesis factors (Brennecke et al., 2007; Houwing et al., 2007; Lim and Kai, 2007) and it also contains components of the miRNA and mRNA degradation pathways, similar to P-bodies in somatic cells.

Mutation in PIWI proteins and other piRNA pathway components abolishes the formation of such granules, indicating that the existence of these granules is under the control of the PIWI pathway (Pillai and Chuma, 2012). A particularly important role in assembling the *nuage* is played by Tudor domain-containing proteins. Tudor domains contain a characteristic aromatic cage that recognizes symmetrically dimethylated arginines (sDMAs) RG/GR repeats in the N-terminus of PIWI proteins. This interaction is believed to act as a binding platform for assembling different RNP complexes during piRNA biogenesis (Kirino et al., 2009; Kirino et al., 2010; Nishida et al., 2009; Reuter et al., 2009; Vagin et al., 2009). Mutation of Tdrd proteins such as Tudor, Tejas, Spindle-E, Krimper and Qin/Qumo, lead to general defects in PIWI protein localisation, with Aub spreading in the cytoplasm instead of concentrating in the *nuage* (Anand and Kai, 2012; Lim and Kai, 2007; Malone et al., 2009; Nishida et al., 2009; Patil and Kai, 2010). Some of these proteins act in primary biogenesis whereas others are essential to maintain the ping-pong pathway (Table 1).

In mouse testes, two main distinct granules are detectable during spermatogenesis: the inter-mitochondrial cement (also called *nuage* or pi-body) and the chromatoid body. The first one is assembled between mitochondria in prenatal prospermatogonia, postnatal spermatogonia and spermatocytes and it contains most of the piRNA biogenesis factors like MILI, TDRD1, MVH and GASZ. The chromatoid body instead, is a single granule that is visible in late meiotic spermatocytes and haploid spermatids. It also contains piRNA pathway components like MILI, MIWI, TDRD1, TDRD5, TDRD9, MVH and MAEL, amongst others. Furthermore, in prenatal testes, secondary biogenesis factors like MIWI2, TDRD9 and MAEL were shown to localise in different granules than *nuage*, called piP-bodies. This name derive from the fact that these granules contain piRNA biogenesis factors and proteins involved in mRNA decay and miRNA silencing typically found in somatic processing bodies

1-Introduction

(P-bodies) (Shoji et al., 2009; Soper et al., 2008). Although piP-bodies often overlap and merge with pi-bodies, they are distinct granules. The exact role of these different granules during piRNA biogenesis remains yet to be elucidated.

Table 1 Factors involved in piRNA pathways in mouse and *Drosophila* with the relative protein domains. Different colours indicate each protein's involvement in **primary biogenesis (green)**, **secondary biogenesis (orange)** or employment of an **effector function (purple)**. Proteins containing a **helicase domain** are highlighted in blue.

<i>Drosophila</i>	Protein domains	Mouse
Piwi	PAZ, MID, PIWI	MIWI
Aub	PAZ, MID, PIWI	MILI
Armi	P-loop NTPase, Uvr_D_C_2	MOV10L
Zucchini	PLD-like	MitoPLD
PAPI	Tudor, KH	TDRD2/TDRKH
Yb	Tudor, DExH, Hel-C	-
BoYb	Tudor, DExH, Hel-C	-
SoYb	Tudor, DExH, Hel-C	TDRD12
Ago3	Paz, Mid, Piwi	MIWI2
Vret	Tudor, RRM, Mynd	TDRD1
Vasa	DEAD box, Hel-C	MVH
Shutdown	ppLASE, tpr	FKBP6
Krimper	Tudor	-
Eggless	Tudor, MBD, SET	SETDB1/SETDB2
Uap56	DEAD box	DDX39
Spn-E	Tudor, DExH, Hel-C, HA2	TDRD9
Qin/Qumo	Tudor, RING, B-box	TDRD4/RNF17
Tejas	Tudor, Lotus	TDRD5
Tud	Tudor,	TDRD6
Rhino	Chromo	-
Cut-off	RAI1	-
Del	NA	-
Squash	Rnase HIII-like	-
Minotaur	GPAT	GPAT1,GPAT2
GasZ	Ankyrin-rpt,SAM	GASZ
Mael	Mael	Mael
Asterix	UPF0224_CHHC-Znf	GTSF1

1.6 RNA helicases

A relevant number of factors involved in piRNA biogenesis contain RNA helicase domains (table 1). RNA helicases participate in various physiological aspects of the RNA

1_Introduction

metabolism, e.g. transcription, translation and mRNA decay. Mechanistically, helicases are proteins that can translocate along a dsRNA molecule, thus separating one strand from its complementary. Alternatively, they can bind a dsRNA and induce local strand separation of the two RNA strands (Linder and Jankowsky, 2011; Ozgur et al., 2015). Beside the different mechanisms of action, these proteins all share the ability to hydrolyse ATP in the presence of RNA, therefore they are often referred more generally as RNA-dependent ATPases. All eukaryotic RNA helicases can be divided into superfamily 1 and 2 (SF1 and SF2) based on sequence, structural, and mechanistic features, with the SF2 containing the majority of the members. The hallmark of both families is the helicase core, which is formed by two RecA domains (RecA1 and RecA2), named for their structural similarity to the bacterial recombinase A (RecA), a protein involved in DNA homologous recombination. The RecA is a globular fold, consisting of a parallel β -sheet surrounded by α -helices. The most conserved motifs in both families is where ATP is bound and hydrolysed, and it is located in the intersection between the two RecA domains. Particularly conserved are motif 1 and 2, also known as WalkerA and WalkerB (Walker et al., 1982) (Figure 1.7). The WalkerA contains the Lysine (GKT motif) that coordinates the α and β phosphates of the ATP, whereas the WalkerB motif contains the universally conserved Asp-Glu residues that catalyse ATP hydrolysis: the aspartic acid coordinates a magnesium ion and the glutamic acid is believed to act as the catalytic base that performs the nucleophilic attack on the γ -phosphate of the ATP (Figure 1.7B), as has been shown for similar ATPase enzymes (Orelle et al., 2003). The proton required to perform the reaction is provided by a molecule of water (catalytic water or lytic water) that is coordinated by the Glu and by other residues forming the binding pocket (Figure 1.7B) (Sengoku et al., 2006). Another important motif is the Arg finger domain and it is positioned on the RecA2. This motif contains two arginines that coordinate the two phosphate molecules of the ATP.

The SF2 family can be further divided in different subfamilies by different criteria. One of them is to consider the amino acids present next to the catalytic Asp-Glu residues, like in the case of Asp-Glu-Ala-Asp (DEAD)-box helicases.

1.6.1 DEAD-box RNA helicases

DEAD-box RNA helicases constitute the larger family of the SF2 and they are mainly composed by the two RecA domains flanked by short N-terminal and C-terminal regions.

1_Introduction

These regions usually contain short regulatory sequences used for interaction with other proteins or to recognize specific RNA sequences. The ATP binding pocket of RecA1 can bind ATP independently but the contribution of RecA2 is essential to form the active site. Binding of ATP and RNA-strand separation induce different conformational rearrangements of the two RecA domains that can be summarized as the on-state and the off-state (Figure 1.7). In the on-state the two RecA domains are closed and bury the ATP in the cleft of the two domains. The RNA molecule is bound perpendicularly to the two RecA domains by its sugar-phosphate backbone. Once the ATP is hydrolysed the helicase expels the resulting hydrolysis products (ADP and P_i) and releases the RNA, switching into the off-state. In the off-state the two domains are far from each other and neither RNA nor ATP are bound. Unwinding of the RNA duplex is promoted by ATP binding and not by its hydrolysis (Hilbert et al., 2011).

DEAD-box proteins can unwind only short RNA duplex by forcing local strand separation upon binding of ATP. Beside their helicase activity, these proteins can often function as RNA clamp, promoting the formation of an RNP complex on a specific RNA sequence. One famous example is the core helicase component of the exon junction complex (EJC) eIF4AIII (Bono et al., 2006). eIF4AIII binds mRNA between the junction of two exons that have been spliced and promotes the assembly of the EJC. Its grip on the RNA is maintained by structural constraints of the other components of the EJC (Mago, Y14 and Berentsz). eIF4AIII binds RNA in the on-state and hydrolyze the ATP but the hydrolysis products cannot be released as Mago and Y14 stall the helicase in the closed conformation. This stalled RNP functions as a platform for recruiting other proteins when the mRNA is exported into the cytosol.

1_Introduction

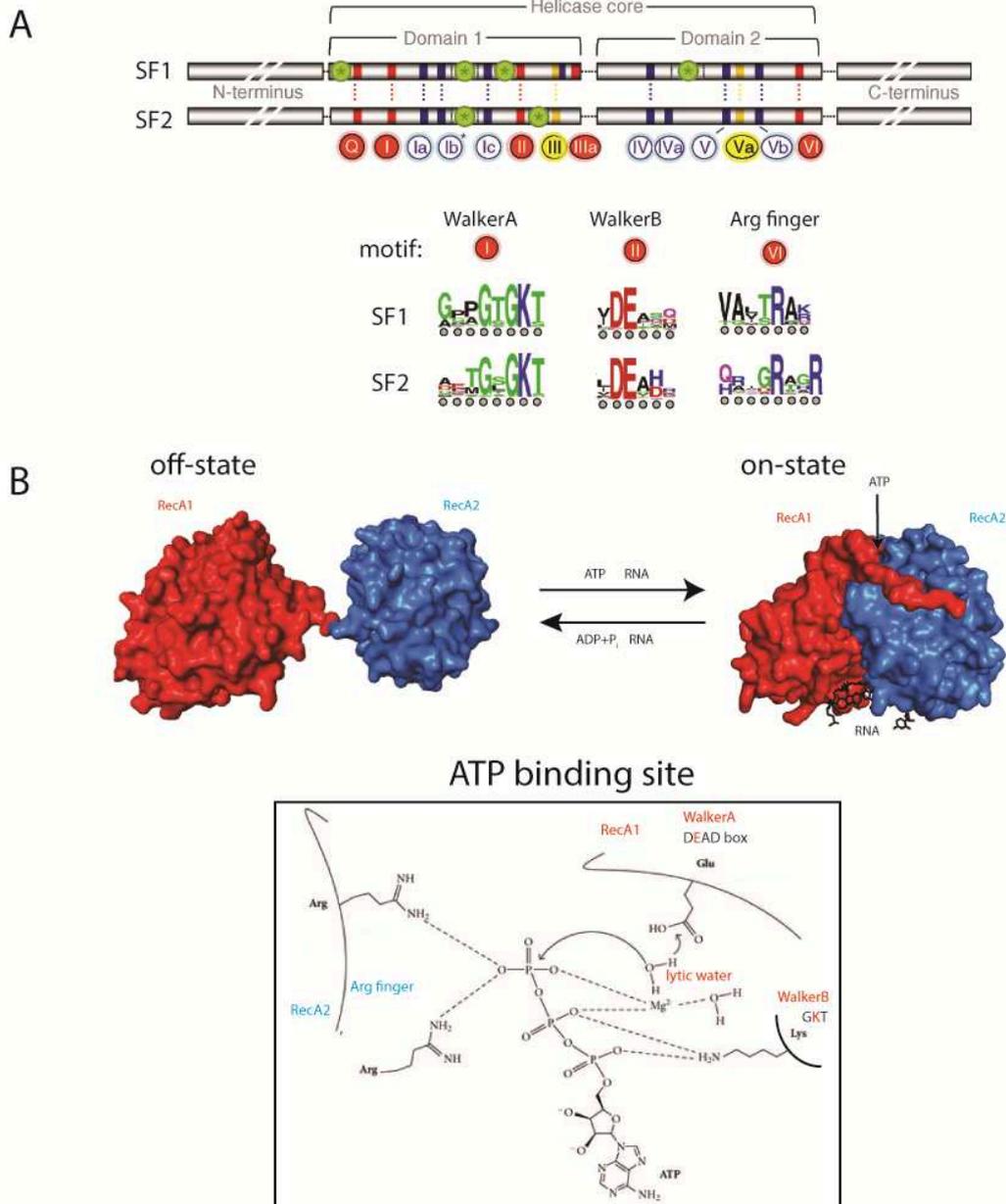


Figure 1.7 RNA helicase families and DEAD-box proteins. (A) Domain architecture of RNA helicases, main motif discussed in this thesis are indicated (adapted from (Fairman-Williams et al., 2010)). (B) Structure of Vasa in the closed conformation (PDB: 2DB3) and in the open conformation (off-state, remodelled on EIF4A1 structure, PDB: 2G9N). Details of the ATP hydrolysis mechanism: a lytic water performs the nucleophilic attack to the gamma phosphate. Residues from RecA1 and RecA2 are involved in the catalysis (adapted from (Irene Briguglio, 2011))

1.6.2 DExH box RNA helicases and other families

DEAD-box proteins have been studied extensively in the late 1990s, mainly for two reasons: they are involved in fundamental cellular processes like translation and splicing and

1_Introduction

they are small and globular. The latter propriety made them suitable substrates for structural and biochemical studies. Much less is known about DExH-box proteins which are multi-domain proteins that usually unwind RNA in a 3'-5' direction and include the DEAH-box proteins (like the spliceosomal protein Prp34) and the Ski2-like proteins (like the DNA helicase Hel308) (Figure 1.8). Structurally they are composed of the two canonical RecA domains plus additional domains that sometimes are referred to a 90aa-long helicase associated domain 2 (HA2) in on-line protein databases. Recent structural studies revealed that this additional domain is longer and comprises a winged-helix (WH) and a helical bundle domain (Ozgur et al., 2015). It retains the two recA domains in a closed conformation that is similar to the on-state of the DEAD-box proteins (Figure 1.8). ATP hydrolysis does not induce major conformational changes in these proteins and the ATP and RNA binding activities are usually independent from each other (Ozgur et al., 2015; Weir et al., 2010). The structure of Hel308 in complex with a partial dsDNA molecule revealed how these proteins may work in unfolding DNA/RNA duplexes (Buttner et al., 2007) (Figure 1.8). The RNA chain is embedded in a channel that is formed in between the RecA and the helical modules and the unwinding element is a β -hairpin present at the top of RecA2. As the single strand 3' end proceeds into the tunnel, the β -hairpin melts the base pairs at the entrance. ATP hydrolysis provides the energy to drive the unwinding forward. Most of the regulation of the DExH-box helicases occurs at the level of the auxiliary domains that flank the two RecA domains. Additional C-terminal domains differ from one protein to another.

1_Introduction

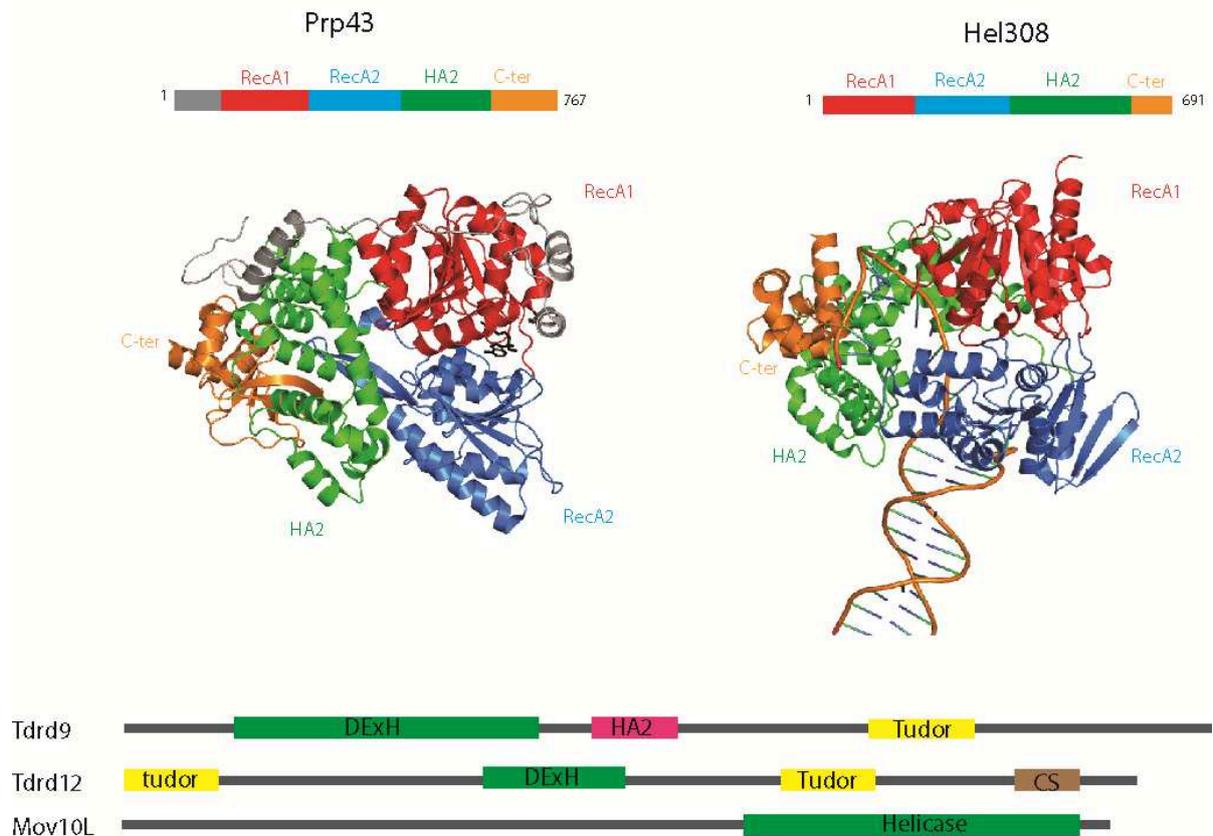


Figure 1.8 Structure and domain architecture of two RNA and DNA DExH-box helicases: Prp43 (PDB 3KX2) and Hel308 (PDB 2P6R). The helicase associated domain 2 (HA2) is closely associated with the two RecA domains and keep them in a closed conformation. At the bottom: domain architecture of three putative RNA helicases involved in the piRNA pathway.

Two DExH-box helicases have been identified in the piRNA pathway by genetic studies, but the contribution of their helicase activity to their function is unknown. These RNA helicases include the mouse protein TDRD9 and TDRD12 (Figure 1.8) (Pandey et al., 2013; Shoji et al., 2009). In addition to the helicase, these proteins contain one or two Tudor domains that mediate the interaction with the N-terminal region of PIWI proteins. The helicase domains of TDRD12 and TDRD9 are similar to the yeast splicing factor Prp43 (figure1.8); TDRD12 is implicated in secondary piRNA biogenesis in mice (Pandey et al., 2013), whereas the role of TDRD9 is still unclear, as it will be discussed in [chapter 4](#). MOV10L1 (the orthologue of *Drosophila* Armi) is another putative RNA helicase involved in primary piRNA biogenesis. It contains the canonical GKT motif for ATP binding and a DEAGQ/H motif that hydrolyse ATP and it is assigned to the SF1 family (like UPF1) (Figure 1.7). A recent study showed that a single point mutation in the ATP binding motif (GKT to

GAT) is sufficient to recapitulate the phenotype of *Mov10l1*-null mouse. This suggests that the helicase activity of Mov10L1 is essential for its function in the mouse piRNA pathway (Frost et al., 2010; Vourekas et al., 2015).

1.7 Aim of the thesis

Beside its fundamental role in development and transmission of genetic information, many aspects of the piRNA pathway remain yet to be elucidated. In the Pillai lab, we aim at understanding the biogenesis and function of piRNAs. In particular, we are trying to dissect the function of the numerous protein factors involved. During my PhD research, I investigated the role of some RNA helicases that are implicated in the pathway. Due to their dynamic nature, I generated a single point mutation in the helicase domain (DEAD box) of the RNA helicases in order to lock them in different functional states.

This proved to be an interesting tool, since mutation of the DEAD box of Vasa (E to Q) *in vivo* locked the protein in the state of a transient RNP biogenesis complex. In the first part of the thesis, I will describe how I contributed to the understanding of the function of Vasa in the piRNA pathway. My first objective was to understand how this mutation alters the biochemical activity of Vasa concerning RNA binding and ATP hydrolysis by using biochemical and structural approaches. Secondly, I aimed at understanding the nature of the RNP complex that Vasa forms in insect cell lines by employing pull-down techniques and recombinant protein complex expression. Finally, I will discuss the effect that is generated by the same mutation in mouse.

The second part of the thesis will focus on the function of Spindle-E/TDRD9, an helicase implicated in the pathway. First, I characterized its ATPase activity *in vitro* and attempted to study its protein partners during spermatogenesis. A point mutation was introduced into the helicase core of Tdrd9 and a knock-in mutant mouse was generated, that expresses a catalytic-dead enzyme. I studied and characterized the phenotype of this mutant mouse to examine, whether its helicase activity was required for efficient transposon silencing in the germ line. Furthermore, I attempted to understand at which step of the piRNA biogenesis it is involved, by analysing piRNA profiles and subcellular localization of different components.

2 - MATERIALS AND METHODS

2.1 Clones and plasmids

pETM-11 (6xHis-fusion; EMBL Protein Expression and Purification Core Facility) has been used for recombinant protein expression in *E. coli*. pACEBac2 (EMBL) has been used for expression in insect cells, which was further modified by inserting an N-terminal Sumo-6xHis-StrepII tag upstream of the protein encoding genes. This plasmid was used independently or fused to pIDC and pIDK for expression of multiple proteins using the multiBac system (Bieniossek et al., 2012). For expression in mammalian cells pCiNeo (Promega) has been used, containing an NHA tag. The plasmid pCAG containing the cDNA of Tdrd9 was a kind gift of Dr. Shinichiro Chuma (Shoji et al., 2009). Plasmids for expression in the BmN4 cell line were pBEMBL-NHA (EMBL) (Xiol et al., 2012) or pizV5-2xFlag-2xStrept (from Invitrogen) while pCR2.1 was used for cloning blunt PCR products obtained from genomic PCR.

Most of the complementary DNAs (cDNAs) used for this study were sub-cloned from plasmids available in the lab. Spindle-E cDNA was obtained by RT-PCR using Superscript one-step RT-PCR with Platinum Taq Kit (Invitrogen) but its 3'end was obtained synthetically and cloned together with the 5'end. A list of all the plasmids used in this study is shown in Table 2.1. The cDNA for mouse Tdrd9 (1383aa; Genbank Accession no. AB362563) was a kind gift of Dr. Shinichiro Chuma (Shoji et al., 2009). cDNAs were amplified by PCR with DNA primers containing flanking restriction sites and cloned into the target vector by standard methods. Pwo polymerase (Pqlab) was used for amplification and T4 DNA ligase (Fermentas) for ligation of the fragments. Plasmids were transformed and amplified in chemocompetent TOP10 cells (Invitrogen). Cloning strategies were designed using VectorNTI software (Invitrogen) and positive clones were sequenced by Macrogen (Korea). For gene mutation, site-directed mutagenesis was performed using the overlapping extension strategy (Ho *et al.* 1989).

For expression studies in *Bombyx mori* BmN4 ovarian cell cultures (Kawaoka et al., 2009), the coding sequence for the full length *Bombyx mori* Vasa (BmVasa; 1-601aa; Genbank Accession Number:NM_001043882.1) or its deletion versions were cloned into the pBEMBL-NHA (Xiol et al., 2012). Point mutations in the BmVasa helicase domain were introduced by an overlap PCR strategy: ATP-binding mutant (K230N; GKT→GNT) or ATP hydrolysis product-trap mutant (E339Q; DEAD→DQAD). To express a recombinant form of the *Bombyx mori* Vasa

2-Materials and methods

helicase core domain (135-564 aa) or its mutated versions in *E.coli*, the required coding sequences were inserted into the pETM-11 vector. To raise antibodies against mouse TDRD9, different protein fragments were cloned into the pETM-11 vector. For the production of full-length proteins we used eukaryotic expression systems based on insect ovary-derived cells: *Spodoptera frugiperda* 21 (Sf21) or the *Trichoplusia ni* HiFive cells. Full-length mouse TDRD9 (1-1383 aa) and the TDRD9^{DQ} ATPase mutant (DEVH→DQVH) were cloned into the NheI and KpnI restriction sites of the vector pACEBac2 to express recombinant proteins as 6xHis-Sumo-StrepII fusions. To express the influenza hemagglutinin epitope (HA)-tagged TDRD9 versions in mammalian cells, the full length coding sequence for mouse Tdrd9 was inserted into the XbaI and NotI sites of the vector pCIneo-NHA.

2-Materials and methods

Table 2.1: List of plasmids used in this study

Organism	Protein	Expression host	Plasmid	Tag	Obtained/subcloned
<i>Bombyx mori</i>	Spindle-E (1-747)	E.Coli	PETM-11	6xHis	RT-PCR+ syntetic cDNA
<i>Bombyx mori</i>	Spindle-E (1-847)	E.Coli	PETM-11	6xHis	RT-PCR+ syntetic cDNA
<i>Bombyx mori</i>	Spindle-E (1-677)	Sf21	pACEBac2	Sumo-6xHis-Strept	RT-PCR+ syntetic cDNA
<i>Bombyx mori</i>	Spindle-E (1-746)	Sf21	pACEBac2	Sumo-6xHis-Strept	RT-PCR+ syntetic cDNA
<i>Mus musculus</i>	TDRD9 (1-723)	Sf21	pACEBac2	Sumo-6xHis-Strept	Shoji et al, 2009
<i>Mus musculus</i>	TDRD9 (1-725)	Sf21	pACEBac2	Sumo-6xHis-Strept	Shoji et al, 2009
<i>Mus musculus</i>	TDRD9 FL	Hi-5	pACEBac2	Sumo-6xHis-Strept	Shoji et al, 2009
<i>Mus musculus</i>	TDRD9 (90-1383)	Hi-5	pACEBac2	Sumo-6xHis-Strept	Shoji et al, 2009
<i>Bombyx mori</i>	Vasa (134-563) WT	E.Coli	PETM-11	6xHis	Xiol et al, 2014
<i>Bombyx mori</i>	Vasa (134-563) GNT	E.Coli	PETM-11	6xHis	Xiol et al, 2014
<i>Bombyx mori</i>	Vasa (134-563) DQAD	E.Coli	PETM-11	6xHis	Xiol et al, 2014
<i>Mus musculus</i>	MVH FL	E.Coli	PETM-11	6xHis	RT-PCR
<i>Bombyx mori</i>	Vasa FL DQAD	BmN4	PiZV5	2xStrept-2xFLAG	Xiol et al, 2014
<i>Bombyx mori</i>	Vasa FL WT	BmN4	PbEMBL	NHA	Xiol et al, 2014
<i>Bombyx mori</i>	Vasa FL DQAD	BmN4	PbEMBL	NHA	Xiol et al, 2014
<i>Bombyx mori</i>	Siwi FL	Sf21 or High Five	pACEBac2	Sumo-6xHis-Strept	Xiol et al, 2014
<i>Bombyx mori</i>	Siwi FL	-	pIDC	-	Xiol et al, 2014
<i>Bombyx mori</i>	Vasa(DQ) FL	Sf21 or High Five	pACEBac2	Sumo-6xHis-Strept	Xiol et al, 2014
<i>Bombyx mori</i>	Vasa(WT) FL	Sf21 or High Five	pACEBac3	Sumo-6xHis-Strept	Xiol et al, 2015
<i>Bombyx mori</i>	Ago3 FL	-	pIDK	-	Xiol et al, 2014
<i>Bombyx mori</i>	Vasa(DQ)-Siwi-Ago3	Sf21 or High Five	pACEBac2	Sumo-6xHis-Strept	Xiol et al, 2014
<i>Mus musculus</i>	Tdrd9 (1-153)	E.Coli	PETM-11	6xHis	Shoji et al, 2009
<i>Mus musculus</i>	Tdrd9 (1063-1383)	E.Coli	PETM-11	6xHis	Shoji et al, 2009
<i>Mus musculus</i>	Tdrd9 (1121-1383)	E.Coli	PETM-11	6xHis	Shoji et al, 2009
<i>Mus musculus</i>	Tdrd9 (1158-1383)	E.Coli	PETM-11	6xHis	Shoji et al, 2009
<i>Mus musculus</i>	TDRD9 FL	HEK cells	pCAG	6xHis	Shoji et al, 2009
<i>Mus musculus</i>	TDRD9 FL	Hela cells	pCIneo-NHA	NHA	Shoji et al, 2009
<i>Mus musculus</i>	Tdrd9 exon5 fragments	E.Coli	PCR2.1	-	genomic PCR

2.2 Cell lines

Human embryonic kidney cell line 293 (HEK293) transformed with the SV40 large T antigen (HEK293T) and HeLa cells were grown in Dulbecco's modified Eagle Medium (DMEM; Invitrogen, cat. No. 21969) supplemented with 10% fetal calf serum (Invitrogen, cat. No. 16000044) and 1% Penicillin-Streptomycin (Life technologies, cat. No. 15140122), and maintained in an environment with 5% CO₂ at 37°C. For transfection, cells growing in a 75 cm² flask were washed with warm (37°C) 1X PBS and incubated with 1 ml of Trypsin (Life technology, cat. No. 25300-054) for one minute to promote removal of cells from the growth

2-Materials and methods

surface. Trypsin was aspirated and 10 mL warm DMEM media was added to harvest the cells. Approximately 500 μ L cell solution were seeded into a 6 cm dish and grown overnight till 50-60% confluent. Plasmid was transfected with Lipofectamin (Invitrogen, cat. No. 18324012) and PlusTM reagent (Invitrogen, cat. No. 11514015). About 2 μ g plasmids was mixed with 30 μ L serum-free media and 1.5 μ L Plus reagent, and incubated at room temperature for 15 min. 1.5 μ L Lipofectamin was mixed with 30 μ L serum-free media. The two mixture systems were combined together, and incubated at room temperature for another 15 min. Media in the petri dish was replaced with pre-warmed serum-free media, and the mixture was added to the petri dish. The serum-free media was replaced with warm media supplemented with serum after 5h. Cells in each 6 cm dish were harvested 48 hours post-transfection.

For immunofluorescence analysis, glass coverslips were placed in 6-well or 12-well plates before seeding cells. Cells were grown till 50% confluence and transfected with 400-600ng of plasmid.

The *Bombyx mori* (Silkworm) ovarian cell line BmN4 (Kawaoka et al., 2009) was maintained in IPL-41 media (Invitrogen catalogue no. 11405057) supplemented with 10% fetal bovine serum (Invitrogen; Catalogue no. 16000044) and 0.5% Pencillin-Streptomycin (Invitrogen). The adherent cells were cultured in a temperature-controlled incubator at 27°C. Cells grow as a monolayer with well-defined, elongated morphology that attach to the plate surface. After reaching confluency, a second layer of rounded cells grow on top of the monolayer. Upon reaching the second layer, cells were passaged and monolayer cells at ~90% confluency were used for transfections. For immunoprecipitations, cells were grown in a 6-cm petri dish and transfected (Fugene; Roche) with 2 μ g of expression plasmids.

2.3 In vivo studies

2.3.1 Mouse strains and genotyping

Generation of *Tdrd9*^{E257Q} knock-in and *Tdrd9* knock out mutant mice

The *Tdrd9* gene locus is located on chromosome 12 of the mouse genome (chr12:111971559-112068854) and is composed of 36 exons, with the translated sequence being

2-Materials and methods

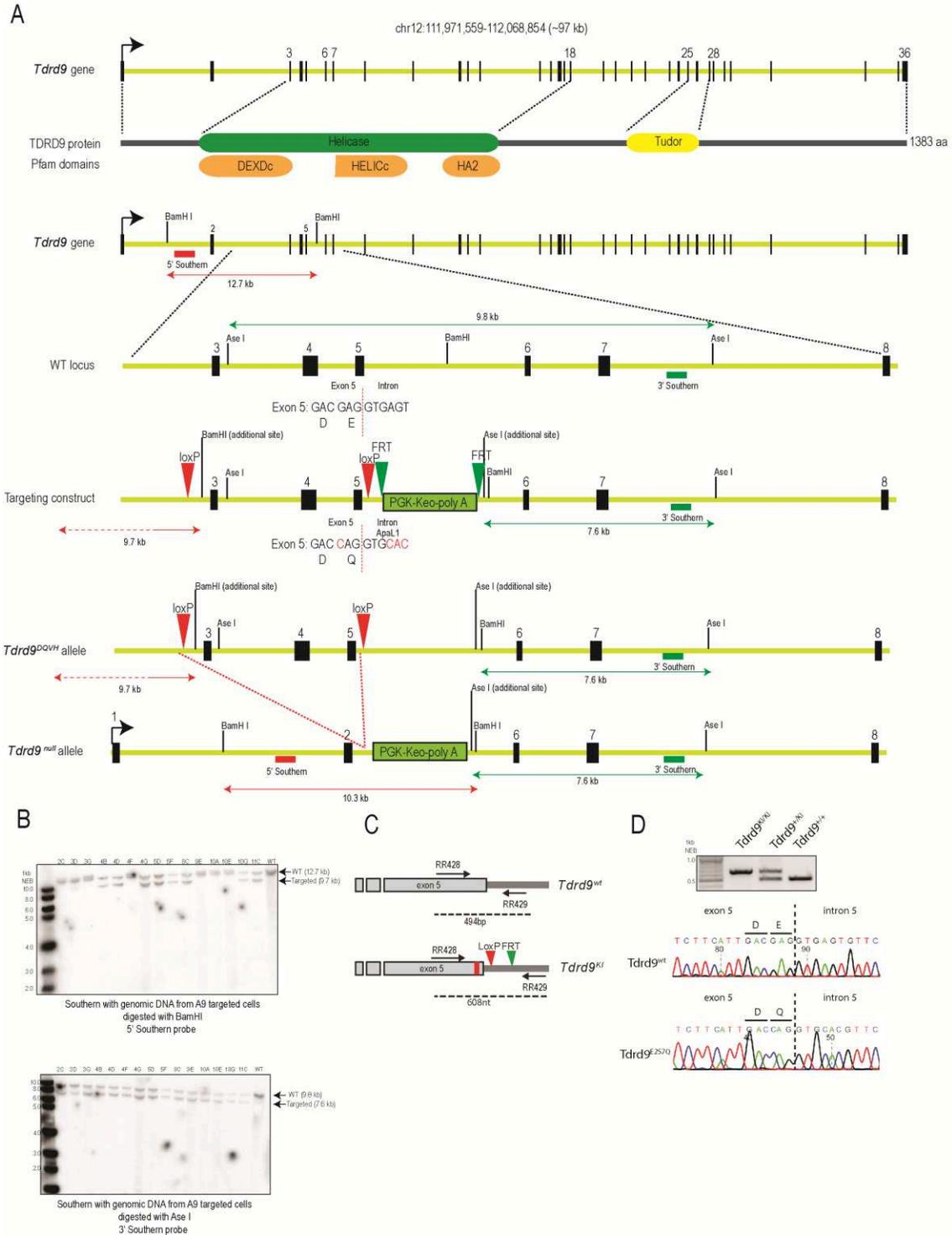
provided by exons 1-36 (Figure 2.1A). The N-terminal RNA helicase domain is constituted by exons 3 to 18, the C-terminal tudor domain by exons 25 to 28. The *Tdrd9* locus in the hybrid C57BL/6 × 129^{SV} embryonic stem (ES) cell line A9 was targeted using a genomic fragment that replaced exons 3 to 8. This fragment carries a single point mutation in the exon 5 (E257Q) within the ATPase motif (DEVH→DQVH), thus generating a putative catalytic-dead RNA helicase. In addition, it contains *loxP* sites flanking the exons 3-5, allowing deletion of the exons to produce a knock-out mutant (Figure 2.1A). The mice were created at the EMBL Transgenic mouse facility, EMBL Monterotondo, Italy. Mice used for experiments were housed at the animal facility of the CEA, Grenoble, France and subject to all national French and European regulations concerning animal research. To enable selection of the targeted ES cells, the targeting construct harbors a PGK-Keo (Kanamycin-neomycin)-polyA cassette in the intron directly downstream of exon 5. The selection cassette is flanked by *frrt* sites to allow its eventual excision from the genome.

Electroporated A9 ES cells were selected (neomycin; G418) and 300 clones screened by Southern blotting with probes recognizing the 5' (BamHI-digested DNA) and 3' (AseI-digested DNA) regions flanking the targeted site. Only 8 ES cell clones that were detected by both probes were retained (Figure 2.1B). These ES cell clones were karyotyped for genome integrity. One clone was used for injection into C57BL/6N host embryos (8-cell stage) for mouse generation. One founder animal was obtained and identified by (brown) coat color and genotyped by PCR of tail genomic DNA for presence of the targeted allele. The founder was crossed with Del-FLPeR female mice to remove the PGK-Keo cassette to generate the *Tdrd9*^{E257Q} allele (hereafter referred to as *Tdrd9*^{KI}). Backcrosses with C57BL/6j Rj (Janvier labs) wild-type females were performed to transfer the mutation to the C57BL/6 background to obtain heterozygous mice carrying the targeted allele. Heterozygous animals for *Tdrd9*^{KI} were intercrossed in order to produce the homozygous *Tdrd9*^{KI/KI} mice. Heterozygous *Tdrd9* null allele (*Tdrd9*^{+/- (Neo)}) was generated by crossing the founder with Del-Cre female mice. This lead to deletion of exons 3 to 5, leading to a frame shift and a stop codon after 2 amino acids in exon 6. *Tdrd9*^{(Neo)/-(Neo)} mice, hereafter referred as *Tdrd9*^{-/-}. Note that the *Tdrd9*-null allele still contains the PGK-Keo cassette which may alter the expression of the locus.

2-Materials and methods

Genotyping was performed with genomic PCR, using specific primers that were positioned around the site of mutagenesis, as illustrated schematically in figure 2.1C. PCR fragments were cloned in PCR2.1 vector using the TA cloning kit (Thermo Fisher) and sequenced (Figure 2.1D).

2-Materials and methods



2-Materials and methods

Southern blot probe sequences

5' probe (603 bp)

CTGGCCTTTGAGCAGACTTCAGTGGCCTTAGAGGACTTCTTGAGCCTTCTCTTGGTTCCTCGGCTTATTG
GACATCTGTCCAGCTGCTACAATTGCTGCTTGTGGCTTTAGTTGCCACCTCAGTCCACACGAGGACTGG
CCTTTGAACTCACTGCACACACATGTTATCTAGTGCTTGGTGTGTGCTACTTACATTACATTTTCATTTTT
CTCCTTTCTCATGAAATTGTTAAGATTGAAGTAGAGACCAAATACCATTTTCTTTAACTATCTCTCCTAC
TACCATGTAGTTAGAGAACATATTTGTAGGTAAGATAAGGCACTTCCCTTTTAAAGTTAATTTTTGCTG
TTTATTGGGGTAAAACCAGATTAGTGTGAGTTTATCAACTATGCAGTAAATTTAGTAAATGGGTGCGCT
CTGATGTATTGGAACCTTATTAGCTATTGTGTATGGAAGGCTTGAAACAACCCAGTCATTCTCTGTGTG
CTGGTTGTCTTGTTTACTTGATATTTTTTCTTTTTTCTTTTGTGGTGGATAGACTGTTGAACACTTTTAC
AGTTTTAACTATGAATATTGTGAAAAGTTACTGTTTTATCATGGC

3' probe (710 bp)

CCCACAAGAACAGTGTGCGAGGGCTTGAGTTCTAGACCTCAGCAGAGTTACTGTGACTCAGCCGAATCC
TTTCCCACATGGCCTGTCTGTAGCAGACTTCTGAGTGGGGTTGGTGGGGTTGTGGAGCACATAGCTTT
TCCTCATTCTGTCTTTGAGTAATTGCGTTACAGAATATCAATCCCTGTTAGGATAGAGTAGGAGATAA
AGAGAATACCAGAGATGACTGGATTAACGCGCTACTGCCCTCTGCTGGTAGTTCTGAGTGTAGCTG
TTGGGATGTCGGTGCTTTTAGAGGCAGGGATTTGATATCCGAGTTCCCTAGGCCTCTAATGCAGTTTGT
CCAGTGCCTTAAAAGTAGTCTCTGCTTGAACCTCTCAGGTTCCATCCACGTTGCCATTGACTTCAACATC
AGCATCTCAGGGATCCCAGGCAGCGAACACTGCCACCAAGTGTTCAGAGTGGCTGTGAACGGAATT
ATTATTATTAG

Tail genomic DNA isolation and genotyping PCR conditions

Tails were digested in 500 µl tail buffer (50 mM Tris/HCl, pH 8.0, 100 mM EDTA, 100 mM NaCl, 1% SDS) with 2.5 µg of Proteinase K at 55°C overnight. After spinning at 16.1g for 10min supernatants were transferred into a new tube and the DNA was precipitated by adding 500 µl isopropanol. Samples were spun in the centrifuge at 16.1g for 10 min, and the resulting pellet was washed with 1 ml 70% Ethanol. The pellet was dried and resuspended for at least 1 hour at 37 °C in 100 to 150 µl of 10mM Tris/HCl, pH 8.0. The extracted genomic DNA (1-1.5 µl) was used for PCR.

PCR primers used (referred to Figure 2.1D):

Wild-type: 494bp

RRoligo428: GACCACTGGAGTCCTGCTTC

RRoligo429: GCCCAGGTTTTGAACCCTAT

KI (Neo): 632 bp

RRoligo428: GACCACTGGAGTCCTGCTTC

RRoligo431: GGGGAACCTCCTGACTAGGG

Reaction mix for 25µl Reaction

1X Taq buffer(-MgCl₂)

2-Materials and methods

2mM MgCl₂
0.5µl dNTPs mix (stock 10mM)
0.5µl primer mix (stock 20nM each)
1.0µl Tail DNA (100-200ng)
0.5µl Taq Pol (from Heidelberg)
ddH₂O to 25µl final volume

PCR cyler program:

1. 94°C-3minutes
2. 94°C-20 seconds
3. 57°C-30 seconds (for primers 428/429), 63°C-40seconds (for primers 428/431)
4. 72°C- 30 seconds
5. GO TO 2, repeat 35 times
6. 72°C-5minutes

Sequenced PCR products from wild-type *Tdrd9* allele (494 nt)

GACCACTGGAGTCCTGCTTCAGAAAATTGTAAGTGCCAAGAGCCTGATGGAGTTCACGCACATCTTCA
TTGAC **GAG**GTGAGTGTTCTGCTGCCTGGCAGGGCCTTATTTCCGGGCAGTGTGAGCTGCCATCCTGTGT
CCTTGCAACAAGGTTCTTCTGATCTTTTTAACCTTGTGAGGGTGAATTATCTGTTTTATTCTTTGTTCTT
TGTCTCTAAAATATAGGTGGGATGTTGAAGACTGGGAAAAGATTGTTGTATAGAGCAGAGATCTATG
GGCAGGTTGTAGTCAGATCTGGCCCTCTGAGATGTCACTAGAGCCTTCCCTCACAGTGTCTTGTGGTT
TGAATTGCCTCTGGCCTGCATCTCAGCAGGGTGCAGCAGTTGCAAGGGACGTTCTCCAGCCTGGTGAA
GATGCTCTCCACAGCCCACATATGGAGGCAGACTTTCTGCCTCATCAATCAAGGATGACA **ATAGGGT**
TCAAAACCTGGGC

Sequenced PCR product from *Tdrd9*^{KI(Neo)} allele (632nt)

GACCACTGGAGTCCTGCTTCAGAAAATTGTAAGTGCCAAGAGCCTGATGGAGTTCACGCACATCTTCA
TTGAC **CAG**GTGCACGTTCTGCTGCCTGGCAGGGCCTTATTTCCGGGCAGTGTGAGCTGCCATCCTGTGT
CCTTGCAACAAGGTTCTTCTGATCTTTTTAACCTTGTGAGGGTGAATTATCTGTTTTATTCTTTGTTCTT
TGTCTCTAAAATATAGGTGGGATGTTGAAGACTGGGAAAAGATTGTTGTATAGAGCAGAGATCTATG
GGCAGGTTGTAGTCAGATCTGGCCCCCGCGGGAGCTCTACCGGCGCGCCTAGTCGACTTCGA **ATAAC**
TTCGTATAGCATACATTATACGAAGTTAT AAGCTG **GAAGTTCCTATTCTCTAGAAAAGTATAGGAACTTC**
AAGCTTAGGTGGCACTTTTCGGGCTACCGGGTAGGGGAGGCGCTTTTCCCAAGGCAGTCTGGAGCATG
CGCTTAGCAGCCCCGCTGGGCACTTGGCGCTACACAAGTGGCCTCTGGCCTCGCACACATTCCACATC
CACCGATAGGCGCAACCGGCTCCGTTCTTTGGTGGCCCCCTTCGCGCCACCTTCTACTCCTC **CCCTAGT**
CAGGAAGTCCCC

Sequenced PCR product for *TDRD9*^{KI} allele (642nt)

GACCACTGGAGTCCTGCTTCAGAAAATTGTAAGTGCCAAGAGCCTGATGGAGTTCACGCACATCTTCA
TTGAC **CAG**GTGCACGTTCTGCTGCCTGGCAGGGCCTTATTTCCGGGCAGTGTGAGCTGCCATCCTGTGT
CCTTGCAACAAGGTTCTTCTGATCTTTTTAACCTTGTGAGGGTGAATTATCTGTTTTATTCTTTGTTCTT
TGTCTCTAAAATATAGGTGGGATGTTGAAGACTGGGAAAAGATTGTTGTATAGAGCAGAGATCTATG
GGCAGGTTGTAGTCAGATCTGGCCCCCGCGGGAGCTCTACCGGCGCGCCTAGTCGACTTCGA **ATAAC**
TTCGTATAGCATACATTATACGAAGTTAT AAGCTG **GAAGTTCCTATTCTCTAGAAAAGTATAGGAACTTC**
CTCGAGTTTAAACTAAGCGGCCGCAATTAATGGTACCTCTGAGATGTCACTAGAGCCTTCCCTCACAGTG
TTCTTGTGGTTTGAATTGCCTCTGGCCTGCACCTCAGCAGGGTGCAGCAGTTGCAAGGGACGTTCTCCA
GCCTGGTGAAGAGTGCTCTCCACAGCCCACATATGGAGGCAGACTTTCTGCCTCATCAATCAAGGATG
ACA **ATAGGGTTCAAAACCTGGGC**

2-Materials and methods

Sequenced PCR product for *TDRD9*^(Neo) allele (403nt)

```
CTGGAGCCAGTGTGTGTCAGGTCTGTTGACTTAGGAGCAGTGTAGGCCAACAATGTCCATGTCTGGGA  
TGGAAGGCCGCAAGCCTTGGTGGCGCGCCTAGTCGACTTCGAATAACTTCGTATAGCATAACATTATAC  
GAAGTTATAAGCTGGAAGTTCCTATTCTCTAGAAAAGTATAGGAACTTC AAGCTTAGGTGGCACTTTTCG  
GGCTACCGGGTAAGGGAAGCGCTTTTCCCAAGGCAGTCTGGAGCATGCGCTTTAGCAGCCCCGCTGGG  
CACTTGGCGCTACACAAGTGGCCTCTGGCCTCGCACGCATTCCACATCCACCGGTAGGCGCCAACCGG  
CTCCGTTCTTTGGTGGCCCCTTCGCGCCACCTTCTACTCCTCCCTAGTCAGGAAGTTCCCC
```

(GAG to CAG): mutated codon at the end of exon 5 that gives the E275Q point mutation

FLP sequence

LoxP sequence

PCR PRIMER POSITIONS

Generation of Mvh knock-in mouse

A similar strategy was used to generate the *Mvh* knock-in/knock-out mouse. Briefly, the *Mvh* gene locus was targeted using a genomic fragment that replaced exons 14 to 17 and carries a single point mutation in the exon 16 (E457Q) within the ATPase motif (DEAD→DQAD), thus generating a putative catalytic-dead RNA helicase (Figure 2.2A). In addition, it contains *loxP* sites flanking the exons 16, allowing deletion of the exon to produce a knock-out mutant. The mice were created and maintained as described for *Tdrd9*.

The founder was crossed with Del-FLPeR female mice to remove the PGK-Keo cassette to generate the *Tdrd9*^{E257Q} allele (hereafter referred to as *Mvh*^{KI}). Backcrosses with C57BL/6j Rj (Janvier labs) wild-type females were performed to transfer the mutation to the C57BL/6 background to obtain heterozygous mice carrying the targeted allele. Heterozygous *Mvh* null allele (*Mvh*^{+/-}) was generated by crossing the founder with Del-Cre female mice. This led to deletion of exon 16, leading to a frame shift and a stop codon in exon 17. *Mvh*^{KI/-} were generated by intercrossing, but not characterised phenotypically. Mice were genotyped with genomic PCR, using specific primers that were positioned around the site of mutagenesis (Figure 2.2B). Southern blot validation and sequencing of PCR products is still undergoing.

2-Materials and methods

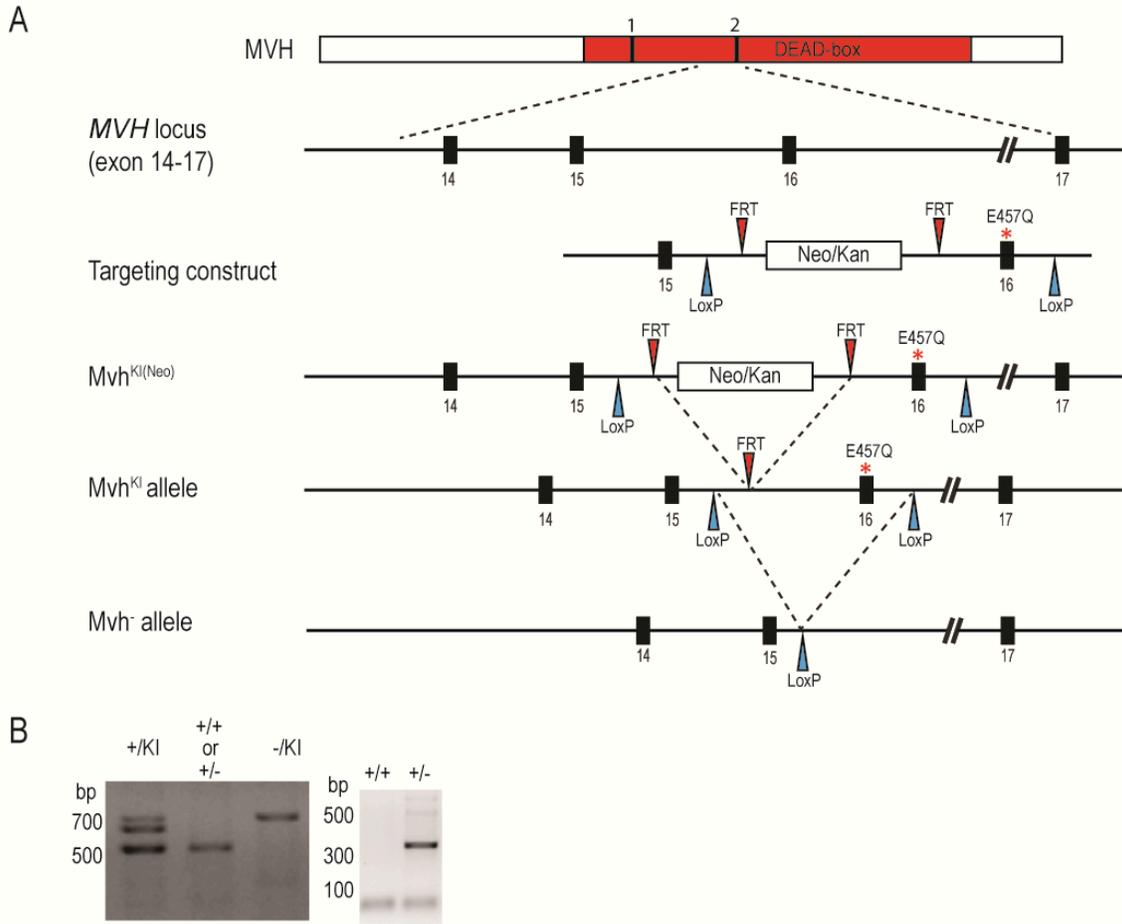


Figure 2.2 Generation of *Mvh* mutant mouse. **(A)** *Mvh* targeting strategy. **(B)** Genomic PCR for the indicated phenotypes. A different set of primers is used to distinguish *Mvh*⁻ allele.

2.3.2 Antibodies

Two different antigens for mouse TDRD9 were used to produce polyclonal antibodies: N-term fragment (1-53 aa) and C-term fragment (1064- 1383 aa). Both of them were cloned in pETM-11 as 6xHis- fusion proteins and used to transform *E. coli* Rosetta strain. Bacteria were grown till OD=1 and protein expression was induced with 1M IPTG at 37°C overnight. The bacterial cell pellet was resuspended in 1 x PBS (containing 1 mM PMSF), and sonicated with MISONIX Sonicator S-4000. Cell lysate was centrifuged at 15000 x g for 40 min at 4°C. Pellet was then resuspended in buffer (100 mM Tris-HCl pH 7.0, 20 mM EDTA, 5 mM DTT), and was further centrifuged at 15000 x g for 40 min at 4°C. High salt wash buffer (50 mM Tris-HCl pH8.0, 60 mM EDTA, 6% (v/v) Triton X-100, 1.5 M NaCl and 5 mM DTT) was used to resuspend the pellet. A final wash with high salt wash buffer without Triton X-100 was

2-Materials and methods

performed and the pellet was finally dissolved in denaturing buffer (8M Urea, 150mM Tris-HCl pH8, 50mM Glycine). Insoluble antigens were further purified on Nickel column in denaturing conditions and eluted with elution buffer (6M Urea, 150mM Tris-HCl pH8, 250mM Imidazole). Antigens were mixed with adjuvant for production of polyclonal antibodies in rabbits. Each antigen was injected into one rabbit to produce a polyclonal antibody.

When indicated, antibodies were affinity-purified from crude immune sera using the respective antigens crosslinked to CNBr activated sepharose beads (GE Healthcare; 17-0430-01). First, antigens were re-solubilized with dialysis in 0.1M NaHCO₃ pH 8.3, 0.5M NaCl overnight at 5°C and crosslinked to CNBr activated sepharose beads (GE Healthcare) in the same buffer. The crosslinking reaction was quenched in blocking buffer (0.1M Tris-HCl pH8.0, 0.5M NaCl) and unbound antigen was washed away with alternative cycles of low Ph buffer (0.1M NaAc pH4.0, 0.5M NaCl) and blocking buffer. Crosslinked antigens were then incubated with 1ml final bleeds crude serum overnight in PBS in 10ml final volume. Beads were washed twice in PBS, twice in PBS (0.1%Tween) and the Ab was eluted with Glycine 0.1M pH2.5 NaCl 150mM. Acidic pH was neutralized with neutralization buffer (0.5M Tris-HCl pH8, NaCl 150mM) and purified antibodies were concentrated to 0.2-0.7 mg/ml and stored at -20°C in 50% glycerol.

Commercial antibodies were purchased: anti-His (Amersham; 27-4710-01), rabbit anti-HA (Santa Cruz; sc-805), anti-actin (Santa Cruz sc-1616-R), anti γ -H2Ax (Abcam AB2893), monoclonal rabbit anti-Flag (F2555). For immunofluorescence studies the following secondary antibodies were used: anti-rabbit or anti-mouse-IgG coupled to fluorescent dyes Alexa Fluor 488/594 (Invitrogen). Antibodies for mouse monoclonal MILI (Reuter et al., 2009), rabbit polyclonal MIWI2 (Pandey et al., 2013) and rabbit polyclonal MIWI (BTO) (Reuter et al., 2011) are described. Mouse LINE1 ORF1p polyclonal antibody was a kind gift of Dónal O'Carroll, EMBL, Monterotondo, Italy. Polyclonal antibodies to *Bombyx* PIWI proteins Siwi and Ago3 were raised against fragments consisting of 1-200 aa of the respective proteins in both rabbits and rats (Xiol et al., 2012).

2.3.4 Western blots

10-15% SDS-PAGE were transferred onto a nitrocellulose membrane using semi-dry blotting equipment (Bio-rad). The membrane was blocked with 5% milk in PBS-T for

2-Materials and methods

approximately 60min and incubated for 1 hour with the primary antibody in 5% milk PBS-T. Dilutions of the antibody ranged from 1/150 to 1/2000 depending on the signal obtained. After extensive washes with PBS-T, the membrane was incubated with a 1/10.000 dilution of anti-mouse or anti-rabbit antibodies coupled to HorseRadish Peroxidase (HRP, GE Healthcare), followed by three washes in PBS-T (at least 10 minutes each). Membranes were developed using ECL+ (GE Healthcare) and chemiluminiscence films (GE Healthcare).

2.3.5 Immunoprecipitations

HEK293 and testes extract were lysed in a buffer containing 50mM Tris pH8, 150mM NaCl, 5mM MgCl₂, 0.01% NP-40, 10% glycerol, 0.5% Triton X-100 and Complete –EDTA free– protease inhibitor cocktail (Roche). For testis extracts, tunica albuginea was carefully removed and in the case that small RNAs had to be extracted after the IP, the buffer was supplemented with 2mM Vanadyl-RC (Sigma) and 50µg/mL tRNA to prevent RNA degradation. Cells/testes were homogeneized in a douncer and the lysate was cleared by centrifugation. For normal western blot examination 40µl of cleared lysate were dissolved in SDS-loading dye, denatured 5 minutes at 95°C and loaded on SDS-PAGE.

For TDRD9 IPs, anti-N-terminal and C-terminal antibodies were coupled to protein G-sepharose beads (GE Healthcare) following the manufacturer's instructions. Briefly 20-50µl of beads were incubated with the affinity-purified antibody (0.2/0.5µg of antibody for µl of beads) in a total volume of 1mL of PBS. After overnight incubation the beads were washed twice in PBS, once in lysis buffer and used for the IP. When required, antibodies were chemically cross-linked to the beads using dimethyl-pimelimidate (Sigma) in order to reduce the background in Western blot and for mass spectrometry experiments. Extracts were incubated with the beads for 2-4h at 4°C and washed extensively with a buffer containing 10mM Tris pH 8, 150mM NaCl, 5mM MgCl₂ and 0.01% NP-40. Immunoprecipitates were resolved by SDS-PAGE and analysed by Western blotting or silver staining. For MIWI2 and MILI IPs three pairs of P0 testes were collected for each genotype and IPs were performed as previously described (Reuter et al., 2009).

2-Materials and methods

For immunoprecipitations from Bmn4 cells, cells were grown in a 6-cm petri dish and transfected (Fugene; Roche) with 2 µg of expression plasmids. For tandem purification of the Amplifier complex, cells were grown in six 6-cm petri dishes and transfected with expression plasmid for 2X-FLAG-Vasa^{DQAD}. 48h post-transfection, cells were removed from their growth surface using a cell scraper (Cell scraper, Costar; catalogue no. 3010) into 1x PBS and collected by gentle centrifugation (0.6 x g). Cells were lysed by douncing (thirty strokes) in lysis buffer and spun at 10,000 rpm. The lysate was incubated with 50µl of Anti-FLAG M2 Affinity Gel (Sigma, A2220) for 2-3h at 4°C. After four washes (20mM Tris, pH 8.0, 150mM NaCl, 5mM MgCl₂ 0.01% NP-40), FLAG-tagged protein were eluted with 150µg/ml of 3x-FLAG peptide (Sigma, F4799) in 800µl wash buffer for 40min in a rotating wheel at 4°C. The beads were spun down and the supernatant was retained. Approximately 50 µl of protein G Sepharose beads (GE healthcare, 17061801) were washed three times with 1x PBS and incubated over-night with 100µl of Siwi polyclonal antibody from rabbit. Siwi-bound Sepharose beads were washed three times with 1xPBS, resuspended in 200 µl of wash buffer and incubated directly with the FLAG-eluted proteins for 1h. Beads were then washed three times with wash buffer and complexes were eluted with SDS loading dye. Complexes were separated by 8% SDS-PAGE and examined by Western blotting using specific antibodies (a rat polyclonal anti-Siwi antibody was used for detecting Siwi by WB).

For the crosslinking experiment dithiobis (succimidyl propionate) (DPS, Thermo Scientific, catalogue no. 22585) was dissolved in DMSO and diluted in 1xPBS at a final working concentration of 2mM. Every 6-cm petri dish transfected with HA-tagged wild-type Vasa was washed three times with PBS and treated with DPS or DMSO control for 2 hours on ice. To avoid precipitations, the media was changed after 1 hour and at the end of the 2nd hour, the plates were incubated for an additional 10 min at 27°C. Crosslinking reaction was stopped adding lysis buffer and cleared lysates were incubated with 15 µl of anti-HA affinity matrix (Roche) for 2-3 hours at 4°C. Beads were collected by gentle centrifugation, 1,000 x g for 1 min at 4°C. After five washes with wash buffer, bead-bound complexes were used for protein analyses.

2-Materials and methods

2.3.6 RNA extraction and Small RNA library preparation

Small RNAs were extracted from the IP performed with anti-TDRD9, anti-MILI, anti-MIWI2 and anti-MIWI, anti-MVH antibodies from different age mouse testes. After the final wash the beads were resuspended in 300 μ L of a buffer containing 10mM Tris pH 7.5, 0.5% SDS, 5mM EDTA. 30 μ g/mL Proteinase K were added and incubated at 55°C during 30 minutes. The RNA was recovered by double phenol-chloroform extraction and precipitated with ethanol and 7.5 M ammonium acetate at -20°C. RNA was dephosphorylated with rAPid Alkaline Phosphatase (Roche; 4898141001), and was phosphorylated at its 5' end using T4 polynucleotide kinase (Fermentas; EK0031) in the presence of [γ -32P]-ATP (PerkinElmer; 10 mCi/mmol). Small RNAs were resolved in a denaturing 15% urea-PAGE, exposed to a PhosphorImager screen (GE Healthcare) after gel drying and visualized with a Typhoon Scanner (GE Healthcare).

The region of interest (20-40nt) was excised from the gel and the acrylamide was crushed with a round pipette tip. Next, 400 μ l of 0.3M NaCl were added and the mixture was left overnight in agitation at 25°C. The pieces of acrylamide were removed by filtration and the RNA was extracted with phenol-chloroform and precipitated with ethanol. Small RNA libraries from immunoprecipitated RNAs were prepared by using the NEB kit [NEBNext Multiplex Small RNA Library Prep Set for Illumina (indexes 1–12); [cat. no. E7300]. Small RNA libraries were sequenced for 50 cycles with Illumina Genome Analyser Iix (EMBL Gene core facility).

Mouse small RNA bioinformatics

Reads were sorted into individual libraries based on the barcodes, the 3' adapter sequences were removed and the reads were mapped to the mouse genome (mm9). The software used for processing the data (genomic coordinates, annotation etc) from the raw data files are in-house tools developed by the Sachidanandam lab (Olson et al., 2008). Only reads perfectly matching the genome were kept for further analysis.

For MILI and MIWI2 IP libraries only the reads of 24-30nt were kept. The piRNA abundances are shown as a fraction of the sequenced libraries (in reads per million – rpm or in %). To estimate the proportion of primary and secondary piRNAs we divided the reads into the categories based on the presence of U at position 1 (hallmark of primary piRNAs) and A at

2-Materials and methods

position 10 (hallmark of secondary piRNAs). To visualize the piRNA production along the repeats we mapped the piRNAs to repeat consensus sequences allowing at maximum 3 mismatches. Read counts were normalized to library sizes and divided by counts of mapped genomic sites. 5' end starts of piRNAs were then plotted. Annotation of piRNAs to different repeat classes was based either on mapping the piRNAs to the genome and genomic annotation of repeats or it was based on direct mapping the piRNAs to repeat consensus sequences. For ping-pong analysis, the product of piRNA counts was used to calculate the score for the 5' end distance Δ : $score(\Delta) = \sum M(i) * N(i+\Delta)$ where $M(i)$ is the count of produced piRNAs (in rpm normalized to number of occurrences in the genome) with 5' end on the plus strand at a particular position i and $N(i+\Delta)$ is the count of piRNAs which have their 5' end position at minus strand at $i + \Delta$. The distance equal to 0 refers to a situation where the piRNAs share the 5' end nucleotide and the distance 9 corresponds to 10nt overlap of piRNA 5' ends – the ping-pong signature. In the following table the number of sequenced and annotated reads for each library is reported:

run	sample	Tdrd9 genotype	stage	mapped_reads	non_mapped_reads	tot_reads	%mapped
RR301	MILI IP	+/KI	PO	10371936	4717214	15089150	45.48
RR302	MILI IP	KI/KI	PO	11651344	5150413	16801757	44.2
RR303	MILI IP	-/-	PO	12228347	3732848	15961195	30.53
RR304	MIWI2 IP	+/KI	PO	16424774	4867541	21292315	29.64
RR305	MIWI2 IP	KI/KI	PO	14003398	4106292	18109690	29.32
RR306	MIWI2 IP	-/-	PO	11133876	4256722	15390598	38.23

2.3.7 Northern blot

Total RNA was extracted from mouse testes using TRIzol RNA-extraction kit (Ambio, cat. No. 15596-026), further purified with double Phenol-Chloroform treatment, precipitated in Ethanol 75% for at least 20 minutes at -20°C and resuspended in milli-Q water. For Northern blotting, 10µg of total RNA were resolved in a 1% agarose gel containing 6.7% formaldehyde (v/v). The quality of the migration was assessed by ethidium bromide staining and the RNA was transferred by capillarity to a Nylon membrane (Hybond N+, Amersham) for at least 16h in in 20× SSC solution (For 1 L: 175 g NaCl, 88.2 g Sodium citrate dihydrate). After the transfer, the RNA was UV cross-linked to the membrane using a Stratagene “cross linker” (120 mJ/cm² in auto-crosslinking mode). Pre-hybridization was performed for 2h in Church buffer (0.25M

2-Materials and methods

NaPO₄ pH 7.2, 1mM EDTA, 1% BSA, 7% SDS) at 65°C. Probes were labelled with [α -³²P]dCTP (Perkin Elmer) using the Random Primed DNA Labeling Kit (Roche), denatured for 5min at 95°C and used for overnight hybridization at 65°C in Church Buffer. Radiolabeled LINE-1 (L1) probe was prepared from L1 DNA fragment (515 bp–1,680 bp; GenBank accession no.M13002) using Random primer DNA labeling kit (Roche; cat no.1004760001) as previously described (Soper et al., 2008). Radiolabeled IAP probe was similarly prepared from the IAP DNA (GenBank accession no. AF 303453). Next day washing was performed at 65 °C as followed: twice 15 min each with buffer-1 (2× SSC, 0.1% SDS), twice 15 min each with buffer-2 (0.2× SSC, 0.1%SDS). The membrane was exposed to a PhosphorImager screen and visualized with a Typhoon Scanner. The signal of 23S and 18S RNA visualized by ethidium bromide staining was used as a loading control.

2.3.8 Southern blot

Analysis of CpG methylation at the level of LINE1 promoter in different genetic backgrounds was performed with methylation-sensitive Southern blot. Probes and primers for these techniques were the same used for northern blot. DNA was extracted from testes of mice of different ages using DNAzol (ThermoFischer) according to the manufacturer instructions. 5µg of DNA were digested with HpaII and MspI restriction enzymes in different reactions. Both enzymes recognize the same DNA sequence for cleavage, but MspI is insensitive to CpG methylation while HpaII cannot cut if the CpGs are methylated. Therefore, HpaII will give different digestion patterns for a given sequence in case the samples have a different methylation status. Digestion was carried out overnight at 37°C in a buffer containing 1mM DTT, 1mM Spermidine, 20U RNase H and the samples were resolved in a 1% agarose gel. After the run, the gel was incubated in 0.25M HCl for 15minutes, followed by two incubations in a buffer containing 0.5M NaOH and 1.5M NaCl (10 and 45 minutes respectively) and finally neutralized in 1M Tris pH 8 and 1.5M NaCl for 20 minutes. The DNA was transferred by capillarity to a Nylon membrane and the following steps were performed as described for Northern blotting.

2.3.9 Histology and immunofluorescence of mouse testes sections

Immunofluorescence of mouse testes cryo-sections

Collected testes were washed in PBS and immediately fixed in 10ml of 2% paraformaldehyde at 4°C for 3 hours on a rotation wheel. Tissues were washed twice in PBS and dehydrated in 15% Sucrose in PBS for almost 3 hour (till the testes sink to the bottom of the falcon tube). After a further dehydration step in 30% sucrose overnight, tissues were embedded in home-made cryomold filled with OCT (Fisher Scientific, cat. No. 14-373-65) frozen on dry ice. Embedded tissues were sent to Histology service in EMBL, Monterotondo, Italy where 7µm tissue sections were mounted on glasses and stored at -80°C. For Immunofluorescent experiments sections were allowed to dry at RT for 30 min and fixed in cold 4% paraformaldehyde (on ice) for 10 min. Slides were then washed in PBS at RT (2 x 5 min) and once in distilled water (5min). Antigen retrieval was eventually performed with Heat-Induced Epitope Retrieval (HIER). Briefly slides were submerged in 600 ml of 10mM Citrate Buffer pH6 and heated in a Microwave at full power (600W) for 20 minutes. Tissues were allowed to cool down at RT for at least 45 minutes, washed in PBS and permeabilised in 0.3% Triton X-100 in PBS at RT for 10 minutes. Slides were washed twice in TBS-0.1%Tween20 (TBS-T) and blocked for 30 minutes at RT in a humidified chamber in 5% normal goat serum in TBS-T. Primary Antibodies were diluted in blocking buffer at different concentration (purified anti-MILI 1:100, third bleed anti-MIWI2 1:50, purified Anti C-ter and anti N-ter TDRD9 1:10, final bleed Anti C-ter and anti N-ter TDRD9 1:50) and incubated overnight at 4°C. Next day slides were washed twice in TBS-T, incubated with secondary antibody conjugated with florescent label (anti mouse 488, anti rabbit 488, anti mouse 594, anti rabbit 594) in a humidified chamber for 45-60 minute (dilution 1:1000). Slides were washed twice in TBS-T and incubated with DAPI (0.5 µg/ml, Bio-Rad, cat. No. 10043282) 5 minutes to conterstain the nuclei. Sections were finally washed twice in TBS-T, once in ddH₂O and mounted with Slowfade Gold Antifade Reagent (Life technology, cat. No. S36942). Pictures were taken using Leica TCS SP2 AOBS, inverted confocal microscope.

2-Materials and methods

Histology of mouse testes sections

To prepare the paraffin sections, the mouse testes were washed in PBS, and fixed in 4% paraformaldehyde overnight at 4 °C. After washing in PBS, testes were dehydrated in 70% ethanol and stored in 70% ethanol at 4 °C. Samples were sent to Histology service in EMBL, Monterotondo, Italy where they were further dehydrated in 80%, 90%, 96% and 100% ethanol (90 min for each step), followed by incubation in xylene (30 min x 3). Xylene was removed and replaced with paraffin, and incubated at 56-58 °C. Testes were then transferred into plastic molds (polysciences, mold S-22) filled with paraffin, and paraffin was allowed to become solid at room temperature. The testes sections (~7 µM thickness) were prepared using microtome and mounted on the Superfrost Plus slides with 10% ethanol. The sections were allowed to stretch at 42 °C and then stored at RT. For histology analysis, the slides containing the paraffin sections were placed in a glass slide holder filled with xylene (5 min x2) to remove the paraffin. For rehydration, the slides were incubated in 100% ethanol, 96% ethanol, 70% ethanol, 50% ethanol and water (2 min for each step). Sections were stained with Hematoxylin solution, Harris modified (Sigma, cat. No. HHS16) for 1 min and rinsed in water. To destain the colorant sections were incubated in acidic alcohol (1% HCl in 70% ethanol) for 5-10 sec and rinsed with water. Sections were stained with Eosin Y solution with phloxine (Sigma, cat. No. HT110332) for 1 min and washed with water. For dehydration, the sections were incubated in 96% (30 sec), 100% ethanol (2 min) and xylene (5 min x 2). Few drops of Permount (Fisher Scientific, cat. No. SP15-100) were deposited on the sections and immediately covered with coverslips. The sections were examined and pictures were taken using Zeiss Axio Imager Z1 fluorescence microscope.

2.4 Expression of recombinant proteins and crystallization

2.4.1 Recombinant protein production in insect cells

Mouse full length TDRD9, *Bombyx mori* full length Vasa and Siwi were cloned in pACEBac2 as a 6xHis-Strep- fusion proteins and the resulting plasmid were transformed to DH10EMBacY competent cells, where the recombination occurs and the bacmid is formed (Bieniossek, 2012). Recombinant baculovirus stocks were generated as described (Bieniossek et

2-Materials and methods

al., 2012) in Sf21 cells and used for infecting exponential growing Sf21 or High five cells. The final amount of virus to perform every infection was determined empirically as the dilution that could arrest cell growth in 24-48 hours. Three to four days after cell proliferation arrest cell pellets were collected with gentle spin (800rcf for 15min) and suspended in lysis buffer supplemented with protease inhibitor (Roche complete EDTA-free, 1 tablet for 50 ml of lysis buffer), or stored at -80°C until further analysis. After 2 minutes sonication lysates were spun down at 21000 rpm for 40minutes and the proteins were purified by Ni⁺²-affinity chromatography.

TDRD9 was purified in a buffer containing (Hepes pH8 20mM, NaCl 300mM, MgCl₂ 5mM, Imidazole 10mM, Tween-20 0,05%, 10% Glycerol, 2-mercaptoethanol 5mM). The protein was further purified on a Streptactin pre-packed column (StrepTrap HP, GE Healthcare), eluted with 2,5mM D-desthiobiotin (IBA) in lysis buffer and mono-dispersed fractions were collected by gel filtration chromatography (Superose6; GE healthcare).

For partial proteolysis experiment 10µg of purified TDRD9^{DQ} protein were digested for 5 to 60 minutes on ice with 0.2µg of chymotrypsin. Reactions were stopped by adding SDS-loading dye and boiling at 95°C for 4 minutes. Domains that are structurally compact remain undigested and can be visualized by SDS-PAGE. Most relevant bands visualized after Coomassie staining were sent for mass spectrometry analysis to EMBL Heidelberg protein expression facility.

Thermal shift assay was performed by HTX facility at EMBL Grenoble as described before (Dupeux et al., 2011). This fluorophore is added to a protein solution, which is then subjected to stepwise increasing temperatures. As the protein denatures, its hydrophobic core is exposed and the fluorophore partitions into it, producing an increase in fluorescence

Vasa and Siwi were purified in a buffer containing 20 mM Tris, pH 7.5, 75 mM NaCl, 5 mM MgCl₂, 1.5 mM DTT and 5 mM 2-Mercaptoethanol using Ni⁺²-affinity chromatography. The multiBac construct containing Vasa, Siwi and Ago3 was generated by mixing 2ug of the following plasmids: pACEBac2-Sumo-His-Strep-Vasa, pIDK-Ago3 and pIDS-Siwi. Following Cre recombinase (NEB) treatment the multigene fusion plasmid was generated and used for bacmid generation as described above.

2.4.2 Recombinant expression in *E.coli* and Crystallization of BmVasa^{DQAD}

The helicase domain of *Bombyx* Vasa (BmVasa (Helicase): 135-564 aa) carrying the K230N (GKT→GNT) or E339Q (DEAD→DQAD) mutations were produced in *E. coli* as His-tagged fusions. The proteins were purified by Ni²⁺-affinity chromatography, followed by removal of the His-tag by TEV protease cleavage. The protein was further purified on a Heparin column (HI-Trap; GE healthcare) and mono-disperse fractions collected by gel filtration chromatography (Superdex 200; GE healthcare). All three proteins were used for biochemical and biophysical studies.

Only the purified helicase domain of Vasa^{DQAD} was used in crystallization experiments. The protein was concentrated to 15 mg/ml in a solution containing 20 mM Tris, pH 7.5, 75 mM NaCl, 5 mM MgCl₂, 1.5 mM DTT and 5 mM 2-Mercaptoethanol. Purified protein was incubated with a 6-mer (UGACAU) RNA and ATP or the ATP analog (AMPPNP) at a molar ratio of 1:1.2:1.2 (protein:RNA:ATP/AMPPNP). The mixture with AMPPNP was incubated for 5 min at 37°C prior to crystallization trials in order to enhance the binding. Crystals of both complexes were obtained using the sitting-drop vapour-diffusion method in a solution containing 0.1M Hepes, pH 7.5 and 25% PEG 3350 after one week at 4°C. For data collection at 100 K, crystals were snap-frozen in liquid nitrogen with a solution containing mother liquor and 30% (v/v) glycerol. Complete data sets were collected to a resolution of 2.1 Å for Vasa^{DQAD}-ADP-Pi-RNA and to 1.9 Å Vasa^{DQAD}-AMPPNP-RNA on beamline ID14-EH4 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France).

Crystals of Vasa^{DQAD}-AMPPNP-RNA belong to the space group $P2_12_12_1$ with unit cell dimensions $a=52.5$ Å, $b=78.6$ Å and $c=104.6$ Å. The data were processed using XDS (Kabsch, 2010). The structure was solved by molecular replacement with PHASER (McCoy et al., 2005) using structure of the *Drosophila melanogaster* Vasa helicase domain (PDB ID: 2DB3) (Sengoku et al, 2006) as a search model. The structure was built in COOT (Emsley and Cowtan, 2004) and refined in REFMAC5 (Murshudov et al., 1997) to final R -factor of 19.6% and R_{free} of 23.5% (Table 2.2) with all residues in allowed (98.1% in favoured) regions of the Ramachandran plot, as analysed by MOLPROBITY (Davis et al., 2004). Crystals of Vasa^{DQAD}-ADP-Pi-RNA belong to the space group $P2_12_12_1$ with unit cell dimensions $a=52.9$ Å, $b=77.8$ Å and $c=107.3$ Å. The structure was solved by molecular replacement with PHASER (McCoy et al., 2005) using

2-Materials and methods

above Vasa^{DQAD}-structure as a search model. After rebuilding with COOT (Emsley and Cowtan, 2004) the structure was refined in REFMAC5 (Murshudov et al., 1997) to final *R*-factor of 23.1% and *R*_{free} of 28.3% (Table 2.2) with all residues in allowed (98.6% in favoured) regions of the Ramachandran plot. The BmVasa^{DQAD} construct we used had an additional unintended mutation (R342C), but this Arg residue normally points out of the ATP-binding pocket, so does not interfere with ATP binding (Sengoku et al., 2006).

Table 2.2 Data collection and refinement statistics of the two crystal structures of Vasa^{DQAD} obtained

	BmVasa^{DQAD}- RNA-ADP-P_i-	BmVasa^{DQAD}- RNA-AMPPNP
Data collection		
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Cell dimensions		
<i>a, b, c</i> (Å)	52.9 77.8 107.3	52.5 78.6 104.6
Resolution (Å)	100-2.1 (2.18-2.1) ^a	100-1.9 (1.98-1.9)
<i>R</i> _{merge}	13.4 (66.9)	12.6 (52.3)
I / σ(I)	8.3 (2.4)	10.9 (2.7)
Completeness (%)	99.4 (95.8)	98.1 (98.4)
Redundancy	4 (4)	3.8 (3.6)
Refinement		
Resolution (Å)	44-2.1	47-1.9
No. reflections	24708	32474
<i>R</i> _{work} / <i>R</i> _{free}	22.4/27.7	17.3/20.8
Average <i>B</i>-factors	30	16.2
R.m.s. deviations		
Bond lengths (Å)	0.006	0.005
Bond angles (°)	1.196	1.041

^a Values in parentheses are for highest resolution shell.

2.5 Biochemical *in vitro* assays

2.5.1 ATPase assay

To test Vasa ATP hydrolysis activity, reactions were performed by mixing 10 µg protein with 0.1 µl of [γ -³²P]ATP (3000 Ci/mmol) in the absence (low substrate condition) or presence (high-substrate condition) of 10 µM cold ATP. The final reaction volume was 20 µl and the buffer contained 20 mM Tris, pH 7.5, 75 mM NaCl, 5 mM MgCl₂, 1.5 mM DTT, 5 mM 2-Mercaptoethanol. For TDRD9 similar conditions were used. Briefly, 5 µg of protein were incubated in a buffer containing (Hepes pH8 20mM, NaCl 300mM, MgCl₂ 5mM, Tween-20 0,05%, 10% Glycerol, 2-mercaptoethanol 5mM), 1 µl of [γ -³²P]ATP (3000 Ci/mmol) in absence or presence of 25nM or 0.25µM cold ATP. Reactions were incubated at 37°C for 30 min (temperature was reduced to 14°C for TDRD9) and stopped by addition of 5 µl of formic acid. 2.5 µl were spotted on a thin layer chromatography (TLC) plate (cellulose fibers on PET foils, Fluka/Sigma-Aldrich; Cat. No. 50488) and migrated in 0.5 M LiCl, 0.5 M formic acid in a migration chamber for 45 min. Free phosphate can be distinguished from ATP because it migrates faster in a TLC. The foil was then dried at room temperature, exposed to Phosphor Storage screens (GE Health) and scanned (Typhoon scanner; GE Health).

2.5.2 UV cross-linking assays

RNA-protein crosslinking: Proteins (10 µg) were incubated for 15 min at room temperature with 0.06 pmol of synthetic 5'-[³²P]-labeled 10-mer poly(U) RNA (Microsynth) in buffer (20 mM Tris-HCl, pH 7.5, 5 mM MgCl₂, 1.5 mM DTT and 75 mM NaCl) with or without 20 pmol of non-hydrolyzable ATP analog AMPPNP (Sigma-Aldrich), in a final volume of 20 µl. The solution mixture was then carefully deposited on the inner side of the cap of a clean eppendorf tube and irradiated for 7 min with a 254 nm UV lamp (1200 µJ) (Stratalinker 1800, Stratagene) at 4°C, from 2 cm distance. After irradiation, samples were boiled for 5 min in SDS loading buffer and resolved by 10% SDS-PAGE. After electrophoresis, the gels were stained with Coomassie Blue (for visualizing proteins to assure similar input of Vasa proteins in each reaction) and dried. Gels were exposed to Phosphor Storage screens (GE Health) and scanned (Typhoon scanner; GE Health).

2-Materials and methods

ATP cross-linking assay: Reaction mixtures containing 10 µg of protein in 20 mM Tris-HCl, pH 7.5, 5 mM MgCl₂, 75 mM NaCl, 1.5 mM DTT, 5% glycerol and 0.5 µl of [γ -³²P]ATP (3000 Ci/mmol) (Perkin Elmer) were incubated for 30 min at 37°C in a final volume of 20 µl. The solution mixture was then irradiated in the presence or absence of 10 pmol (2 µl) 10-mer poly(U) RNA (Microsynth) with a 254 nm UV lamp at 1200 µJ (Stratalinker 1800, Stratagene) at a distance of 2 cm and at 4°C. Samples were subjected to 10% SDS-polyacrylamide gel electrophoresis and visualized as mentioned above.

2.5.3 *In vitro* footprinting assay

In vitro footprinting assay was performed as already described (Bonneau et al. 2009). Briefly, body-labeled RNA was generated by *in vitro* transcription in presence of 5µl of 800Ci/mmol [α -³²P] UTP (Perkin-Elmer, Cat. No. NEG507T001MC) with the MAXIscript transcription kit (Ambion, Cat No.AM1312M) and gel-eluted on a 15% denaturing UREA gel. Template of 200nt random sequence RNA was obtained by annealing of DNA oligonucleotides containing the T7 promoter sequence. RNA binding reaction was performed at RT for 1h using 2 µl of RNA, 1mM ATP, and 100-500pmol of BmVasa^{DQAD} (Helicase domain) in 20 µl final volume reaction. RNase treatment was conducted at 20°C for 45 minutes using 1 µl of RNaseA/T1 mix (Fermentas, Cat. No. EN0551) per reaction. Reactions were treated with proteinaseK for 20 min at 55°C, extracted in Phenol Chloroform, precipitated in Ethanol, resuspended in loading dye and loaded on a 20% denaturing UREA gel.

2.5.4 Kinetic studies with Surface plasmon resonance (SPR)

Surface Plasmon Resonance (SPR) experiments were carried out on a Biacore 3000 system (GE Healthcare) using a streptavidin-coated Sensor Chip SA (GE healthcare). The chip was washed with 1 M NaCl, 50 mM NaOH solution prior to the immobilization of RNA. A single-stranded 5'-biotinylated 20-mer (5'-GAGCACCCUGUGUUCAUGUCA-3') synthetic RNA (Microsynth) was dissolved in a solution containing 20 mM Tris-HCl, 0.5M NaCl, 5 mM MgCl₂,

2-Materials and methods

and 5 mM 2- β -Mercaptoethanol at a concentration of 0.05 μ M and applied to Fc2 channel of the chip. To reduce mass transport limitations for reliable kinetics assays, repetitive injections of 10 μ l of the above buffer were applied (flow rate = 10 μ l/min) until the response units (RU) reached 150-200 RU (Fc2-Fc1). BmVasa or BmVasa^{DQAD} (helicase domains) were prepared in 20 mM Tris-HCl, 150 mM NaCl, 5 mM MgCl₂, 5 mM 2- β -Mercaptoethanol with 100 μ M ATP and directly applied to a Fc2 channel.

A multicycle kinetic procedure was utilized to measure the binding. Analytes at an appropriate range of concentrations (20 μ M – 1 μ M) were injected through Fc1 and Fc2 channels to measure the binding. The flow rate was set at 20 μ l/min. Data for a period of 600 seconds of association and 3000 seconds of dissociation were collected. The sensor surface was regenerated after every injection with 3 washes of a solution containing 0.1% SDS, 1M NaCl until baseline was reached. The sensorgrams obtained from assay channel (Fc2) were subtracted by the buffer control and the Fc1 channel control, and then overlaid for kinetic fitting to obtain the binding *on* (k_{on}) and *off* (k_{off}) rates, and affinity ($K_D = k_{off}/k_{on}$). The kinetic fitting was carried out with the Biacore 3000 evaluation software using 1:1 Langmuir binding model ($A + B = AB$). χ^2 is used as a statistical measure to evaluate how closely a model fits the experimental data. In general, a χ^2 value lower than 5 is considered as a good fit to sensorgrams with normal noise level.

3-THE ROLE OF VASA IN THE piRNA PATHWAY

Résumé

Le rôle de Vasa en voie secondaire de biogenèse des piRNAs

La voie piRNA réprime les éléments transposables, assurant ainsi l'intégrité du génome ainsi que la fertilité des individus. Chez la drosophile, la plupart des piRNAs ont une séquence anti-sens des transposons. Ils seraient générés par le « processing » primaire des transcrits simple-brin issus des groupements de piRNAs. Leur biogenèse secondaire est initiée par le clivage d'un ARNm de transposon en deux fragments, dont l'un va produire le piRNA secondaire d'orientation sens. Des études génétiques ont impliquées différents facteurs protéiques, comme l'hélicase à ARN de type DEAD-Box, Vasa, dans la biogenèse secondaire des piRNAs. Cependant, leur fonction reste inconnue.

Les hélicases à ARN avec une boîte de type DEAD utilisent l'ATP pour interagir dynamiquement avec l'ARN, elles sont omniprésentes dans complexes multi-protéiques tout au long de l'expression génique. Ici, nous décrivons une mutation (DQAD) dans la boîte DEAD de l'hélicase Vasa de lignée germinale, qui empêche la libération du produit d'hydrolyse de l'ATP, la bloquant ainsi dans une conformation fermée. Ceci fige *in vivo* Vasa avec le complexe transitoire de biogenèse des piRNAs et ses ARN cibles. Nous montrons que la fixation de l'ARN par Vasa assemble dynamiquement, sur des ARNm de transposons, une machinerie de biogenèse ATP-dépendant pour générer de nouveaux piRNAs secondaires d'orientation sens. En accord, des souris dont l'homologue de Vasa (*Mouse Vasa Homologue*, Mvh) porte la mutation DQAD sont stériles. Nos résultats révèlent la base moléculaire de l'amplification des petits ARN qui confère une immunité adaptative contre les transposons.

3.1 Introduction

3.1.1 The role of Vasa in embryogenesis: an old story

Sexually reproducing metazoans segregate and establish their germ line early in development. During early embryogenesis, a few cells are set aside as primordial germ cells (PGCs), which establish the germ line lineage by generating germ line stem cells and eventually differentiate to form the sex cells. A few molecular determinants identify PGCs across the animal kingdom, the most common one being Vasa. Germ cell determination in *Drosophila* is achieved by maternal inheritance of gene products that localise to the posterior pole (pole plasm or germ plasm) of late-stage oocytes and early embryos. These RNA-protein complexes form what are called polar granules and these are essential for development of pole cells, which are progenitors of the *Drosophila* germ line (Illmensee and Mahowald, 1974; Jazdowska-Zagrodzinska, 1966; Mahowald, 1968). *Vasa (vas)* was originally identified as a maternal-effect gene and is required for development of the female germ line in *Drosophila* (Schupbach and Wieschaus, 1986). Subsequent cloning of *vasa* by chromosomal walking (Lasko and Ashburner, 1988) identified its gene product as a protein similar to ATP-dependent helicases, especially to the eukaryotic translation initiation factor 4A (eIF4A).

One of the early suggested roles of Vasa is in translation control during *Drosophila* body plan formation. Early events that lay down the body patterns are controlled by maternally contributed mRNAs that have to be spatially localised in the developing embryo. Among these, *nanos (nos)*, *oskar (osk)* and *gurken (gurk)* mRNAs become localised to the posterior pole of embryos. *Nos* and *gurk* mRNA localisation is dependent on Vasa (Ephrussi et al., 1991; Ephrussi and Lehmann, 1992; Tinker et al., 1998; Wang et al., 1994). These mRNAs are normally repressed in the oocyte, but once at the posterior of the embryo, they get translationally activated. Since recombinant Vasa bound double-stranded RNA and was shown to have ATP-hydrolysis/RNA unwinding activities (Liang et al., 1994), it was hypothesized that Vasa mediates unravelling of secondary structure elements that keep *nos* mRNA repressed during transport (Gavis et al., 1996).

3-The role of Vasa in the piRNA pathway

Apart from the above mentioned role in the *Drosophila* embryo, Vasa is also implicated in the biogenesis of piRNAs. PIWI proteins are gonad-specific Argonautes, that, when guided by piRNAs, silence transposable elements to ensure integrity of germ line genomes. In *Drosophila* three PIWI proteins are expressed: Piwi, Aub and Ago3. Two biochemically distinct piRNA biogenesis pathways are recognized: primary and secondary. The initial population of piRNAs within a germ cell is generated by converting long single-stranded RNAs into 25-31 nt piRNAs via a poorly understood primary processing pathway. Primary piRNAs guide the *Drosophila* nuclear Piwi member, Piwi, for mediating transcriptional silencing on target transposon loci. Alternatively, when associated with the cytoplasmic PIWI protein Aub, the mainly antisense-oriented primary piRNAs identify transposon mRNAs which are destroyed by the endonuclease activity of Aub.

Secondary biogenesis is initiated by cleavage of a transposon mRNA into two fragments, one of which is destined to become the future sense-oriented secondary piRNA that enters Ago3. Slicer cleavages, by the RNA-induced silencing complex (RISC), would normally result in complete destruction of the mRNA. How Aub- and Ago3-guided cleavages escape this fate is unknown. Since different protein factors were genetically implicated in this pathway, it is likely that they protect the freshly-cleaved ends and orchestrate transfer of a cleavage fragment from one PIWI complex to another for maturation as a new piRNA. RNA helicases are involved in a variety of cellular functions ranging from splicing, translation to mRNA localisation and decay (Linder and Jankowsky, 2011). As part of multiprotein complexes, they perform local RNA unwinding, disassembly or remodelling of RNP complexes. Among the factors implicated in secondary biogenesis is the RNA helicase Vasa. Loss of *Vasa* causes a severe drop in piRNA population of germ cells, with a loss of the ping-pong signature between Aub and Ago3 and upregulation of transposons (Malone et al., 2009).

The role of Vasa in piRNA biogenesis is also conserved in mouse since, in the absence of *mouse vasa homologue* (*Mvh*), secondary piRNAs loaded into MIWI2 are not produced. Furthermore, spermatogenesis is blocked at the first meiotic cell division in *Mvh*-deficient mice, with a male-infertility phenotype similar to *Mili*- and *Miwi2*-deficient mice (Kuramochi-Miyagawa et al., 2010).

3-The role of Vasa in the piRNA pathway

To explore the molecular role of Vasa in the piRNA biogenesis pathway, we exploited the *Bombyx mori* BmN4 ovarian cell line that mimics every known aspect of the *Drosophila* Ping-pong cycle (Kawaoka et al., 2009). In BmN4 cell line two PIWI proteins are expressed: Siwi and Ago3. Siwi is the homologue of Aub in *Drosophila* and it is loaded with anti-sense transposon-derived piRNAs. Ago3 is loaded with sense transposon-derived piRNAs that contain the typical 10A-bias and it is engaged in the ping-pong cycle together with Siwi (Kawaoka et al., 2009; Xiol et al., 2012).

3.1.2 A piRNA amplifier complex is assembled on the DEAD box RNA helicase Vasa

To understand the role of Vasa in the piRNA pathway we isolated complexes from transfected BmN4 cells with an HA-tagged version of *Bombyx mori* Vasa. Initial studies revealed the absence of piRNAs or PIWI proteins in Vasa complexes, suggesting potential transient/dynamic interactions. This led us to examine ways to freeze Vasa during the course of its normal function in the cell. Vasa is a member of the DEAD-box family of RNA helicases (Linder and Jankowsky, 2011). DEAD box RNA helicases bind tightly to RNA in the presence of ATP, but require ATP hydrolysis for the release of bound RNA (Liu et al., 2008). To trap Vasa in various functional states, Jordi Xiol (a previous member of the lab) introduced point mutations in the helicase domain of *Bombyx mori* Vasa to hinder either ATP binding (GKT→GNT) or ATP hydrolysis (DEAD→DQAD) (Figure 3.1A) (Pause and Sonenberg, 1992).

Transient expression of Vasa^{DQAD} (Vasa^{DQ}) mutant in BmN4 cells, surprisingly, showed that Vasa^{DQ} but not its WT counterpart could pull down a protein complex consisting of the two PIWI proteins implicated in the ping-pong cycle: Siwi and Ago3 (Figure 3.1B). Vasa^{DQ} was also found to be clamped with RNA *in vivo* (Figure 3.1C). Subsequent, deep sequencing revealed that these were a mixture of transposon-derived mature piRNAs (27-nt RNAs) and piRNA intermediates (12nt RNAs) (Xiol et al., 2014).

These first findings suggested that Vasa could transiently promote the assembly of the ping-pong biogenesis machinery where Siwi slices its transposon target and one of the sliced products is loaded into Ago3. The single amino acid mutation (DEAD to DQAD) in the DEAD box of Vasa appeared to be sufficient to block the natural disengagement of

3-The role of Vasa in the piRNA pathway

this complex. Due to its involvement in piRNA amplification in ping-pong cycle this complex was called ‘Amplifier’.

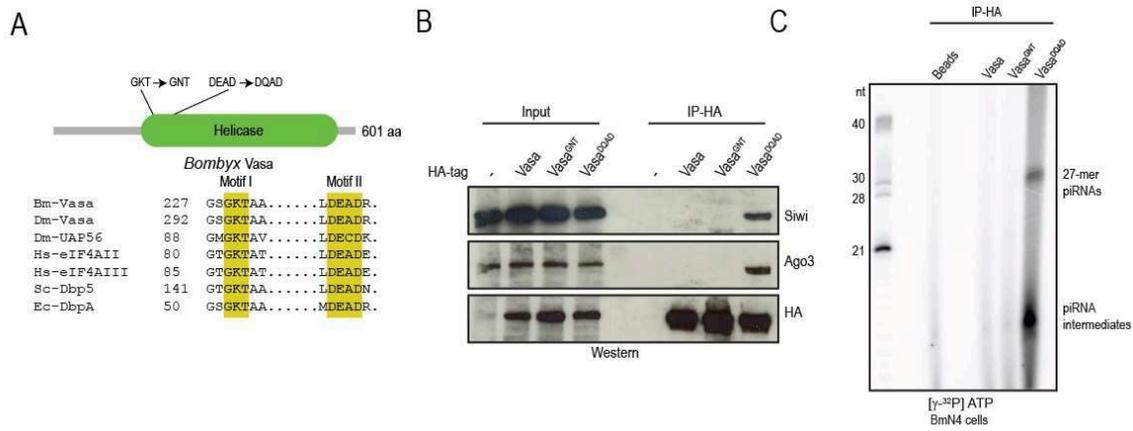


Figure 3.1 Dominant phenotype of Vasa^{DQ} in BmN4 cells. (A) Mutation sites in Vasa, the GKT and DEAD motifs are highly conserved in the DEAD-box RNA helicase family. (B) Vasa transiently interacts with the two ping-pong PIWI proteins Siwi and Ago3, as seen by co-IPs. (C) Vasa^{DQ} is bound to two distinct RNA species, mature piRNA population (27-mer) and smaller 12-mer piRNA intermediates (see section 3.2.4 for details). These experiments were performed by Jordi Xiol.

When I joined the lab in January 2012 this model was being tested but biochemical evidence explaining why the DEAD to DQAD mutation in the helicase domain of Vasa could block the disengagement of the complex was missing. Thus I performed biochemical and structural studies on Vasa and further characterised the Amplifier core complex formed by Vasa, Siwi and Ago3. Most of these results were included in a recent publication entitled: “**RNA Clamping by Vasa Assembles a piRNA Amplifier Complex on Transposon Transcripts**”, together with important experiments performed by other people in the lab. The paper contains the *in vivo* data, which were essential to fully characterise the function of Vasa, and it is reported entirely in the [appendix](#) of this thesis (Xiol et al., 2014).

Furthermore, I will briefly summarise the *in vivo* work that I am conducting in collaboration with Zhaolin Yang (a post-doc from our lab). We introduced a point

3-The role of Vasa in the piRNA pathway

mutation in the *Mvh* gene that encodes for the DEAD to DQAD mutation, generating a knock-in mutant mouse that expresses the mutant protein during spermatogenesis. Expression of *Mvh*^{DQ} has a dominant negative effect on spermatogenesis and MVH mutant protein (hereafter referred as MVH^{DQ}) is blocked on an RNP complex containing PIWI proteins and piRNAs in adult mouse testes.

3.2 Results

3.2.1 Protein purification and biochemical studies

Cell culture experiments in BmN4 point to an important role for Vasa's ATPase cycle in regulating Amplifier assembly, as we were able to trap this complex only by using a mutation that is predicted to abolish its ATPase activity. We therefore took an in vitro approach to gain molecular insight into this process.

To study the biochemical activity of Vasa, I cloned the helicase domain (HD) of *Bombyx mori* Vasa WT (Vasa^{WT}) for expression in *E. coli*. Two point mutants of Vasa (Vasa^{DQ}, Vasa^{GNT}) that should abolish either ATP binding (GKT to GNT) or ATP hydrolysis (DEAD to DQAD) were also produced in the same way. The boundaries of the helicase domain (residues 134-563) were designed based on previous structural studies of the *Drosophila* Vasa (Sengoku T, 2006). All three constructs were expressed as 6xHis tagged and purified using Nickel-NTA affinity resin. Protein preparations were often found contaminated by nucleic acids, as deduced by a high 260/280nm absorbance ratio.

To separate the protein from contaminating bacterial nucleic acids, I applied the purified proteins to a heparin affinity column, which is negatively charged and mimics the RNA sugar-phosphate backbone. A fraction of purified protein devoid of nucleic acid could be separated from contaminants. To further purify, the three proteins were applied to a gel filtration column, which resulted in single peaks eluting at around 15mL (Figure 3.2A).

From the first purification steps we noted that Vasa^{DQ} appeared to behave differently from Vasa^{WT} and Vasa^{GNT}. Loading the fractions on SDS-PAGE, a band corresponding to the size of Vasa was detected in the flow-through of the heparin column when Vasa^{DQ} was purified but this band could not be observed with Vasa^{WT} and Vasa^{GNT}.

3-The role of Vasa in the piRNA pathway

(Figure 3.2A). The fact that this portion of Vasa^{DQ} did not bind to heparin suggests it may be strongly bound to nucleic acids and consequently forming species that cannot attach to the heparin resin. To test the nucleic acid content of the bound fraction, I further purified the two fractions of Vasa^{DQ} from the heparin column (bound and flow-through) by gel filtration (Figure 3.2B). I collected the fraction that eluted at 15mL, which corresponds to the free protein as observed before, and extracted nucleic acids. These were radiolabelled with [γ -³²P] ATP and visualised on a denaturing urea gel. As expected, the protein eluted from the bound fraction of the heparin column does not contain nucleic acids (see 'bound' lane on the gel on Figure 3.2B). On the contrary the protein extracted from the flow-through fraction of the heparin column contains nucleic acids. Interestingly, treatment with RNase generates a band shift, indicating that Vasa^{DQ} is mainly bound to bacterial RNA during purification.

Concluding, I optimised a purification protocol that allowed us to obtain pure Vasa protein preparations carrying different mutations in the DEAD-box of its helicase domain. Vasa^{DQ} but not Vasa or Vasa^{GNT} partially co-purifies with bacterial RNA, suggesting strong RNA binding, nevertheless, these contaminants could be removed with a heparin affinity purification step.

3-The role of Vasa in the piRNA pathway

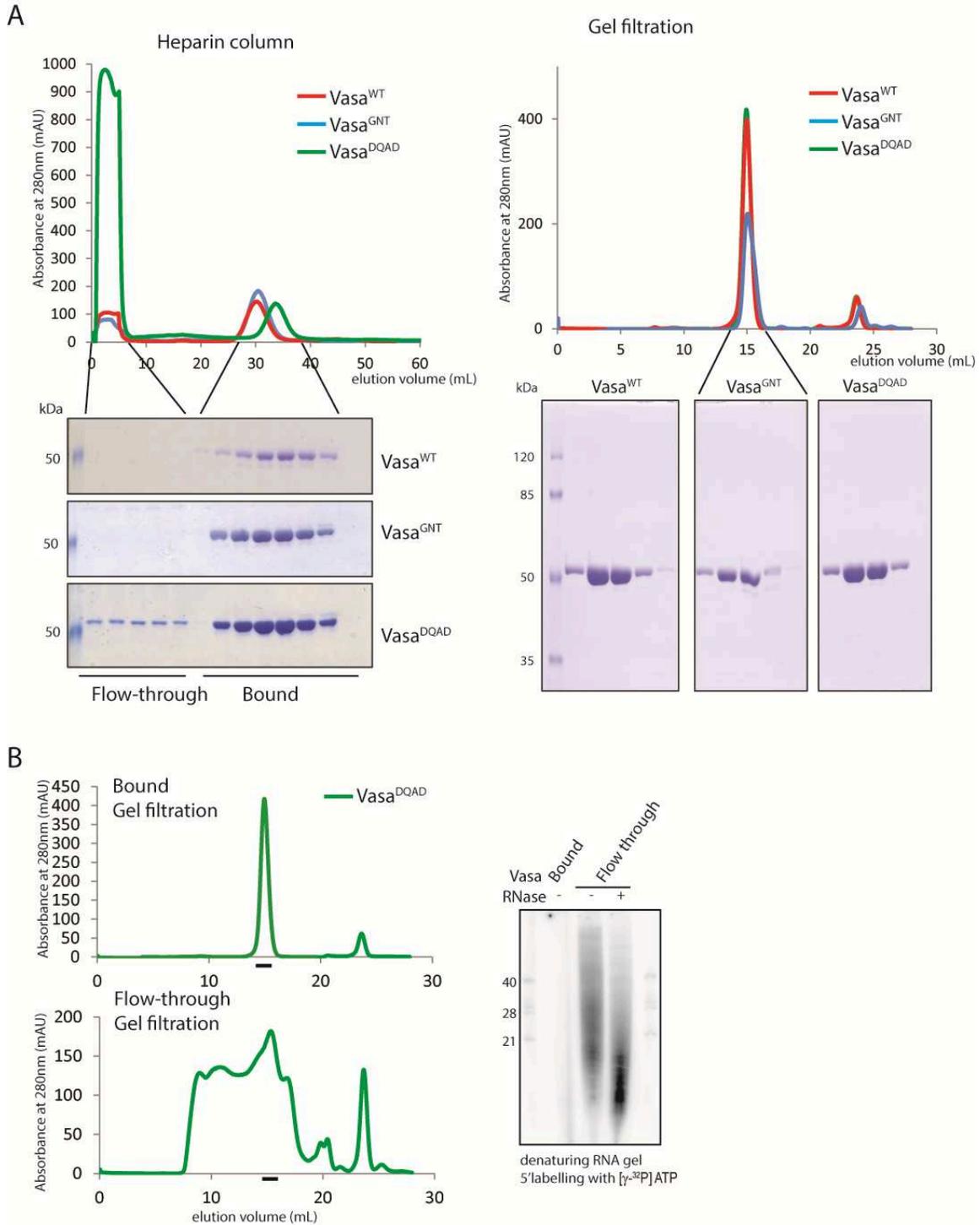


Figure 3.2 Purification of Vasa. (A) The HD of Vasa^{WT}, Vasa^{GNT} and Vasa^{DQ} were purified with a Nickel-NTA affinity column followed by Heparin affinity purification and gel filtration. (B) Vasa^{DQ} is also present in the flow-through of the heparin column, indicating that it forms aggregates with bacterial RNA. Vasa^{DQ} from the bound and flow-through fractions of the heparin column was purified on gel filtration and nucleic acids were extracted and radiolabelled. Protein

3-The role of Vasa in the piRNA pathway

from the flow-through but not from the bound fraction is associated with RNA, since RNase treatment generates a band shift.

The purified proteins were tested for their RNA and ATP binding activities using UV-crosslinking experiments. First Vasa, Vasa^{GNT}, Vasa^{DQ} and bovine serum albumin (BSA) were separately incubated with [γ -³²P] ATP and, after UV-crosslinking, the protein-ATP complexes were resolved on SDS-PAGE. The gel was fixed, dried and exposed to autoradiography. Formation of protein-ATP complexes can be visualised by the presence of a band around the size of Vasa (55kDa) only if the protein is cross-linked with radioactive ATP. As expected from studies on other DEAD box RNA helicases where similar mutations were characterised (Pause and Sonenberg, 1992), Vasa^{GNT} failed to bind ATP, since a very weak band could be observed at the expected size (Figure 3.3A). The Lysine (Lys) present in the GKT motif is responsible for interaction with the γ and β phosphates of the ATP (Sengoku et al., 2006); mutation of this Lys to a shorter Asparagine (Asn) in Vasa^{GNT} completely abolishes its ATP binding activity. Both Vasa and Vasa^{DQ} showed binding in this UV crosslinking assay, with a stronger signal observed for Vasa^{DQ} (Figure 3.3A). Furthermore, addition of RNA to the assay does not influence the ATP-binding activity of Vasa, since bands of similar intensities are observed in presence or absence of an unlabelled RNA substrate (cold Poly-U).

With a similar approach I incubated Vasa, Vasa^{GNT} and Vasa^{DQ} with 5'- γ ³²P labelled poly-U RNA and checked the formation of protein-RNA complexes with autoradiography (Figure 3.3B). The three proteins have weak RNA-binding activity by their own, but this was strongly enhanced by the presence of the ATP non-hydrolysable analogue AMPPNP in the case of Vasa and Vasa^{DQ}, since a strong band corresponding to the size of Vasa could be detected in the lanes where AMPPNP was added. Vasa^{GNT} failed to show such an enhanced binding, indicating that inability to bind ATP also precludes RNA binding.

Next, I examined ATPase activity of each of the recombinant proteins by incubating them with [γ -³²P] ATP. ATP hydrolysis can be followed by measuring the amount of free [γ -³²P] formed after migration on a thin layer chromatography (Figure 3.3C). While Vasa displayed robust hydrolysis of ATP, Vasa^{GNT} did not show activity in

3-The role of Vasa in the piRNA pathway

any of the tested conditions. Vasa^{DQ} had different activities in different experimental settings: when only a radioactive ATP substrate was used (low substrate), the protein was behaving as an active ATPase since a strong signal for the free P_i could be detected, similarly to Vasa^{WT} (Figure 3.3C). When the radioactive substrate was competed by an excess of unlabelled (cold) non-radioactive ATP (high substrate), its activity was strongly reduced (Figure 3.3D). This suggests that Vasa^{DQ} is able to bind ATP and induce its hydrolysis but with a much slower rate compared to the WT protein. The ATPase activity of all the proteins tested is not affected by the presence of RNA, since similar signals are observed in presence or absence of cold poly-U.

In conclusion, Vasa^{GNT} fails to bind ATP, mimicking the situation before binding RNA, while Vasa^{DQ} binds RNA in an ATP stimulated manner as the WT protein, but its ability to hydrolyse ATP appears to be reduced. The RNA binding activity of Vasa and Vasa^{DQ} is dependent on ATP binding but not on ATP hydrolysis, since a non-hydrolysable ATP substrate is sufficient to promote RNA binding. The DQAD mutation in eIF4A, a prototypic member of the DEAD box helicase family, is reported to abolish ATPase activity (Pause and Sonenberg, 1992), so we were surprised to observe that Vasa^{DQ} maintained a residual ATPase activity in our low-substrate assay condition. Overall it is not clear why Vasa^{DQ} has enhanced RNA binding activity *in vivo*. To have a better understanding of the molecular mechanism we undertook X-ray crystallography of the RNP complex formed by Vasa^{DQ}.

3-The role of Vasa in the piRNA pathway

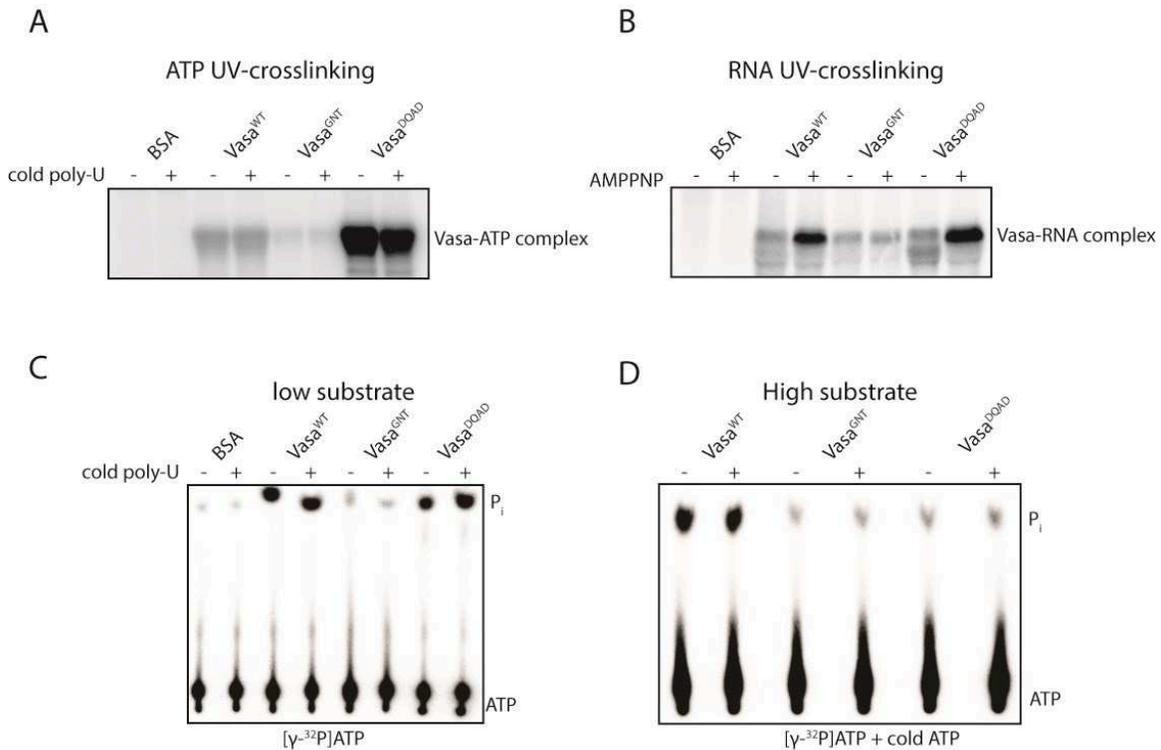


Figure 3.3 Biochemical assays on BmVasa (A) UV crosslinking assay to detect ATP binding and (B) RNA binding with recombinant Vasa carrying the indicated mutations. Bovine serum albumin (BSA) serves as a negative control. Reactions were carried out in the presence (+) or absence (-) of unlabelled (cold) poly-U RNA or the ATP non-hydrolysable analogue AMPPNP. (C) ATP hydrolysis assays with only radioactive ATP as a substrate to reveal generation of inorganic phosphate (P_i). (D) ATPase assay with excess of cold ATP. Reactions were run in presence or absence of unlabelled (cold) poly-U RNA.

3.2.2 Vasa^{DQ} is an ATPase product release trap mutant

Most of the structures of DEAD-box RNA helicases solved so far used an ATP analogue (for example the non-hydrolysable AMPPNP) to block the protein in a closed and stable conformation, which helps the crystallization process (Sengoku et al., 2006; von Moeller 2009). Since Vasa^{DQ} stably binds ATP, as described in section 3.2.1, and since RNA binding is ATP-dependent, I attempted to crystallize the protein in complex with ATP and a 6-nt synthetic RNA (UGACAU). Crystals were obtained using the sitting-drop vapour-diffusion method in a solution containing 0.1M HEPES, pH 7.5 and 25% w/v PEG 3350 at 4°C (Figure 3.4C). Xtals were flash frozen and used for data collection on beamline ID14-EH4 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) where they diffracted to 2.1Å resolution (see Table 2.2 in materials

3-The role of Vasa in the piRNA pathway

and methods for complete statistics). The structure was solved in collaboration with Jan Kadlec and Stephen Cusack, using molecular replacement from the structure of *Drosophila Melanogaster* Vasa (DmVasa, PDB: 2DB3) (Sengoku T, 2006) (Figure 3.4A). The structure shows that the N-terminal domain (NTD) and the C-terminal domain (CTD) enclose the ATP molecule in a deep interdomain cleft and they both interact with the RNA that runs perpendicular to the cleft (Figure 3.4D). The RNA is bound by the NTD with its 3' end and by the CTD with its 5' end. The phosphate-ribose backbone of the RNA lies in a positively charged pocket (Figure 3.4B) formed by both domains but the protein does not interact with the base moiety of the RNA (Figure 3.4A).

Compared to the previously solved structure of DmVasa the main differences reside in the active site implied in ATP binding and hydrolysis (Figure 3.4D). Surprisingly the ATP was found hydrolysed in the BmVasa^{DQ} structure, with the hydrolysis products ADP and P_i both present in the active site. Within the protein, the electron density (in green) showed a defined space between the free P_i and ADP (Figure 3.4E). The catalytic water (W_a) that is believed to mediate the catalysis (Sengoku T, 2006) is absent in our structure and substituted with the free phosphate (Figure 3.4D). This led us to hypothesize that the DEAD to DQAD mutation could impair the positioning of the W_a in the active site, slowing down the hydrolysis process.

To test this hypothesis I crystallized Vasa^{DQ} in complex with a non-hydrolysable ATP analogue (AMPPNP) and the same RNA molecule, in order to lock the protein in its pre-hydrolysis mimicking state. Crystallization drops were set manually and crystals were obtained in the same conditions described above. This structure was solved as before (solved at 1.9Å; see table 2.2 for statistics) and perfectly aligns with the one of BmVasa-ADP-P_i. However the catalytic water was now in the expected position as found in the structure of the DmVasa (compare the active sites of the three structures in Figure 3.4D). This suggests that the first hydrolysis step is indeed happening correctly in Vasa^{DQ}, whereas the release of the hydrolysis products is evidently blocked.

Our structure provides the basis for the retention of the hydrolysis products within Vasa^{DQ}. First, the mutation (E339Q) converts the negatively-charged side chain of

3-The role of Vasa in the piRNA pathway

glutamate (E) to a neutral charge in glutamine (Q), removing a repulsive force that would normally be used to eject the negatively-charged γ -phosphate. Second, the free phosphate is held in place by hydrogen bonding with a number of residues, with the mutated residue Q339 in Vasa^{DQ} providing an additional interaction not available in the native context (Sengoku et al., 2006) (Figure 3.4D). Failure to release hydrolysis products can prevent turnover of the enzyme (Liu et al., 2008), perhaps explaining the increased crosslinking of ATP to the protein (Figure 3.3A), and reduced ATPase activity under high substrate concentrations (Figure 3.3D). Thus, we conclude that Vasa^{DQ} is a product-release-trap mutant.

3-The role of Vasa in the piRNA pathway

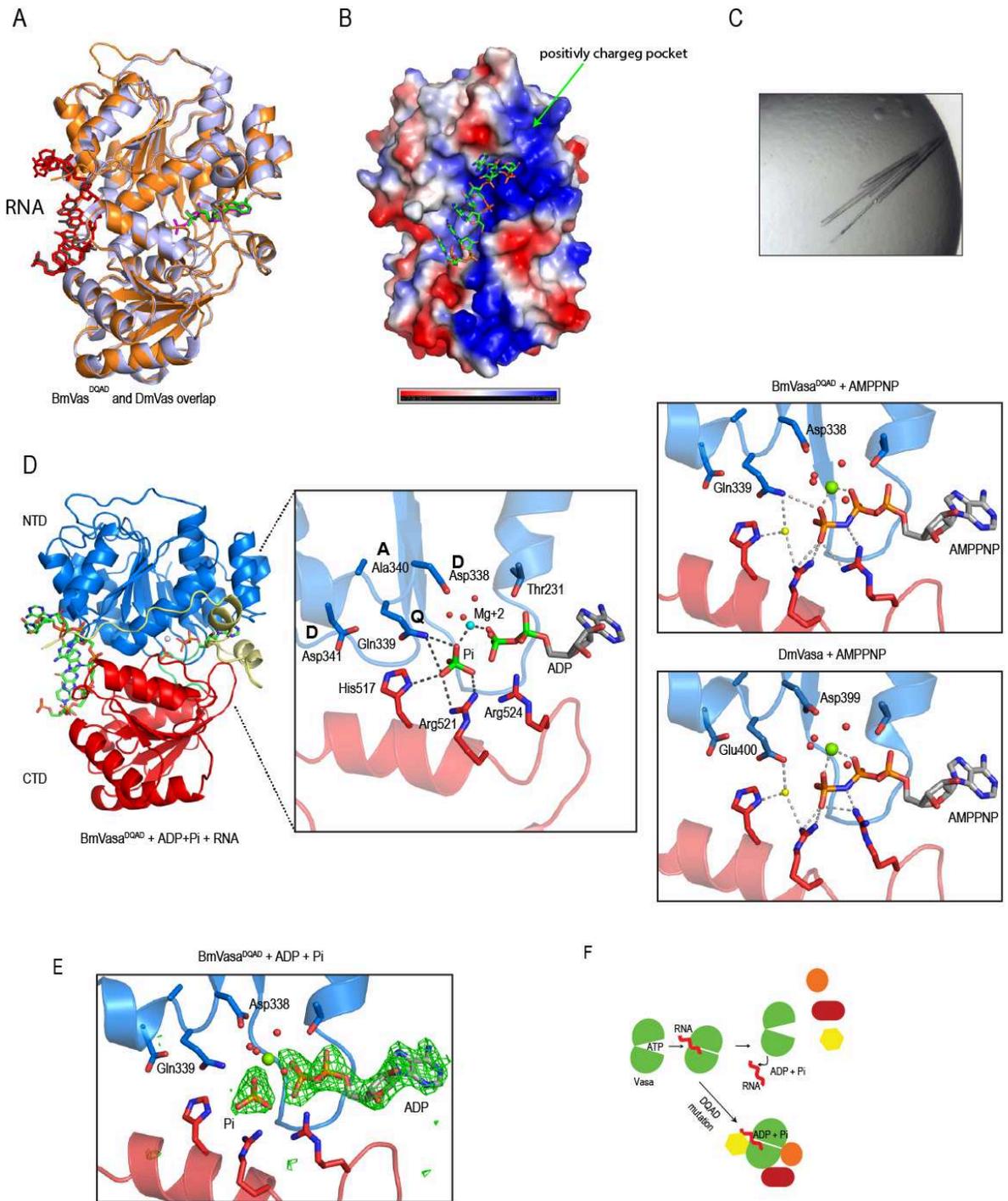


Figure 3.4 Structure of the Helicase domain of *Bombyx mori* (Bm)Vasa^{DQ} in complex with ATP (PDB: 4D26) and AMPPNP (PDB: 4D25). **(A)** Overlap of BmVasa^{DQ} in orange and DmVasa (PDB: 2DB3) in grey. **(B)** The RNA is bound in a positively charged pocket shared by the NTD and CTD. **(C)** Picture of the crystal of Vasa^{DQ} in complex with ATP and RNA. **(D)** Crystal structure of BmVasa^{DQ}-RNA-ADP-P_i complex (left). The two RecA-like domains (blue and red) of the helicase module are represented as ribbons, whereas RNA and ATP are indicated as ball-

3-The role of Vasa in the piRNA pathway

and-sticks. ATP added during crystallization with BmVasa^{DQ} becomes hydrolysed into ADP and P_i. Magnesium ion (Mg²⁺) (blue ball) and water molecules (red balls) are indicated. Interaction of the glutamine (Q339) (grey dash lines) traps P_i in the protein, creating a product-release mutant. On the right the structure of Vasa^{DQ} in complex with AMPPNP is shown and the catalytic water (W_a) is in yellow. The active site of DmVasa in complex with AMPPNP is also shown (PDB: 2DB3; Sengoku et al., 2006). **(E)** Actual electron density map (in green) in the active site of Vasa^{DQ}-ADP-P_i. **(F)** Cartoon illustrating the effect of the mutation on Vasa functions *in vivo*.

3.2.3 Insights into the ATP hydrolysis mechanism of DEAD-box helicases

The structure of Vasa^{DQ} in complex with ADP+P_i shows a post-hydrolysis stage where the free P_i derived from ATP hydrolysis is trapped inside the catalytic pocket. The comparison of the pre-hydrolysis state-mimicking structure (Vasa^{DQ}-AMPPNP) and the post-hydrolysis structure (VASA^{DQ}-ADP-P_i) reveals that after hydrolysis the phosphate replaces the water molecule that was found in the lytic position in the pre-hydrolysis stage (Figure 3.5A). Further regularization of these structures followed by the side-chain minimization and H-bond optimization procedures by standard Internal Coordinate Modelling (ICM-Pro) protocols allowed us to visualize a putative “histidine switch,” that could be involved in the sensing of the released phosphate and participates in the energy transmission pathway in the protein (molecular dynamics experiments conducted by Mikhail Grigoriev) (Figure 3.5). Because the heat transfer in proteins is mediated by the H-bond-forming residues, H517 could be the first residue in the energy transmission path that connects the ATPase CTD and regulatory NTD of Vasa. Indeed, in the pre-hydrolysis stage, His517 forms the H-bond with T375 located in the N-term domain, linking the two domains and probably promoting a closed conformation. After hydrolysis, His517 switches its H-bond to the free phosphate, releasing the connection to the NTD (Figure 3.5B). This switch, together with other interactions that are still to be identified, may help in releasing the energy that brings the two domains in a closed conformation.

It is important to note that both structures were solved using the Walker B (E339Q) mutant protein. Accordingly, we do not know the nature of the residue that has served as a proton acceptor during the lytic water activation mechanism, which is the first and essential step in ATP hydrolysis. Further experiments will be necessary to resolve

3-The role of Vasa in the piRNA pathway

this interesting puzzle, especially because the histidine-threonine pair that we think is actively involved in the reaction is not only conserved among a wide variety of RNA helicases from SF1 and SF2 families, but also among other ATPases that belong to the AAA+ superfamily, suggesting that this interaction may generally be conserved in evolution (Linder and Jankowsky, 2011).

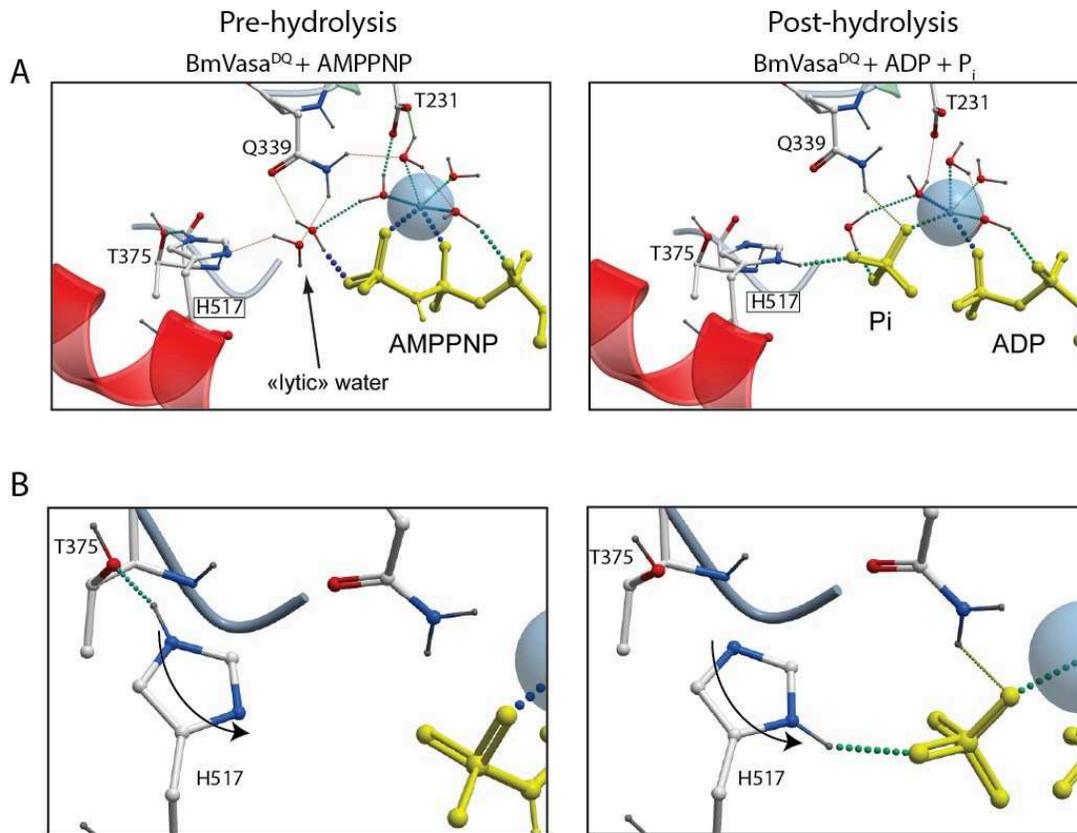


Figure 3.5 A possible role in sensing the organic Phosphate by Histidine517 is proposed (prepared by Mikhail Grigoriev and re-adapted). (A) Details of the possible interactions between the amino acids and the water molecules involved in the catalysis. Oxygen are indicated as red dots and Nitrogen as blue dots. ATP and AMPPNP are in yellow. H517 on the CTD of the protein (in red) forms an H-bond with T375 that is situated on a loop in the NTD. After hydrolysis H517 switch its H-bond to the free Phosphate originated from the hydrolysis. The additional H-bond between the mutated Q339 and the Pi is indicated as a yellow dashed line, the other H-bonds are indicated as green dashed lines. (B) Details of the histidine switch model without water molecules.

3.2.4 Blocking phosphate release impairs RNA binding kinetics of Vasa

Our structural analysis revealed that Vasa^{DQ} is an ATPase product release mutant and this probably affects its RNA binding kinetics, since failure to eject ATP hydrolysis products prevents also RNA release (Liu et al., 2008). To investigate in this direction I monitored the kinetics of ssRNA-binding using Surface Plasmon Resonance (SPR) (Figure 3.6A). Briefly, a ssRNA molecule is immobilised on a surface plasmon chip to which a light wave is applied. When the protein binds the ssRNA on the chip, the reflection angle of the light source changes and this is measured in response units (RU) (Figure 3.6A). The intensity of this shift is measured as protein is added at increasing concentration and consequently the affinity for the ssRNA is calculated (association constant (k_{on})). Similarly, when the protein is washed away from the chip, the dissociation constant (k_{off}) can be measured.

In the presence of ATP, the k_{on} for RNA-binding was similar for both Vasa and Vasa^{DQ} (55.1 and 15.7 M⁻¹s⁻¹ respectively) but the latter dissociated from the RNA with a rate constant (k_{off}) ~40-fold lower than that of Vasa, indicating a significantly tighter binding (Figure 3.6A). This confirms our first hypothesis that ATP binding and consequently RNA binding have similar kinetics in Vasa and Vasa^{DQ} but the failure to release hydrolysis products in Vasa^{DQ} restrains the protein in the closed conformation bound to RNA.

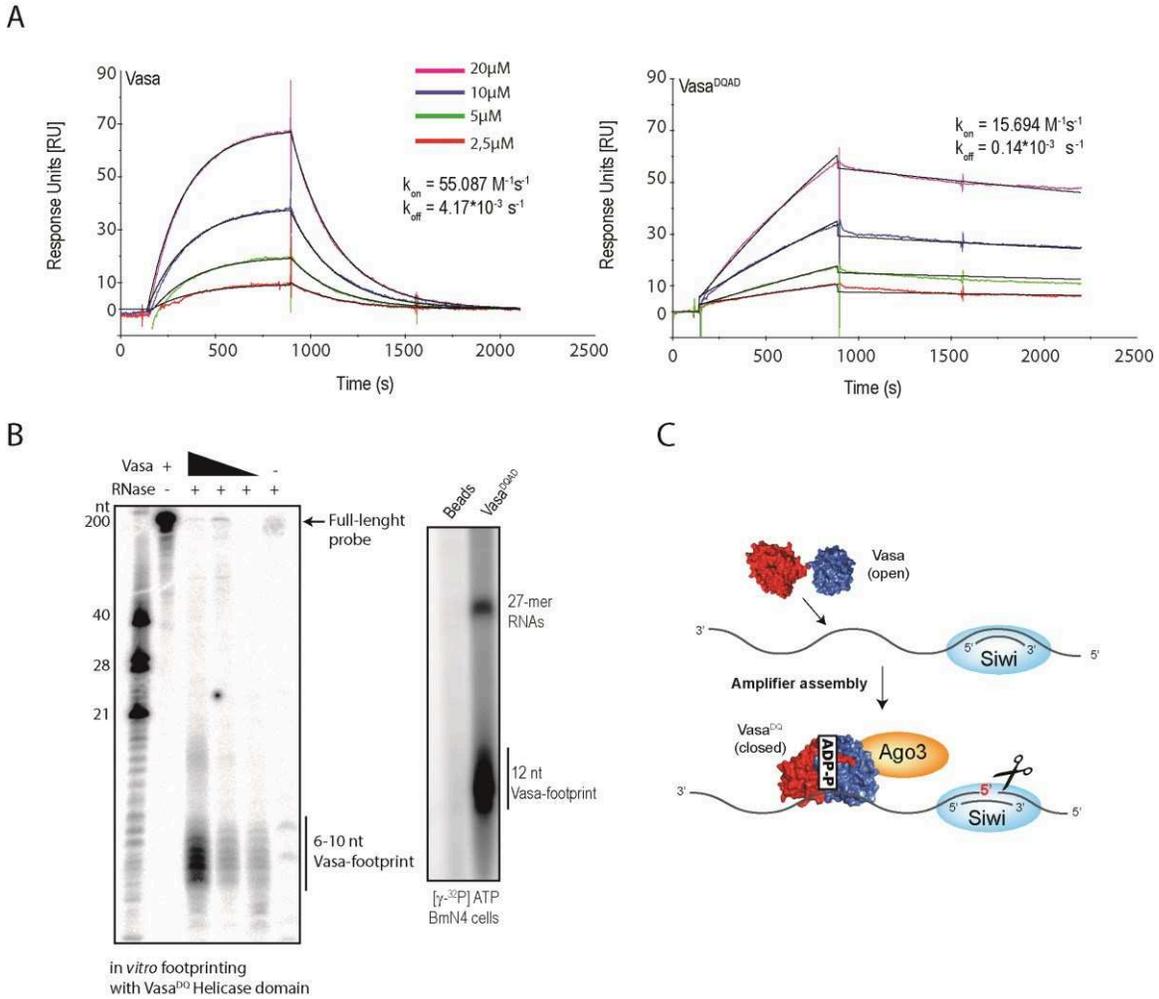
This change in the dissociation constant through RNA explains why Vasa^{DQ} is found clamped on its RNA targets *in vivo*. As mentioned before, deep sequencing of the small RNA population pulled down by HA-tagged Vasa^{DQ} in BmN4 cells showed two distinct families of RNA (27nt and 12nt) (Figure 3.6B). Deep-sequencing of these bands suggested that the 12-mer is the portion of RNA that is physically bound to Vasa (footprint), whereas the 27nt band is bound by Siwi and it is only indirectly pulled down by Vasa (see cartoon in Figure 3.6C for clarity). We speculate that since the 12-mer is tightly bound to the helicase core of Vasa^{DQ}, it is protected from being further processed. We wondered if this RNA protection property could be reproduced *in vitro* and performed an RNase protection assay using a body-labelled RNA probe and the purified

3-The role of Vasa in the piRNA pathway

Vasa^{DQ} protein. The protein was incubated with ATP and radioactively labelled RNA and the resulting complex was subjected to RNase treatment. After treatment, the RNA was extracted and visualised on a denaturing urea gel (Figure 3.6B). Fragments of RNA that could be visualised are protected from RNase treatment by Vasa. Vasa^{DQ} was found associated with a 6-10nt RNA fragment that was protected from degradation, which is similar to the 6nt RNA molecule used in crystallization experiments. Interestingly *in vivo*, Vasa protects a slightly longer fragment (9-12nt) (Figure 3.6C right). This difference can be explained by the presence of other factors in the Amplifier that protect few additional nucleotides *in vivo*. Since I used the helicase domain of Vasa^{DQ} for the assay, it is also possible that additional protection *in vivo* is given by the N-terminal domain of Vasa.

Overall we demonstrated that Vasa^{DQ} can protect an RNA fragment *in vitro* and this is an indirect proof that the similar size fragments detected *in vivo* are the bona fide RNA binding sites of the protein. Interestingly, a weak band of 16nt can also be observed in the RNase protection assay (Figure 3.6B). We can speculate that this band represents a longer protected RNA fragment where two proteins are bound in close proximity to each other. Accordingly, this additional band was identified only when a high concentration of protein was used.

3-The role of Vasa in the piRNA pathway



3.2.5 *In vivo* studies

As mentioned above, IP of HA-tagged Vasa^{DQ} pulls down the two PIWI proteins expressed in BmN4: Siwi and Ago3 (Figure 3.1), suggesting that Vasa, Siwi and Ago3 are part of the same RNP complex. Nevertheless we could not exclude the possibility that Vasa is forming two independent complexes with Siwi and Ago3 and that these are co-purified in the IP. To determine if this was the case, I performed a tandem IP where first I pulled down FLAG-Vasa^{DQ} and eluted the complex with the FLAG peptide. From the resulting eluate, I performed an IP with Siwi and checked for presence of Ago3 by WB. Detection of Siwi and Ago3 signals in the final elute demonstrated that the Amplifier complex is trimeric (Figure 3.7A).

This complex could be identified only with Vasa^{DQ} but not with Vasa, suggesting that its transient nature precludes its identification in a wild-type environment. We wondered if alternative methods, other than single amino acid mutation, could be used to block Vasa in the Amplifier complex. This would exclude that the formation of the Amplifier is solely caused by the anomalous RNA binding of Vasa^{DQ}. I transfected BmN4 cells with HA-Vasa and 48 hours after transfection I cross-linked the cells with DSP (dithiobis(succinimidyl propionate)). From the resulting lysate I performed an IP of HA-Vasa and tested by WB if Siwi and Ago3 could be co-purified (Figure 3.7B). Crosslinking with DSP preserved the interaction of Vasa with Siwi but not with Ago3. This suggests that the interaction of Vasa and Siwi is direct or at least that these two proteins are in very close proximity (the spacer of DSP is 12.5Å long). Several attempts to crosslink Ago3 to Vasa failed, suggesting that Ago3 interaction in the Amplifier complex is even more transient than Siwi. Alternatively, it is also possible that Ago3-Vasa interaction requires additional protein factors, thus precluding the identification in these experimental setting.

To investigate in this direction, I co-expressed the three proteins Vasa, Siwi and Ago3 in a heterologous system, and attempted to reconstitute the Amplifier core complex *in vitro*.

3-The role of Vasa in the piRNA pathway

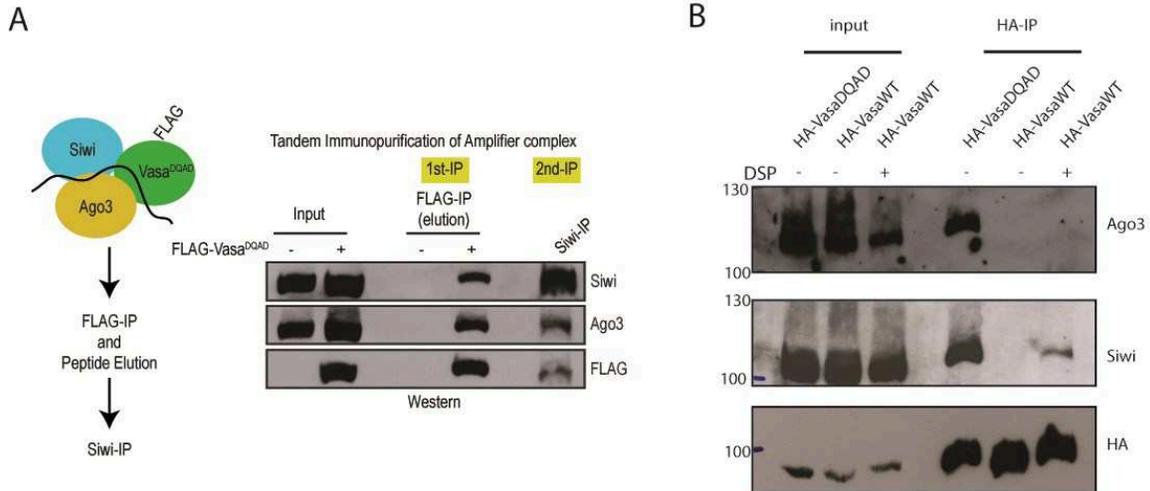


Figure 3.7 Vasa forms a trimeric complex with Siwi and Ago3 *in vivo*. (A) Tandem affinity purification showing presence of Vasa^{DQ} and the ping-pong PIWI partners in the same complex. (B) Co-precipitation of Ping-pong PIWI partners with HA-Vasa^{DQ}, but not with HA-Vasa isolated from transfected BmN4 cells. However, treatment of cells with the *in vivo* crosslinking agent DSP (+) allows detection of Siwi with HA-Vasa

3.2.6 Purification of the recombinant Amplifier complex (Vasa-Siwi-Ago3)

Previous members of the lab attempted to express PIWI proteins from different species in *E. coli* but failed in obtaining soluble proteins. Thus, I decided to express Vasa, Siwi and Ago3 in insect cells. First I cloned full length Vasa and Siwi independently, as Sumo-His-Strep-tagged proteins, and tested their expression. Vasa was well expressed in insect cells and could be purified with Ni²⁺-NTA affinity resin. It was identified by presence of a single band at around 120kDa when the elution of the Ni²⁺ resin is loaded on an SDS-PAGE and stained with Coomassie (Figure 3.8A). Siwi is also expressed but at a very low level and with high contamination (Figure 3.8B).

Protein expression issues can be solved by using combinatorial vectors that consist of just one vector expressing several components. Different subunits of a complex that are not soluble on their own may become soluble when co-expressed with a protein partner (Bieniossek et al., 2012). Following this strategy, I cloned Vasa, Siwi and Ago3 in three different plasmids that were later recombined in the MultiBac system using the CRE recombinase *in vitro*. The resulting plasmid contains the cDNA sequences of 6xHis-Strep-tagged Vasa and untagged Siwi and Ago3. The recombinant plasmid was used to

3-The role of Vasa in the piRNA pathway

express the trimeric complex in insect cells. Initial small scale experiments showed that only Vasa could be eluted from Ni²⁺-NTA affinity resin and could be visualised in Coomassie stained gel (Figure 3.8C). Since by Coomassie I could only detect Vasa, the same eluate was also analysed by WB for presence of Siwi and Ago3 (Figure 3.8C). Both PIWI proteins could be identified by immune staining, suggesting that they were co-purified with Vasa. In Ago3 staining a strong unspecific band was observed at a molecular weight of around 70kDa, probably reflecting protein degradation (Figure 3.8C). The fact that the two PIWI proteins are not visible in Coomassie staining but only by WB indicates that they are not in the same stoichiometry with Vasa.

I tried to scale up the protein production in order to increase the amount of the protein complex but this approach was unsuccessful (Figure 3.8D). Vasa was still the predominant protein and the other bands that were visible on a Coomassie staining after Ni²⁺-NTA purification were identified to be Sf21 contaminants by mass spectrometry (data not shown) (Figure 3.8D). The elution from the nickel resin reported in figure 3.8D was further purified via streptactin affinity column and the TAG was cleaved with TEV protease. Cleavage of the tag generates a band shift visible in Coomassie staining, confirming that the main band identified on the gel is indeed Vasa, since it is the only tagged protein in the complex (Figure 3.8D). The resulting elution was further purified by size exclusion chromatography, which shows two main peaks at 7 and 14 mL (Figure 3.8E). The peak elution from the gel filtration (~ 7mL) usually represents the column's void volume, which contains large molecular complexes and soluble aggregates. Analysis of the resulting elution fractions by Coomassie staining and WB indicated that the main peak at 14mL is predominantly constituted by Vasa alone, since a single band corresponding to the size of Vasa could be identified in Coomassie stained gel (Figure 3.8E). WB for Siwi identified a signal in fractions that elute between 12 and 13mL. This suggests that Vasa and Siwi are forming a complex that elutes earlier than Vasa alone (around 12mL vs 14mL for Vasa), suggesting that a protein complex of higher molecular weight than Vasa alone, which contains Vasa and Siwi, can be purified. For Ago3 staining only the degradation band at 70kDa but not the full length protein was detected (not showed), so we were not sure that it was forming the complex. Unfortunately the

3-The role of Vasa in the piRNA pathway

amount of protein complex obtained was again very low, preventing further biochemical experimentations.

Overall, further structural studies on the Amplifier complex were not possible due to the poor solubility of the PIWI proteins. It is possible that an improved construct would produce a better yield of soluble protein. For example the tag used for purification (which contains the Sumo polypeptide that increases solubility) could have been cloned on the less expressed protein component of the complex. Alternatively, for a cleaner preparation, different tags could have been cloned on both Siwi and Vasa and the complex could have been purified with tandem affinity purification. Finally, other components may be required for complex formation like the Tudor protein Qin, which was identified to be a protein component of the Amplifier complex *in vivo* (Xiol et al, 2014).

3-The role of Vasa in the piRNA pathway

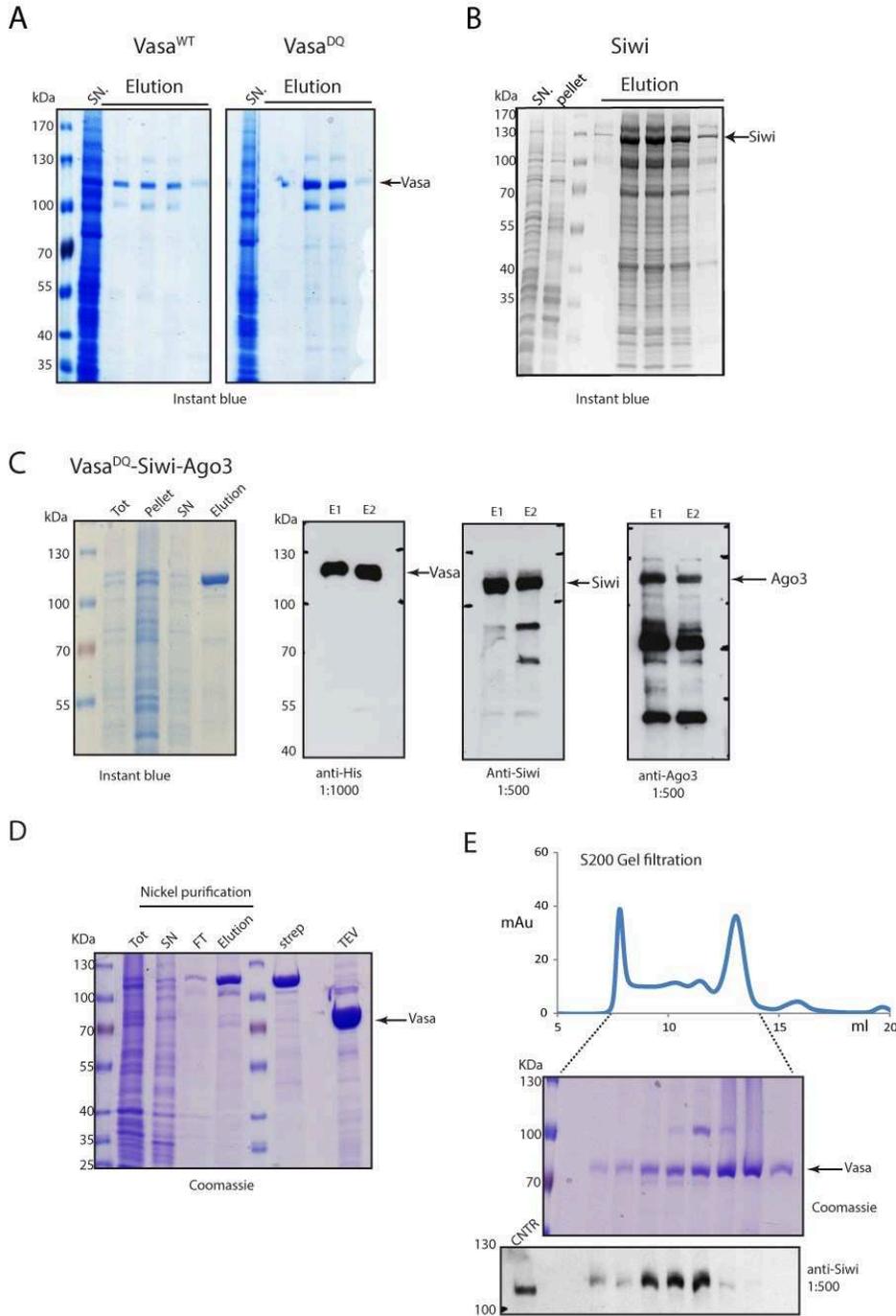


Figure 3.8. Recombinant production of the Amplifier complex. Ni²⁺ purification of Vasa (**A**) and Siwi (**B**) expressed in Sf21 cells. (**C**) Co-expression of Vasa, Siwi and Ago3 in insect cells using the multiBac system. WB after elution from the Nickel resin shows that the three proteins can be co-expressed. (**D-E**) Large scale purification of Vasa-Siwi-Ago3 complex. The TEV-cleaved complex eluted in Figure D was further purified on gel filtration (**E**) and eluted fractions analysed. SN, supernatant; Tot, total lysate; FT, Flow-through.

3.2.7 DQAD mutation in mouse Vasa homologue (MVH) blocks a pachytene piRNA biogenesis complex

Depletion of mouse vasa homologue (*Mvh*) in mouse and mutation of *Vasa* in *Drosophila* shows a similar phenotype in terms of transposon activation and defects in secondary piRNA biogenesis. In *Mvh*^{-/-}, MILI is correctly loaded with primary piRNAs but MIWI2-loaded piRNAs are absent (Kuramochi-Miyagawa et al., 2010). Similarly, in *vasa* fly mutant, Aub is loaded with primary piRNAs and Ago3 with maternally-derived piRNAs but the ping-pong amplification is abolished (Malone et al., 2009). Overall, the role of Vasa in secondary piRNA biogenesis is conserved from insect to mouse. Thus, we wondered if DEAD to DQAD mutation in MVH can also block the protein on a transient piRNA biogenesis complex.

The genome of mouse embryonic stem (ES) cells was targeted to introduce a single nucleotide mutation in the exon 14 of *Mvh* (this knock-in allele will be hereafter referred to as *Mvh*^{KI}). This results in the creation of a single amino acid mutation (E to Q) in the helicase domain of the MVH protein (Figure 3.9A). Mutant mice were generated from one of the targeted ES cell lines (see Materials and Methods). This project was initiated only recently and it is currently on-going. I am working on it in collaboration with Zhaolin Yang, a post-doc in our lab.

First phenotypic analysis of mutant mice revealed that heterozygous *Mvh*^{+KI} mice of both sexes are viable. Furthermore, females of all genotypes are fertile. It was previously described (Kuramochi-Miyagawa et al., 2010) that animals of the *Mvh*^{-/-} genotype display male-specific infertility whereas *Mvh*^{+/-} are fertile. Interestingly, *Mvh*^{+KI} males are infertile, indicating a dominant negative effect of the DQAD mutation. I analysed the histology of adult (P60; 60 days old) testes sections from the *Mvh*^{+KI} mutant and observed the presence of seminiferous tubules devoid of late stage germ cells, as identified by Hematoxylin/Eosin staining. Meiosis can proceed through pachytene stage as identified by the presence of spermatocytes in histological sections (Figure 3.9B). Spermatogenesis appears to be uniformly arrested at the round spermatid stage, since no spermatid undergoes cellular elongation or nuclear condensation. Although more immunostaining experiments need to be performed to understand the exact stage when

3-The role of Vasa in the piRNA pathway

germ cell development is arrest, it seems that *Mvh*^{+DQ} has a similar phenotype to *Miwi* KO (Deng and Lin, 2002) (Figure 3.9B).

Furthermore, immunoprecipitation of MVH from mutant adult testes and extraction of RNA shows the presence of piRNAs bound to MVH, as identified after 5' labelling with [γ -³²P]ATP (Figure 3.9C) (experiment conducted by Zhaolin Yang). This suggests that MVH^{DQ} is clamped together with its RNA targets *in vivo*.

The IP of MVH from adult mouse testes of different phenotypes was also sent for mass spectrometry analysis in order to identify putative protein interactors of MVH (Figure 3.8D) (experiment conducted by Zhaolin Yang). Interestingly, both MVH and MVH^{DQ} pull-down TDRD6 whereas only MVH^{DQ} is present in a complex which contains MILI, MIWI and TDRD1 (Figure 3.8D). TDRD1 is a secondary piRNA biogenesis factor that directly interacts with MILI and it is necessary for the loading of MIWI2 in prospermatogonia but it is also expressed in adult spermatocytes (Reuter et al., 2009) where no evident secondary biogenesis is present. This finding indicates that MVH is forming a transient complex with MILI and MIWI in pachytene spermatocytes and that the disassembling of such complex depends on the ATPase activity of the protein. We still ignore the nature and the function of such biogenesis complex and further experiments are required to validate its exact protein composition.

3-The role of Vasa in the piRNA pathway

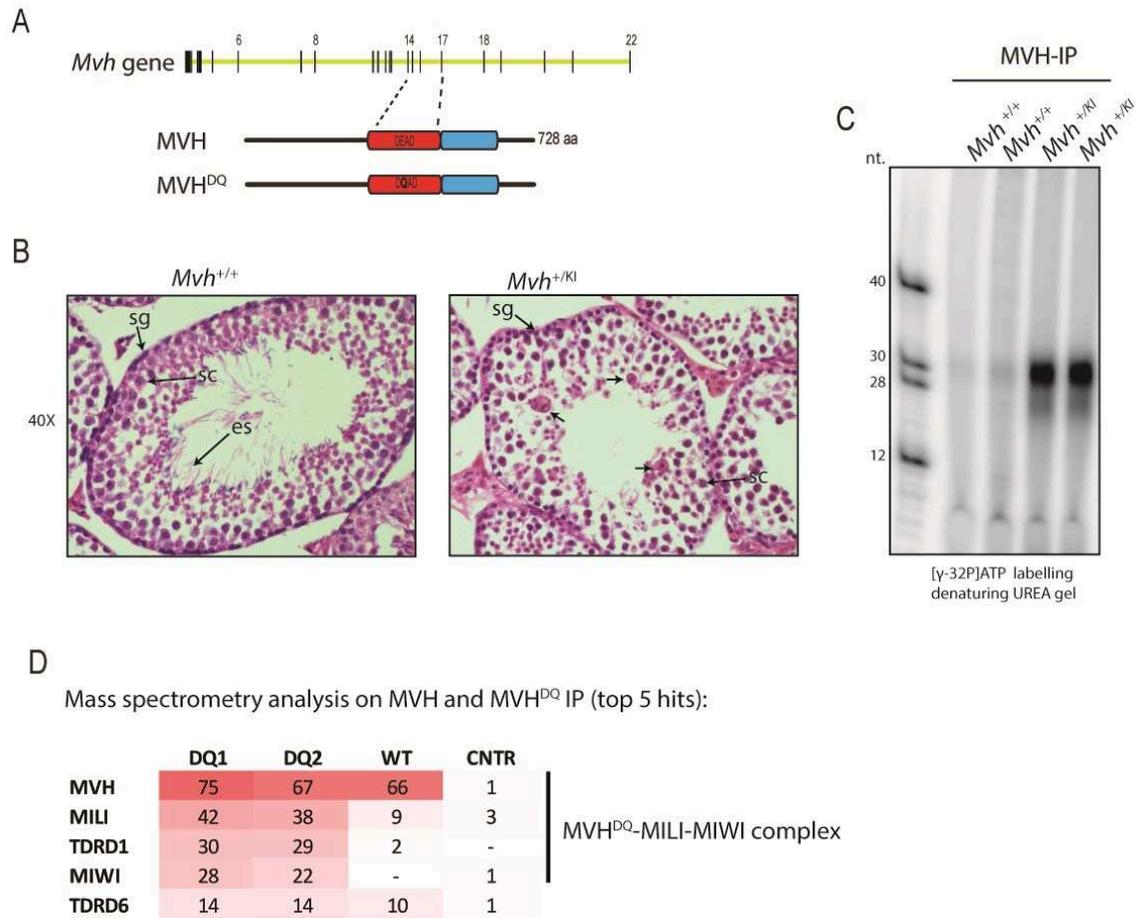


Figure 3.9 Generation of Mvh DQAD knock-in mouse (*Mvh^{KI}*). **(A)** Cartoon showing that a mutation in exon 14 of *Mvh* leads to the E457Q single amino acid substitution in MVH protein (for more details see [Chapter 2.3.1](#)). **(B)** Hematoxylin/Eosin staining from testes sections from adult *Mvh^{+/+}* and *Mvh^{+/KI}* mice. Sg spermatogonial cells, sc spermatocytes, es elongating spermatids. Arrows indicate spermatid-like cells with pycnotic and fragmented nuclei. **(C)** MVH IP shows that piRNAs are pull-down when the mutant protein is expressed. **(D)** Mass spectrometry analysis of MHV IP from different genotypes revealed a presence of MILI and MIWI in the IP from *Mvh^{+/KI}* testes but not from *Mvh^{+/+}*. Abundance of peptides for each protein interactor is reported.

We next wondered if the DQAD mutation in MVH has the same biochemical effects as observed for BmVasa in terms of ATPase activity. I cloned and expressed full-length MVH from *E. coli* as a 6xHis tagged protein and purified it via Nickel-NTA affinity column. Coomassie analysis of the resulting purification steps indicate that MVH was predominantly detected in the insoluble pellet fraction (Figure 3.10). Nevertheless, a

3-The role of Vasa in the piRNA pathway

small fraction of soluble protein could be eluted from the resin and further purified with heparin affinity column and size-exclusion chromatography (Figure 3.10). The protein appears to be in a grade of purity sufficient to perform biochemical assays that will be performed in the future.

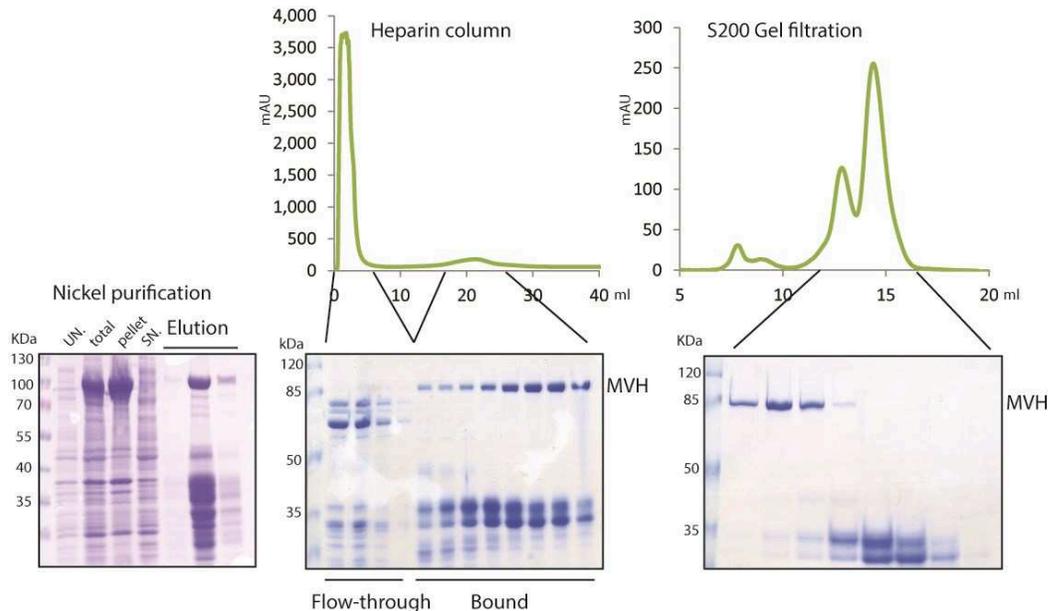


Figure 3.10 MVH purification. MVH was expressed in *E. coli* as 6xHis-tagged protein and purified with Nickel-NTA. The eluted at around 100kDa was TEV cleaved and further purified via heparin affinity column and size exclusion chromatography.

3.3 Discussions and perspectives

In secondary piRNA biogenesis the cleavage product of one PIWI protein (*Bombyx mori* Siwi or *Drosophila* Aub) is loaded into Ago3. Vasa was genetically involved in this process but its biochemical contribution was ignored (Malone et al., 2009). In this study we dissected the function of Vasa in the pathway. Antisense piRNAs loaded into Siwi recognise a transposon transcript through sequence complementarity. This is closely followed by deposition of Vasa on the transcript and by assembly of the Amplifier by the joining of Qin/Kumo and Ago3 within the *nuage*. Once Siwi slices its target, the resulting fragment carrying a free 5' end is loaded into Ago3. ATP hydrolysis and release of the hydrolysis products by Vasa is necessary to transfer the cleaved RNA into Ago3. The resulting sense-oriented new piRNA is further processed on its 3' end and guide Ago3 in the feed-forward step of the ping-pong cycle.

3-The role of Vasa in the piRNA pathway

Identification of this complex was made possible due to a point mutation in the active site of Vasa. DEAD to DQAD mutation generates a mutant that has a reduced ATP hydrolysis activity. The crystal structure of VASA^{DQ} in complex with ATP or AMPPNP revealed the molecular details of this inhibition. Vasa^{DQ} has a reduced ATP hydrolysis rate because after hydrolysis the free Phosphate is blocked inside the active site due to an additional H-bond formed with the mutated Gln. The reduction in ATP hydrolysis rate is mirrored by an impaired RNA binding activity. The mutant can bind RNA with a similar affinity to the WT protein but the dissociation rate is forty times slower. Combination of the two structures and molecular modelling revealed a histidine-switch that happens after hydrolysis of the ATP. His517 from the C-terminal domain loses its H-bond with Tre375 situated on the N-terminal domain and forms an H-bond with the free phosphate generated from ATP hydrolysis. This switch could be one of the first events in the energy transmission path that connects the ATPase C-terminal and regulatory N-terminal domains of Vasa.

Many open questions remain on how this machinery is assembled and how the helicase domain of Vasa can function as a platform for the formation of the Amplifier complex. One important factor that was not considered in this study is the helicase activity of Vasa. Vasa can unwind short RNA duplexes (Sengoku et al., 2006). Biochemical studies previously showed that the unwinding activity of DEAD-box helicases is dependent on ATP binding but not on the release of the ATP hydrolysis products, which is indeed necessary for protein turnover (Fei Liu, 2008). We can speculate that Vasa promotes local unwinding of Siwi sliced product to facilitate its loading into Ago3.

Additional structural studies on the Amplifier complex will be necessary to fully understand the molecular mechanism underneath. These will probably require first a better understanding of which protein domains are directly involved in the complex formation. For example, in our attempts to purify Vasa-Siwi-Ago3 from insect cells, we did not consider the contribution of Qin. Qin plays an important function in *Drosophila* ping-pong by forcing Aub cleavage products into the correct PIWI protein. This prevents Siwi-slicer products from becoming the precursors of PIWI-bound piRNAs but promoting their loading into Ago3 (Zhang et al., 2014a). Thus, Qin may be necessary to

3-The role of Vasa in the piRNA pathway

place Ago3 in a position where its MID domain can accept the 5' end of the new secondary piRNA.

Recently we obtained the first experimental mice expressing MVH^{DQ} and we reported that the mutation has a dominant negative effect on mouse male fertility. Interestingly, a similar dominant phenotype is observed in BmN4 cell line, when BmVasa^{DQ} is expressed (Xiol et al., 2014), indicating a functional conservation of Vasa in the piRNA pathway of different species. Accordingly, as identified for BmVasa^{DQ}, MVH^{DQ} is blocked in an RNP complex containing piRNAs and PIWI proteins (MIWI and MILI). Secondary piRNA biogenesis in adult spermatocytes has been recently proposed to play a role in silencing certain mRNAs that contains piRNA binding sites (Goh et al, 2015). Further characterisation of this mutant mouse may give a better understanding of the ping-pong pathway in pachytene piRNA biogenesis. Interestingly we could not observe a shorter band of 12nt corresponding to the MVH footprint when we extracted RNA from the IP of MVH^{DQ}, contrary to what observed in Bmn4 cells. Why the footprint is detectable in in BmVasa^{DQ} IP but not in MVH^{DQ} IP is still not clear and need further investigations. We can speculate that different expression levels of the mutant protein can influence the detection of the footprint: in BmN4, Vasa^{DQ} is transiently expressed at high level whereas in the knock-in mouse mutant MVH^{DQ} has the same expression profile as the WT protein.

**4 - FUNCTIONAL
CHARACTERIZATION OF TDRD9 IN
THE piRNA PATHWAY**

Résumé

Caractérisation fonctionnelle de l'activité catalytique de Tdrd9 chez la souris

Dans les lignées germinales, les protéines PIWIs et les piRNAs protègent le génome contre l'activation non contrôlée des transposons. Dans la spermatogenèse de la souris, cette voie est particulièrement active durant les premiers stades du développement prénatal des cellules germinales, durant laquelle, deux protéines PIWI sont exprimées : MILI et MIWI2. MILI réside dans les granules péri-nucléaires où elle réprime la transcription de transposons par clivage endonucléolytique. Son activité « slicer » génère des ARN substrats pour MIWI2, qui va se charger avec des piRNA secondaires et migrer dans le noyau, où elle réprime les transposons à leur *loci* génomiques par méthylation *de novo* de l'ADN. Des études génétiques ont impliqué différents facteurs protéiques dans la biogenèse des piRNAs dont TDRD9, une protéine à domaine Tudor. TDRD9 est une putative hélicase/ATPase à ARN, et malgré son rôle essentiel dans fertilité, sa fonction dans la spermatogenèse est inconnue. Dans ce chapitre, nous avons étudié le rôle de TDRD9 dans la répression des transposons.

En premier, nous avons exprimé TDRD9 *in vitro* et prouvé qu'elle possède une activité ATPase par sa boîte DExH. Nous avons ensuite introduit une mutation ponctuelle dans la boîte DExH (DEVH à DQVH) de la protéine qui abolit son activité ATPase et nous avons généré une souris mutante *knock-in* portant cette mutation. Les souris mâle *knock-in* qui en résultent sont stérile avec un blocage dans les stades méiotiques précoces de la spermatogenèse. Ceci prouve que l'activité ATPase de TDRD9 est essentielle pour sa fonction. Ce blocage est probablement une conséquence de dommages non contrôlés de l'ADN générées par l'activation de transposons comme les éléments LINE-1, qui ne parviennent pas à être réprimé pendant méthylation *de novo* de l'ADN. Bien que TDRD9 est important pour la répression des transposons LINE1, elle ne l'est probablement pas *via* un rôle dans la biogenèse piRNAs, vu que les protéines PIWI sont correctement chargées avec les piRNAs dérivés de LINE1. Fait intéressant, une baisse de piRNAs issus des éléments SINE dans MIW2 est observée chez ce mutant, ce qui reflète probablement

un rôle dans le tri des transcrits primaires par TDRD9 vers MIWI2 pendant la méthylation *de novo* de l'ADN.

4.1 Introduction

piRNA biogenesis is a multi-step process that is commonly divided into primary and secondary processing. Primary biogenesis describes the conversion of long single-stranded precursor transcripts from large (50-100 kilobases) genomic regions into thousands of piRNAs. Final maturation of piRNAs involves their loading into PIWI proteins which in turn can post-transcriptionally silence complementary RNA targets by direct endonucleolytic cleavage. Alternatively, some PIWI proteins can migrate in the nucleus and silence transposons at transcriptional level, by promoting DNA methylation or establishment of repressive chromatin marks. In post-transcriptional silencing, one of the cleavage fragments, derived from PIWI slicer activity, is transferred to a new PIWI protein, followed by 3' end trimming to a new mature secondary piRNA. In the fly germline, slicing by the PIWI protein Aubergine (Aub) generates the RNA fragment that is loaded into Ago3 as a secondary piRNAs. As described earlier, the handover of Aub sliced products into Ago3 is dependent on Vasa, a DEAD-box RNA helicase which drives the assembly of the RNP complex formed during secondary piRNA biogenesis in insects. The role of Vasa appears to be conserved in higher eukaryotes since a point mutation in mouse Vasa homologue (MVH) freezes Vasa in an RNP complex containing MIWI, MILI and TDRD1 (chapter 3.2.7).

Other factors containing putative RNA helicase domains such as Tdrd9, Tdrd12 and Mov10l1 play important functions in piRNA biogenesis in mouse but their precise role is not known (Pillai and Chuma, 2012). Mouse knock-out for these genes generates a male-sterility phenotype, with early block in spermatogenesis during prophase 1 of meiosis and de-repression of different transposon families (Pandey et al., 2013; Shoji et al., 2009; Zheng et al., 2010). Mov10l1 is considered a primary biogenesis factor since its depletion leads into loss of all piRNAs (Zheng et al., 2010) whereas Tdrd12 is referred as a secondary biogenesis factor since in its absence MIWI2 fails to be loaded (Pandey et

4 - Functional characterization of TDRD9 in the piRNA pathway

al., 2013). TDRD9 also function in transposon repression both in mouse and flies but its function in the pathway is unknown.

TDRD9 is a well conserved protein that contains a DExH-box RNA helicase/ATPase domain and a tudor domain. It is expressed in the gonads of insects and mice where it plays essential roles in transposon control and gametogenesis (Malone et al., 2009; Shoji et al., 2009). The TDRD9 orthologue in *Drosophila* is known as Spindle-E (Spn-E). The flies mutant for *spn-E* are female sterile and produce embryos with defects in anterior-posterior and dorsal-ventral axes formation (Gillespie and Berg, 1995). Transposon are derepressed in both male and female germ cells (Aravin et al., 2001; Vagin et al., 2006). In the absence of *spn-E*, piRNA pathway components Vasa, Ago3 and Aub fail to localise in perinuclear granules (*nuage*) and remain dispersed in the cytosol (Lim and Kai, 2007; Malone et al., 2009; Snee and Macdonald, 2004).

Sequencing analysis of small RNAs in *spn-E* fly mutants revealed a severe drop in certain classes of piRNAs, which suggested that Spindle-E has a direct role in piRNA biogenesis (Malone et al., 2009). piRNAs produced from germ cell-specific *42AB* cluster are strongly impaired in the mutant with a strong reduction of the ping-pong signature. However, follicle cell-specific piRNAs derived from *flamenco* cluster are unaffected, suggesting that *spn-E* is important for biogenesis of germ cells specific piRNAs only. Germ cells have primary and secondary biogenesis whereas follicle cells have only primary. The reduction of ping-pong signature suggests that Spindle-E play a role in the secondary biogenesis pathway (Malone et al., 2009).

On the other hand, Spn-E was recently linked to primary biogenesis from *in vivo* studies in the BmN4 cell line (Nishida et al., 2015). This cell line is derived from ovarian germ cells of the insect *Bombyx mori* and it expresses most of the factors involved in primary and secondary piRNAs biogenesis (Kawaoka et al., 2009; Kawaoka et al., 2011; Xiol et al., 2014). Immunoprecipitation of endogenous Spn-E from BmN4 cells identified Siwi, the *Bombyx mori* orthologue of Aub, as a protein partner of Spn-E. Siwi fails to get loaded with piRNAs upon *spn-E* knock-down in BmN4 cells (Nishida et al., 2015). This effect was rescued by transient expression of Myc-tagged Spn-E, but not by the ATP hydrolysis mutant *spn-E* (DEVH→DQVH). This indicates that ATPase activity of Spn-E is essential for its function in *Bombyx* germ cells, which is probably to load primary

4 - Functional characterization of TDRD9 in the piRNA pathway

piRNAs into Siwi (Nishida et al., 2015). The importance of the putative helicase activity of Spindle-E in transposon repression was also highlighted in *Drosophila*. Flies harbouring a point mutation in the DExH-box of the helicase domain (DEVH→DKVH and DEVH →DEVQ) phenocopy the *spnE* null mutant, in terms of female sterility and transposons de-repression (Ott et al., 2014).

Similar to its fly orthologue, mouse TDRD9 is also essential for correct male germ cells development (Shoji et al., 2009). *Tdrd9* mouse mutant is male-infertile, with an early block in spermatogenesis during prophase I of meiosis (Shoji et al., 2009). The block appears to be caused by a failure in silencing transposons, like the non-LTR element LINE-1, during *de-novo* DNA methylation. Methylation of LINE-1 in fetal prospermatogonia is controlled by the two PIWI proteins MILI and MIWI2 (Pillai and Chuma, 2012). The molecular link between MIWI2 and DNA methylation is unknown but when MIWI2 is depleted of LINE-1 derived piRNAs, LINE-1 elements are up-regulated and germ cells undergo cell cycle arrest as soon as they enter prophase I of meiosis (Aravin et al., 2008; De Fazio et al., 2011). This happens for example when essential secondary piRNA biogenesis factors involved in MIWI2 loading like *Tdrd1*, *Mvh* or *Tdrd12* are depleted (Kuramochi-Miyagawa et al., 2010; Pandey et al., 2013; Reuter et al., 2009). *Tdrd9* knockout mice have the same phenotype in term of block in spermatogenesis and LINE-1 de-repression, however, MILI and MIWI2 proteins are loaded with piRNAs. MIWI2 nuclear localisation is also preserved. TDRD9 was shown to partially co-localises with MIWI2 in piRNA processing bodies (piP-bodies) and in the nucleus of prospermatogonia. Its interaction with MIWI2 appears to be direct, since the two proteins can be co-purified when ectopically expressed in a human cell line (Shoji et al., 2009). These evidences suggest that TDRD9 acts downstream of MIWI2 in establishing DNA methylation on transposons, perhaps by mediating the interaction between MIWI2 and the DNA methylation machinery.

Overall, the contribution of Spindle-E/Tdrd9 to piRNA biogenesis pathway is controversial. To understand its function it would be important to identify protein interactors and RNA targets of TDRD9. RNA helicases are dynamic enzymes which switch from an off-state to an on-state rapidly and the analysis of their RNA targets and protein interactors is challenging.

4 - Functional characterization of TDRD9 in the piRNA pathway

A point mutation in the DEAD box motif of Vasa (DEAD to DQAD), a DEAD-box RNA helicase involved in secondary piRNA biogenesis, traps Vasa with its *in vivo* RNA targets and transient PIWI-containing biogenesis complex. Characterization of this mutant in insects has helped elucidating the function of Vasa in the piRNA pathway (Xiol et al, 2014 and chapter 3). We designed a similar mutation in the DEVH box of the helicase domain of TDRD9 and generated a conditional knock-in/ knock-out mouse bearing the E257Q mutation (DEVH to DQVH).

The mouse TDRD9 protein is composed of 1383 amino acids that encode for an N-terminal RNA helicase and a C-terminal Tudor domains (Figure 4.1A). Tudor domains bind methylated ligands, which in the case of Tdrd proteins is shown to recognize symmetrically dimethylated arginines (sDMA) present on PIWI proteins (Mathioudakis et al., 2012). The interaction with MIWI2 (Shoji et al., 2009), is likely to involve its Tudor domain. In contrast, the role of the RNA helicase domain of TDRD9 is not currently known.

4.2 Results

4.2.1 Generation and validation of antibodies against mouse TDRD9

During spermatogenesis Tdrd9 is expressed both in pre-natal spermatogonia and in adult spermatocytes. In adult spermatocytes, it is reported to be mainly cytoplasmic, with enrichment in processing bodies in round spermatids (Yabuta et al., 2011). Different studies further reported that, in pre-natal prospermatogonia, TDRD9 localises in perinuclear piP-bodies and in the nucleus (De Fazio et al., 2011; Shoji et al., 2009; Tanaka et al., 2011; Yabuta et al., 2011). At the same developing stage, another work detected TDRD9 only in perinuclear granules but not in the nucleus (Aravin et al., 2009). The localisation of TDRD9 has been tested always with the same antibody batch produced by Chuma and colleagues in 2009 (Shoji et al., 2009). The nuclear localisation of TDRD9 in prospermatogonia fits with its potential role in promoting *de-novo* DNA methylation downstream of MIWI2.

4 - Functional characterization of TDRD9 in the piRNA pathway

To investigate on the localisation of TDRD9 during spermatogenesis, I produced polyclonal antibodies against TDRD9 in rabbit. First I expressed different antigens that map to regions of the protein not conserved among paralogue proteins. Four different antigens (antigens 4 to 7) were expressed in *E. coli* as 6xHis-tagged proteins and purified from inclusion bodies (Figure 4.1A). Purity of the preparations was checked by loading the eluted antigens on SDS-PAGE. Antigen 4 and antigen 7 appeared to be better expressed and were further purified on a Nickel-NTA affinity column. The eluted peptides revealed prominent bands corresponding to the expected size on SDS-PAGE (Figure 4.1B) and were used to raise antibodies against the C-terminal (antigen4) and N-terminal (antigen 7) part of the protein. The C-terminal antigen has the same amino acid sequence (aa 1064-1383) that was used to characterize Tdrd9 KO mouse (Shoji et al., 2009). Once I obtained the two rabbit polyclonal antibodies we tested if they could specifically recognise TDRD9 protein.

By western blot, I could identify a single band at the expected size of 156kDa corresponding to the His-tagged TDRD9 protein transiently expressed in HEK cells. A band of a similar size was also detected in wild-type mouse testis lysate using anti C-ter and anti N-ter TDRD9 antibodies (crude serum) (Figure 4.1C). Interestingly, sometimes the antibody recognizes a second band in testis extract, probably reflecting the presence of a post-translational modification on the protein or a splicing variant (Figure 4.1C).

To test if the antibodies were suitable for IF experiments, I tested the localisation of transiently expressed HA tagged TDRD9 (HA-TDRD9) in HeLa cells. Staining of HA-TDRD9 with anti N-ter TDRD9 antibody gave cytoplasmic specific signal that cannot be observed in untransfected cells (Figure 4.1D). The same localisation is observed with an anti-HA antibody. A weak nuclear signal is observed with anti N-ter TDRD9 staining but this is probably only background, since it is not observed with the more specific anti-HA antibody. These experiments suggest that TDRD9 is a cytoplasmic protein when expressed in a heterologous cell culture system derived from human cells. Accordingly, I could not identify any nuclear localisation signals (NLS) using on-line research software (NLS mapper) on TDRD9 sequence.

4 - Functional characterization of TDRD9 in the piRNA pathway

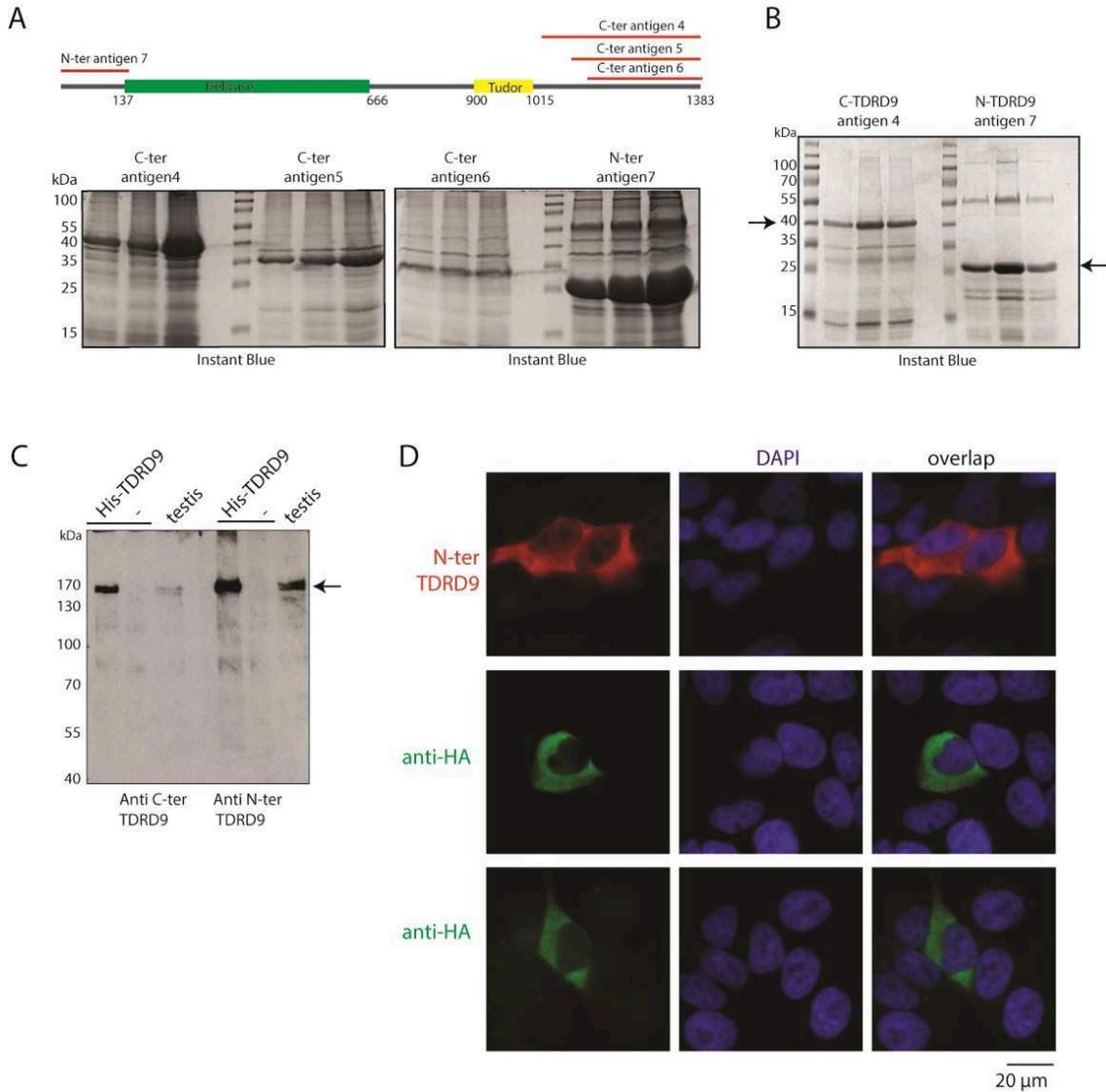


Figure 4.1 Production and validation of antibodies against TDRD9. (A) Antigens from C-terminal and N-terminal end of TDRD9 were purified from inclusion bodies and the purity was checked on a SDS-PAGE. (B) Antigens 4 and 7 were further purified with Nickel-NTA affinity column and purity was checked on SDS-PAGE. Arrows indicate protein bands corresponding to the antigens. (C) C-terminal and N-terminal Ab specifically recognize a band of 156kDa from mouse testis, similar to transfected His-TDRD9 in HEK cells. A double band is often detected from testis lysate (arrow). (D) N-terminal Ab specifically recognizes HA-TDRD9 transfected in HeLa cells.

4.2.2 TDRD9 localisation in mouse testes

In adult mouse testes, the two PIWI proteins that are expressed are MIWI and MILI, whereas MIWI2 expression is restricted to fetal prospermatogonia during

4 - Functional characterization of TDRD9 in the piRNA pathway

epigenetic reprogramming of the germline (Carmell et al., 2007; Deng and Lin, 2002; Kuramochi-Miyagawa et al., 2004) . I tested TDRD9 localisation in WT testis sections from different developmental stages and compared it to the localisation of PIWI proteins.

Spermatogenesis in adult mice is a continuous process where spermatogonial stem cells asynchronously differentiate into meiotic cells, which are committed to become spermatozoa (Figure 4.2B). All the different stages of germ cell maturation can be found in adult testis sections. Every seminal tubule section can contain a different composition of cells with stem cells being adjacent to the interstitial space and the developing spermatocytes and spermatids concentrically disposed in the seminal tubule. TDRD9 was reported to be expressed from spermatogonial stem cells to round spermatids in adult testis (Shoji et al., 2009). In line with what is published I reported TDRD9 localisation in spermatocytes with a very weak staining in round spermatids (Figure 4.2A). I also show that TDRD9 localises in spermatocytes with a weak staining in the nucleus (Figure 4.2C), as reported by Yabuta and colleagues (Yabuta et al., 2011). I checked the localisation of TDRD9 in P0 testes where spermatogonial cells express MILI and MIWI2. Whereas MIWI2 is clearly nuclear, diffused TDRD9 signal is detected only in the cytosol. TDRD9 was reported to be both cytoplasmic and nuclear (Shoji et al., 2009) but in our hands it was repeatedly detected in the cytoplasm, both in P0 (Figure 4.2E) and in E16.5 and E18.5 prenatal testes sections (not shown). Perhaps differences in the staining protocols and in the antibodies used for this study may explain these differences. A cytoplasmic staining of TDRD9 would exclude its involvement in nuclear processes like DNA methylation of transposons promoter.

To exclude the possibility that our antibody was cross-reacting with spermatocyte-specific proteins, I tested it on cryo-sections obtained from adult *Tdrd9*^{-/-} mice. The signal appears to be specific for TDRD9 since only a weak background can be detected in *Tdrd9*^{-/-} testes in IF experiments (Figure 4.2F). I further tested our antibodies in adult testis lysate derived from wild-type animals and from *Tdrd9*^{-/-} animals (Figure 4.2D). A specific band corresponding to the molecular weight of TDRD9 (156 kDa) can be detected in WT mouse testes but not in *Tdrd9*^{-/-} testis lysate.

4 - Functional characterization of TDRD9 in the piRNA pathway

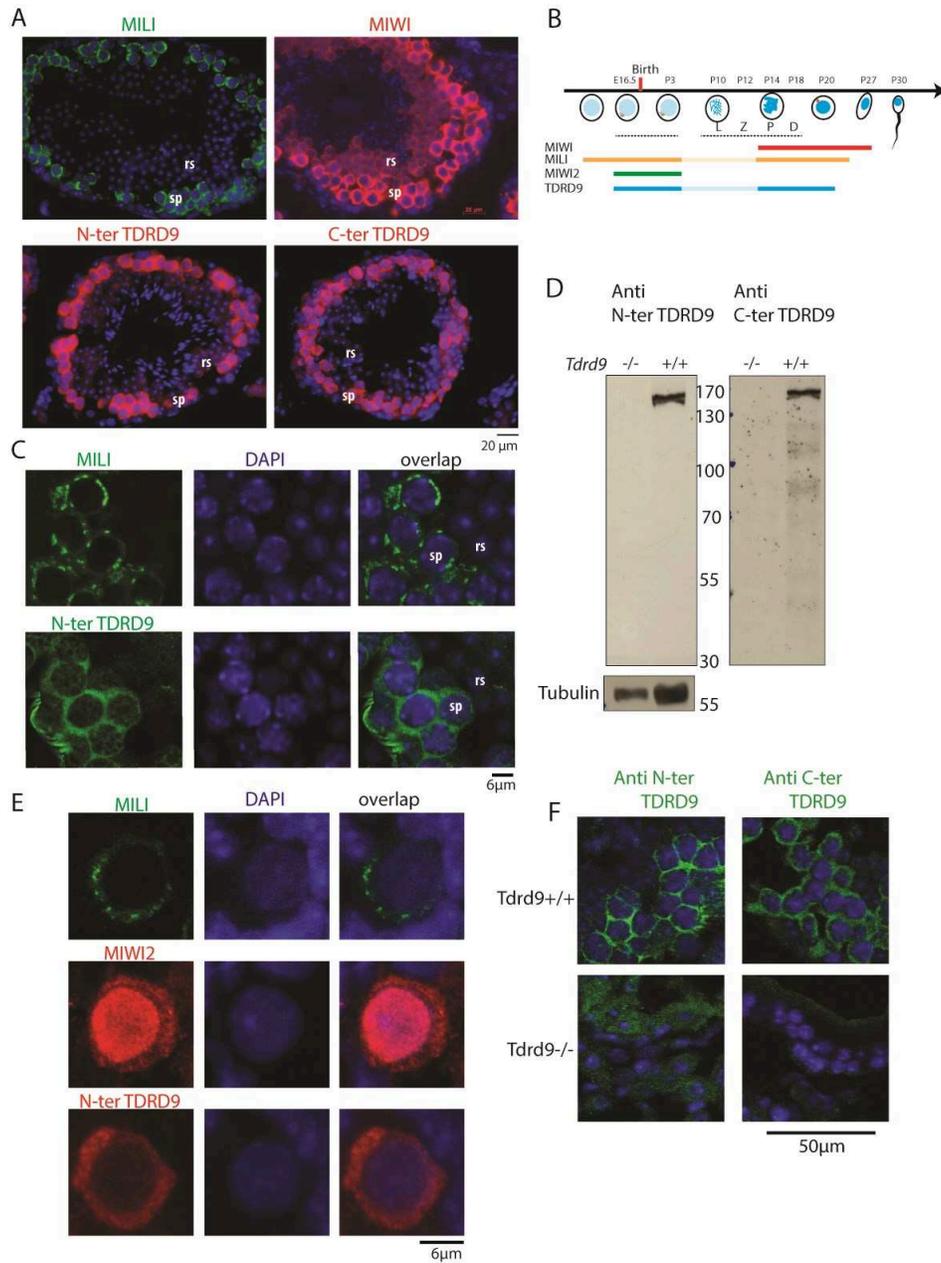


Figure 4.2 Tdrd9 localisation in mouse testes. (A) Adult seminiferous tubules sections stained for the indicated antibodies (sp, spermatocytes; rs round spermatids). (B) schematic representation of the first cycle of spermatogenesis and expression windows of PIWIs and TDRD9 (C) Zoom of TDRD9 localization in adult spermatocytes compared to MILI. (D and F) TDRD9 Ab validation in *Tdrd9*^{-/-} negative control testes by WB and IF. (E) TDRD9 localises in the cytoplasm in P0 mouse testes, contrary to nuclear MIWI2.

4.2.3 Immunoprecipitation of TDRD9 and search for interactors

The role of TDRD9 in transposon repression during the early stages of spermatogenesis has been extensively investigated (Shoji et al., 2009) but its role in pachytene spermatocytes is unexplored. I attempted to investigate the role of TDRD9 in pachytene piRNA biogenesis by immune-purifying TDRD9 from adult testis extract, using the two different purified antibodies (N-ter and C-ter). I probed for the presence of piRNA pathway protein and piRNAs but I could not detect PIWI-proteins MILI and MIWI and piRNAs associated with them (Figure 4.3A).

Next, I tried to find new TDRD9 protein interactors by immunoprecipitating TDRD9 from mouse adult testis using purified TDRD9 antibodies cross-linked to Protein-G sepharose beads. Separation of TDRD9 IP on an SDS-PAGE stained with silver staining shows additional bands beside the TDRD9 protein band identified around 160 kDa, indicating the potential presence of TDRD9 protein partners (Figure 4.3 left). Two biological replicates of N-ter TDRD9 IPs and C-ter TDRD9 IPs were sent for total mass spectrometry analysis, using an IP performed with Protein-G sepharose beads alone as a control (Figure 4.3B). I sorted the list of TDRD9 interactors by considering only proteins with a spectra count ratio (sample/CNTR) higher than 2 and a total number of peptides identified higher than 3. Furthermore, I considered only the proteins that were identified in the two different biological replicates. The only high confident hit was the bait protein, TDRD9, whereas very few proteins were found in both the IPs, none of them being implicated in piRNA biogenesis. None of the interactors identified was present in both C-ter and N-ter IPs, except for Hsp90a (figure 4.2B). Chaperone proteins are often identified in total mass spectrometry analysis due to their high abundance in the cytoplasm, thus we considered it as a contaminant. Overall immunoprecipitation of TDRD9 did not identify any protein interactors. This result is probably a consequence of the biochemical activity of the TDRD9 protein. As already mentioned helicases are dynamic enzymes that consume ATP to unwind RNA molecules. Thus, it is hard to identify protein partners with standard pull down techniques.

TDRD9 often appears as a double band, suggesting the presence of post translational modification. This prompted us to check for the presence of phosphorylation sites by mass spectrometry. We found that the Ser1382 at the very C-terminal end of

4 - Functional characterization of TDRD9 in the piRNA pathway

TDRD9 is phosphorylated (Figure 4.3C). The peptide score of 160.27 is far above the confident threshold of 60 commonly used to validate phosphorylation sites (Marchini et al., 2011). Other phosphorylated sites may be present on the C-terminal end of the protein but they were not detected in this experiment. Further validations are required to understand to which extent the protein is phosphorylated and what is the biological function of such a modification. I could not find any consensus sequences predicted to be bound by specific protein kinases.

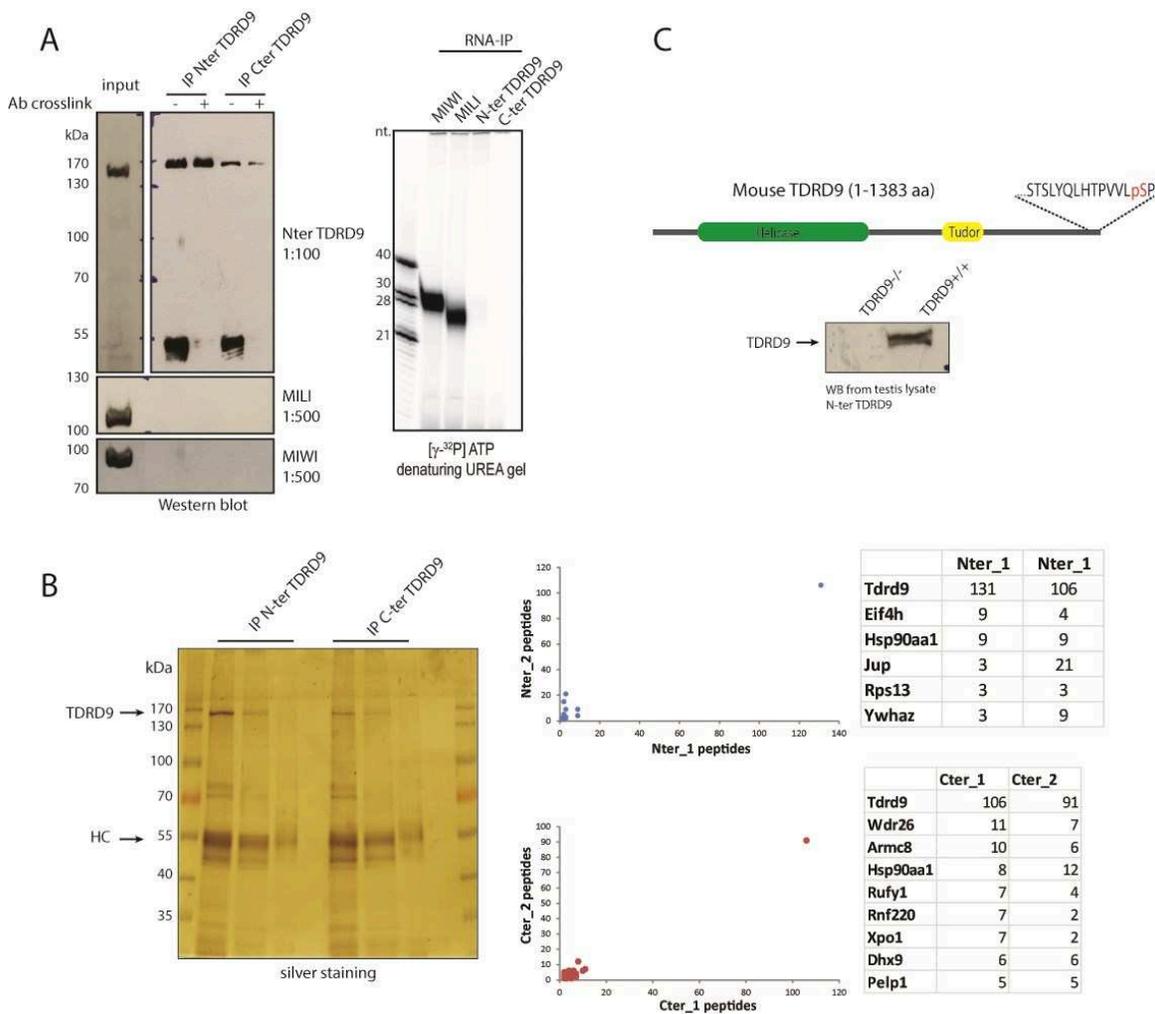


Figure 4.3 IP of TDRD9 from WT adult testis. (A) (on the left) IP of TDRD9 with or without Ab crosslinking did not identify any interaction with MILI and MIWI. (on the right) TDRD9 RNA IP with N-ter and C-ter antibodies: no small RNAs are identified, compared to the 26nt and 29nt piRNA bands observed in MILI and MIWI IP. (B) Total MS analysis of TDRD9 IP (SDS-PAGE on the left, HC heavy chain); peptide numbers of protein interactors were plotted for the two different biological replicates (N-ter_1 vs N-ter_2 and C-ter_1 vs C-ter_2), only TDRD9 is strongly enriched. The other proteins are listed in the tables on the right, only Hsp90 is shared

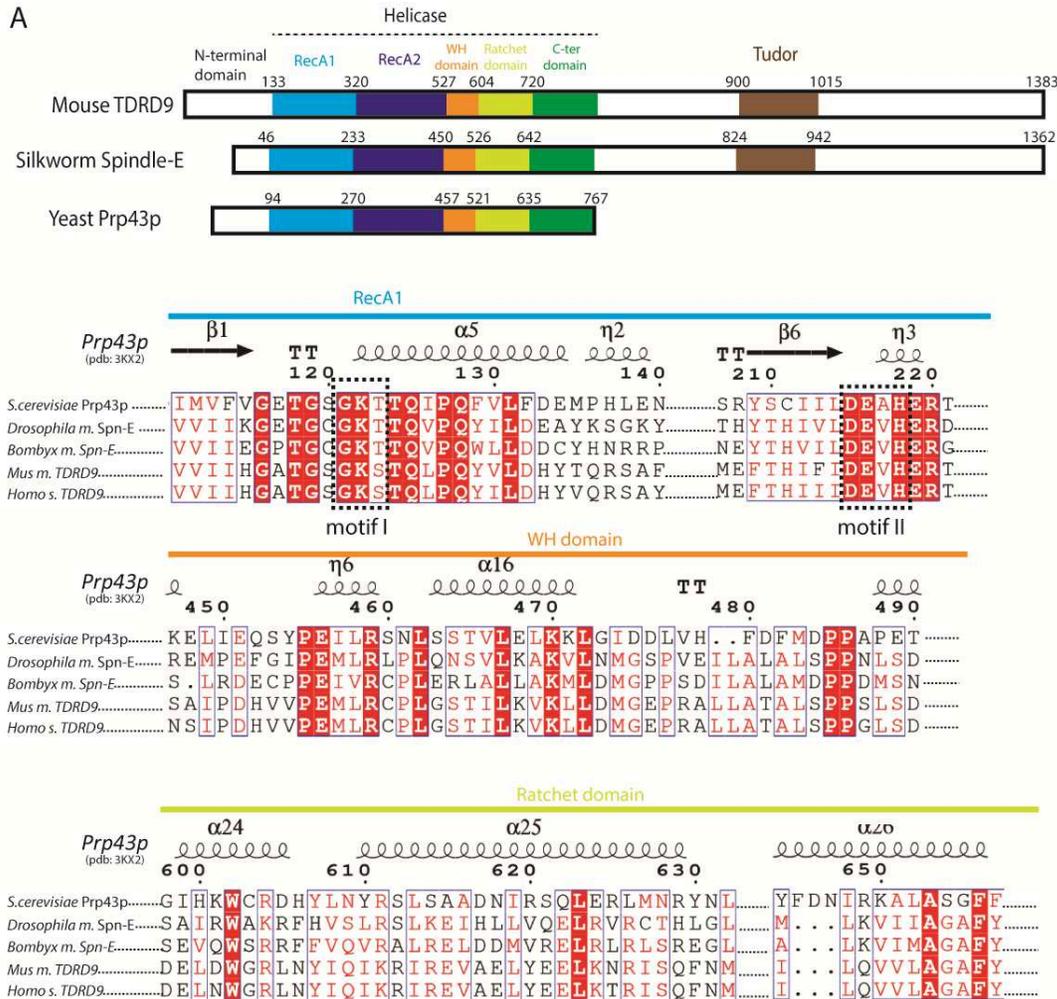
4 - Functional characterization of TDRD9 in the piRNA pathway

between the two IPs (C) Search for phosphorylation sites on TDRD9 identified that Ser1382 on the C-terminal of the protein is phosphorylated. Typical double band in Tdrd9 WB from testis lysate is reported.

4.2.4 Mouse TDRD9 is an active ATPase

In order to assess the biochemical activity of TDRD9 I tried to express and purify recombinant mouse TDRD9 or its orthologue protein Spindle-E from *Bombyx mori*. Sequence analysis of TDRD9 revealed that its helicase domain has a high sequence similarity to Prp43p, a yeast RNA helicase involved in mRNA splicing, whose structure is known (Walbott et al., 2010). I aligned TDRD9 sequences from different eukaryotes with Prp43p. As all members of DExH-box family of RNA helicases, TDRD9 contains the conserved motif I (GKT) implied in ATP binding and motif II (DEVH) responsible for ATP hydrolysis (Ozgun et al., 2015) (Figure 4.4A). The alignment revealed that there are additional domains beside the two RecA domains that appear to be conserved, including the winged-helix motif (WH) and Ratchet domains (Figure 4.4A). Based on this alignment, conservation of residues in different species and secondary structure prediction, I designed different constructs to express the helicase domain of TDRD9 or Spindle-E in bacteria and insect cells (Figure 4.4B). Only two deletion constructs of Spindle-E were efficiently expressed in insect cells but the protein formed soluble aggregates (data not shown). We codon-optimized the helicase domain of mouse TDRD9 for expression in insect cells but this attempt also proved unsuccessful. Curiously, when I finally cloned the full length mouse TDRD9, without codon-optimization, as a Sumo-His-Strep-tagged protein, I managed to obtain a soluble protein.

4 - Functional characterization of TDRD9 in the piRNA pathway



B

protein	boundaries	host	expression	solubility
Bm_Spindle-E	1-747	<i>E.coli</i>	no	-
Bm_Spindle-E	1-847	<i>E.coli</i>	no	-
Bm_Spindle-E	1-677	Sf21	yes	aggregation
Bm_Spindle-E	1-746	Sf21	yes	aggregation
Mm_TDRD9	1-723	Sf21	no	-
Mm_TDRD9	1-725	Sf21	no	-
Mm_TDRD9	Full-Lenght	Hi-5	yes	soluble
Mm_TDRD9	90-1383	Hi-5	no	no

Figure 4.4 Domain analysis of TDRD9. (A) Alignment of TDRD9 from different species with yeast Prp43p identified extensive similarity spanning outside of the two RecA domains. Sequence alignment of motif I and motif II contained in the RecA1 domain is shown (B) List of the constructs tested and relative expression yields.

Mouse full-length (FL) TDRD9 was expressed recombinantly with a 6xHis-Strep-SUMO-tag and purified via two-step affinity chromatography: Nickel followed by Streptactin. Final polishing of the protein prep was achieved by gel filtration (Superose 6

4 - Functional characterization of TDRD9 in the piRNA pathway

10/300 GL) column which gives a single peak at 15mL (Figure 4.5A). The fractions corresponding to the main peak were loaded on an SDS-PAGE, revealing the presence of a single band corresponding to the expected size of His-Sumo-Strept-TDRD9 (173 kDa). The same expression and purification strategy was applied to purify a single point mutant in the ATP hydrolysis catalytic site (DEVH to DQVH), hereafter referred as TDRD9^{DQ}.

To test the ATPase activity of TDRD9, I incubated the wild-type protein with radioactive ATP (³²P- γ ATP) and this resulted in liberation of radioactive free phosphate as observed by thin-layer chromatography (Figure 4.5B). This is specific to the TDRD9 protein, as a point mutation (E257Q) in the catalytic motif (DEVH→DQVH) abolished this activity, indicating that mouse TDRD9 is an active member of the DExH family of ATPases. Interestingly, I did not observe any basal ATPase activity for TDRD9^{DQ}: even in the presence of ATP (³²P- γ ATP) without addition of cold-ATP (low-substrate), the protein was inactive (Figure 4.5B). This is in contrast to what has been observed for Vasa, where under similar conditions (low substrate, Figure 3.1), the Vasa^{DQ} mutant was able to hydrolyse ATP as Vasa wild-type (Figure 3.1). This difference suggests that TDRD9^{DQ} is not a slow-turnover mutant like Vasa^{DQ} but a catalytically-dead enzyme. Note that the point mutation does not affect the overall folding of the protein as the mutant elutes similarly to the wild-type protein on a size-exclusion chromatography (Figure 4.5A).

We wondered if solving the crystal structure of TDRD9 may help in understanding its molecular function. First, I tested the thermal stability of the protein by using a fluorescence-based thermal shift assay. The melting temperature of a purified protein was shown to be a good predictor of its likelihood to form crystals: proteins with a T_m above 44°C have statistically a 50% chance to crystallize (Dupeux et al., 2011). TDRD9 has a T_m of 47°C and since it elutes as a monodisperse protein in gel filtration with a rather high purity grade (estimation from the gel in Figure 4.5A) it is a suitable substrate for crystallography. Nevertheless, attempts to concentrate the protein above 1.3-1.5 mg/ml resulted in protein aggregation and precipitation. To improve protein solubility at higher concentrations I tried hypotonic solutions (dialysis in 30% PEG 33K) or different additives but none of these approaches was successful. I performed a high-

4 - Functional characterization of TDRD9 in the piRNA pathway

throughput crystallization test at a protein concentration of 1.3mg/ml but did not obtain any crystals.

We thought that perhaps the protein consists of unstructured regions at its N-terminal or C-terminal ends, which affect solubility and promote aggregation above a certain protein concentration. Thus, I performed a limited proteolysis assay, where the protein was subjected to chymotrypsin digestion and samples were collected at different time points. Protease-digested polypeptides were separated on an SDS-PAGE and analysed. Over time, I could observe the disappearance of the FL input protein band and the appearance of other discrete bands (mainly bands 1, 2 and 3) (Figure 4.5E). These bands were excised from the gel and sent for mass spectrometry analysis to obtain the boundaries (Figure 4.5E). Identification of the N-terminal and C-terminal regions of the trimmed polypeptides reveal core structured subunits of the protein that may have higher chances to crystallise. Interestingly, we observed that the N-terminal region is rapidly digested by chymotrypsin (band 1 fig.4.5E) whereas the C-terminal part remains intact until the end of the protease treatment (30min). The RecA1 domain is also rapidly separated from the rest of the protein (band 2), suggesting that the linker between the two RecA domains is exposed to the solvent and can be cleaved by the protease. This result suggests that the helicase domain of TDRD9 is not an autonomously folding unit, but an integral part of the entire protein with the entire C-terminal end forming a compact unit with the RecA2. This may explain why the constructs that I tested which contained the helicase domain alone, were hardly expressed: perhaps the resulting proteins expose parts to the solvent that are usually buried in the core. I re-cloned TDRD9 with the first 90 amino acids deleted (band 1) but unfortunately failed to obtain a more stable protein (data not shown).

4 - Functional characterization of TDRD9 in the piRNA pathway

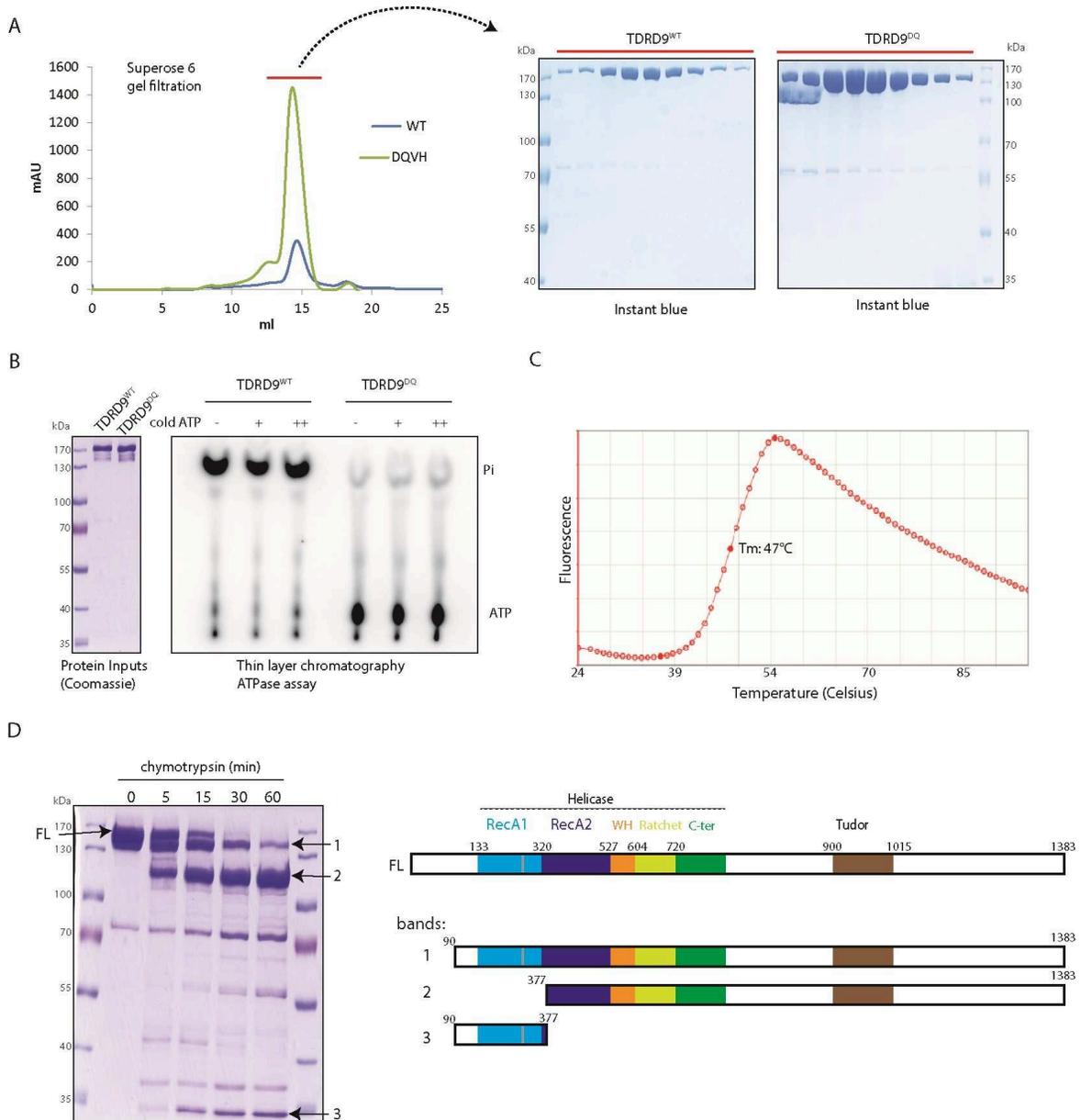


Figure 4.5 Purification and in vitro activity of TDRD9. (A) The final gel-filtration step (Superose 6 10/300) of TDRD9^{WT} TDRD9^{DQ}; 9 fractions of the collected peak were loaded on a gel and stained with Coomassie (on the right) (B) ATPase assay of TDRD9 and with the relative protein inputs used for the experiment. Non-radioactive ATP (cold) was added in increasing concentration to the [γ -³²P]ATP (C) Thermal shift assay on TDRD9^{DQ} purified protein. (E) Limited proteolysis of TDRD9^{DQ}; with chymotrypsin digestion. The resulting bands were analysed by mass spectrometry in order to identify the boundaries.

4.2.5 Loss of TDRD9 ATPase activity results in male-specific infertility in mice

To examine whether the ATPase activity of TDRD9 is critical for its function *in vivo*, we targeted the genome of mouse embryonic stem (ES) cells to introduce a single nucleotide mutation in the exon 5 of *Tdrd9* (this knock-in allele will be hereafter referred to as *Tdrd9^{KI}*). This generates a single amino acid mutation (E257Q) in the ATPase domain of the TDRD9 protein (DEVH to DQVH), resulting in a catalytically-dead protein (Figure 4.6A). Mutant mice were generated from one of the targeted ES cell lines (see Materials and Methods). The gene targeting strategy also introduced *loxP* sites flanking exons 3-5, allowing the generation of the knock-out allele when crossed with mice expressing the Cre recombinase (Figure 4.6A).

Heterozygous or homozygous *Tdrd9* knock-out or knock-in mice of both sexes are viable. Furthermore, females of all genotypes are fertile. However, as previously described (Shoji et al., 2009), animals of the *Tdrd9^{-/-}* genotype display male-specific infertility. Interestingly, *Tdrd9^{KI/KI}* males are also infertile. Visual inspection of the *Tdrd9^{KI/KI}* mutant testes reveals strong reduction in size, similar to that from *Tdrd9^{-/-}* mice (Figure 4.6B). Histology of adult (P60; 60 days old) testes sections from the *Tdrd9^{KI/KI}* mutant show the presence of narrow seminiferous tubules that appear devoid of late-stage germ cells (Figure 4.6C). This is in contrast to the larger seminiferous tubules in the *Tdrd9^{+KI}* control samples that are full of germ cells at different stages of development, including late-stage elongated spermatids that mature into sperm (Figure 4.6C). These differences are barely visible in P26 post-natal testes sections (Figure 4.6D) even if the overall seminiferous tubules size is reduced. During prophase I of meiosis, homologous chromosomes pair and undergo synapsis that allows DNA recombination between the non-sister chromatids. This recombination event is initiated by introduction of double-stranded breaks (DSBs) that are marked by the phosphorylated form (γ -H2AX) of the histone variant H2AX. During synapsis, the unpaired sex chromosomes form a specialized structure called the XY body or sex body that is also decorated by γ -H2AX. When compared to the control, immunofluorescence analysis of the *Tdrd9^{KI/KI}* mutant spermatocytes reveals an increased signal for γ -H2AX, indicating a failure to complete the chromosome crossover event (Figure 4.6E). Also, while the XY body is clearly visible in the control germ cells, spermatocytes in the *Tdrd9^{KI/KI}* mutant evidently lack

4 - Functional characterization of TDRD9 in the piRNA pathway

such structures. This indicates pre-pachytene arrest in germ cell development, prior to completion of the synapsis process in the *Tdrd9*^{KI/KI} mutant. These data indicate that the ATPase activity of TDRD9 is essential for proper progression of spermatogenesis in mice.

4 - Functional characterization of TDRD9 in the piRNA pathway

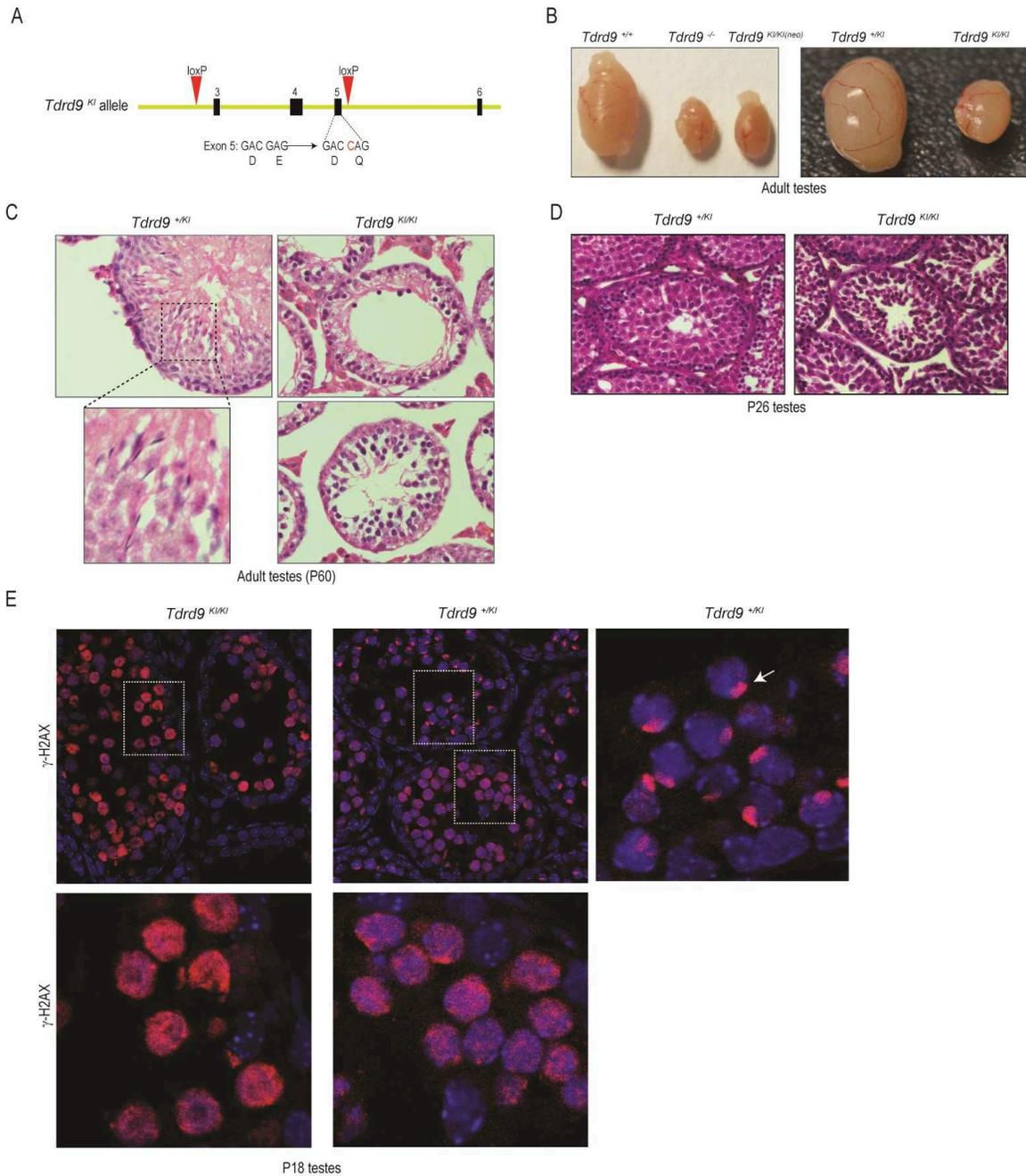


Figure 4.6 Phenotype of *Tdrd9* knock-in mouse mutant. (A) Schematic strategy used to generate the *Tdrd9*^{KI} allele, a point mutation at the end of exon 5 gives the DEVH to DQVH mutation and the *LoxP* sites allows generation of *Tdrd9*^o allele (see materials and methods). (B) Picture of mouse adult testes from the indicated phenotypes. (C-D) Hematoxylin/eosin staining of testes sections from P26 and P60 mouse mutants, differences between knock-in and het control are barely visible at P26 but at P60 *Tdrd9*^{KI/KI} testes are devoid of late-stage germ cells. (E) γ -H2Ax staining on P18 testes section. The characteristic sex body identified in pachytene cells is not detected in the mutant, indicating a pre-pachytene cell arrest.

4.2.6 *Tdrd9* catalytic-dead mutant mice fails to suppress LINE1 retrotransposons

Retrotransposons like IAP and LINE1 elements are under the control of the piRNA pathway in mice (Chuma and Nakano, 2013). Examination of their levels by Northern blotting revealed de-repression of the LINE1 transcripts in the adult (P60) *Tdrd9*^{KI/KI} mutants (Figure 4.7A). This is similar to the observations in our knock-out line (Figure 4.7A), and consistent with what has been previously reported for the *Tdrd9* knock-out mutant generated by Chuma and colleagues (Shoji et al., 2009). The de-repression of LINE1 transcripts can also be detected in younger (P18) *Tdrd9*^{KI/KI} mutants. Notably, IAP levels are not elevated in both the *Tdrd9* knock-in and knock-out animals. These LINE1 transcripts constitute functional copies of the retrotransposons, as a translation product of the ORF1 (an RNA-binding protein) can be detected by Western analysis in *Tdrd9*^{KI/KI} mutant testes lysates (Figure 4.7B). Similar expression of LINE-1 (L1)ORF1p is also seen in the *Tdrd9*^{-/-} mouse (Figure 4.8A), as previously reported (Shoji et al., 2009). The produced L1ORF1p protein accumulates in the cytoplasm of *Tdrd9*^{KI/KI} mutant germ cells (Figure 4.7D), where it likely forms large ribonucleoprotein (RNP) complexes with L1 RNA, required for retrotransposition (Goodier and Kazazian, 2008). It has been reported that in *Tdrd9*^{-/-} DNA *de-novo* methylation of LINE1 promoter in prospermatogonia is affected (Shoji et al., 2009). To investigate if the same defects could be detected in our KI mouse mutant I performed methylation sensitive southern blot from genomic DNA extracted from P14 and adult mouse mutants. The extracted DNA was digested with two different enzymes (MspI and HpaII) that recognise the same DNA sequence (CCGG) but have different sensitivity to CG methylation: MspI is insensitive to the modification whereas HpaII is sensitive, meaning that it cannot cut if the sequence is methylated. Accordingly, after probing the digested DNA with a probe against the 5'UTR sequence of LINE1(L1 MD-A2), wt samples and *Tdrd9*^{+KI} samples show a cleavage band only in the lane corresponding to the DNA digested with MspI but not in the lane of HpaII (Figure 4.7C). On the contrary, DNA extracted from *Tdrd9*^{KI/KI} and *Tdrd9*^{-/-} testes show presence of a band in both digestion condition (Figure 4.7C, red arrows), indicating that LINE1 promoter are not methylated in the knock-in mutant .

Taken together, these results demonstrate the importance of the ATPase activity of *Tdrd9* in active suppression of LINE1 elements in the mouse male germline. Failure to

4 - Functional characterization of TDRD9 in the piRNA pathway

establish proper methylation on their genomic loci leads to arrested germ cell development and infertility.

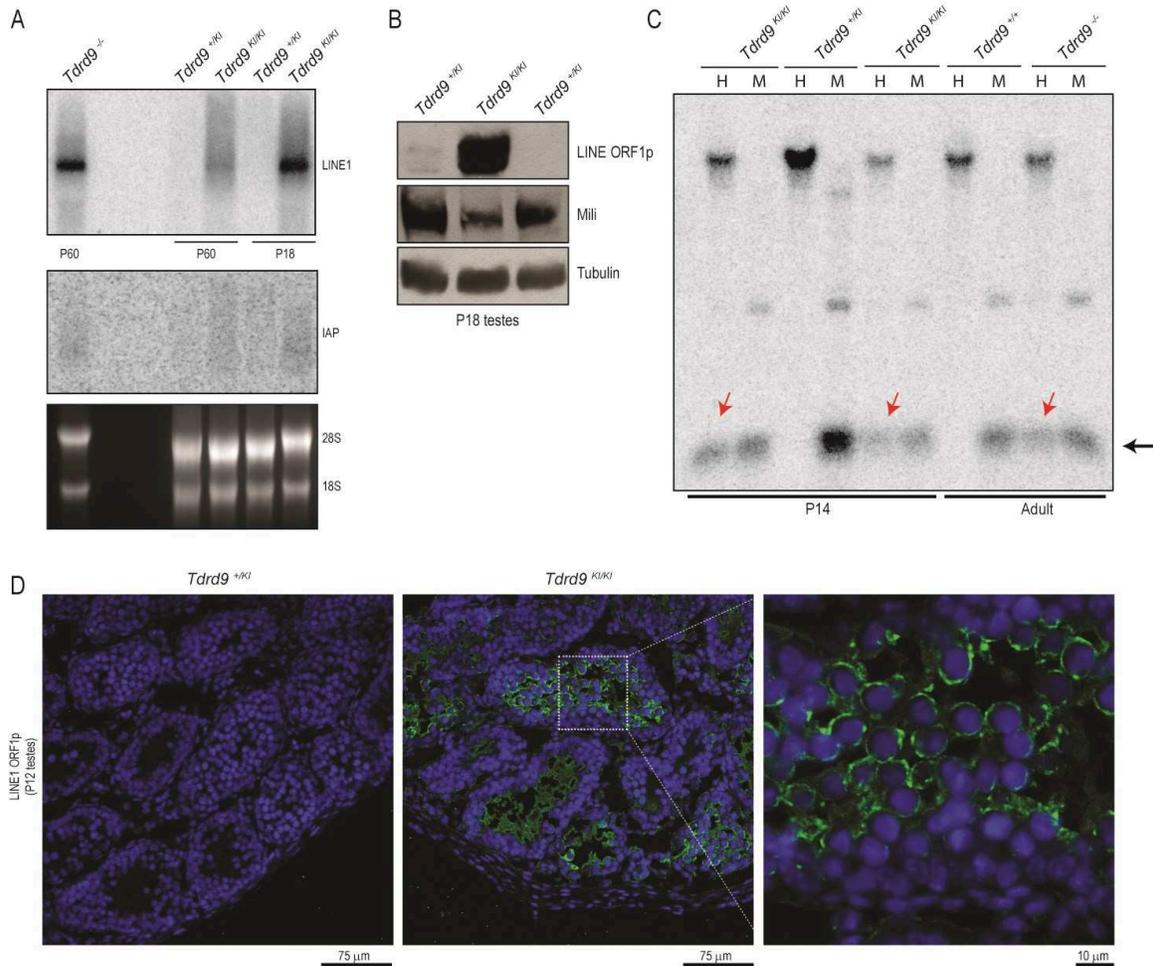


Figure 4.7 LINE1 de-repression in mutant mouse testes. (A) Northern blot analysis identifies LINE1 transcript in mutant testes, IAP is not upregulated. (B) LINE-Orf1 is strongly expressed in mutant testes as identified by WB. (C) Methylation-sensitive Southern blot analysis of testis LINE1 genomic DNA. Genomic DNA was digested with methylation sensitive (HpaII, H) and methylation insensitive (MspI, M) enzymes. Red arrows indicate the position of methylation-sensitive restriction products exclusively detected in both knock-in and knock-out mutants. (D) Identification by IF of LINE1-Orf1 protein in P12 testes sections.

4.2.7 TDRD9 mutant protein is not detectable in adult mouse testes.

As shown in chapter 3, the failure of Vasa^{DQ} mutant to release its ATP hydrolysis products, blocks the protein in a stalled RNP complex. We wondered if the same phenotype could be observed for TDRD9^{DQ}. Unfortunately, the mutant protein was not

4 - Functional characterization of TDRD9 in the piRNA pathway

detectable in adult mouse testes, as indicated by western blot analysis from testis lysate, precluding us to perform any biochemical studies on TDRD9^{DQ} *in vivo* (Figure 4.8A). I noted that expression of MILI is also strongly reduced in mutant adult testes, in both the knock-in and knock-out mutants. This is probably reflecting the massive loss of germ cells observed by Haematoxylin/Eosin staining (Figure 4.8A). Thus, I checked the expression of TDRD9 at earlier stages, but even in P18 testes the TDRD9 mutant was not detected (figure 4.8B). We wondered if the mutation in the catalytic site of TDRD9 (DEVH to DQVH) affects protein stability. This appears unlikely however, as *in vitro* TDRD9^{DQ} shows the same solubility as its wild-type counterpart (Figure 4.5C). Furthermore, a similar mutation in the active site of *Drosophila* Spindle-E (DEVH to DKVH) was shown to be stably expressed in fly ovaries (Ott et al., 2014) This leads to hypothesis that the block in pre-pachytene stage caused by the expression of catalytic-dead TDRD9 in prospermatogonia precludes the formation of those germ cells that would express TDRD9 in later stages. Accordingly, I could successfully detect TDRD9^{DQ} protein by IF in E16.5 mutant embryonic testes (Figure 4.8C). Interestingly, TDRD9 appears to be enriched in perinuclear foci in both *Tdrd9*^{+/+} and *Tdrd9*^{+/KI} testes. This signal is instead lost in the *Tdrd9*^{KI/KI} testes, where only a diffuse cytoplasmic staining could be observed. Nevertheless, additional experiments are required to confirm this observation.

4 - Functional characterization of TDRD9 in the piRNA pathway

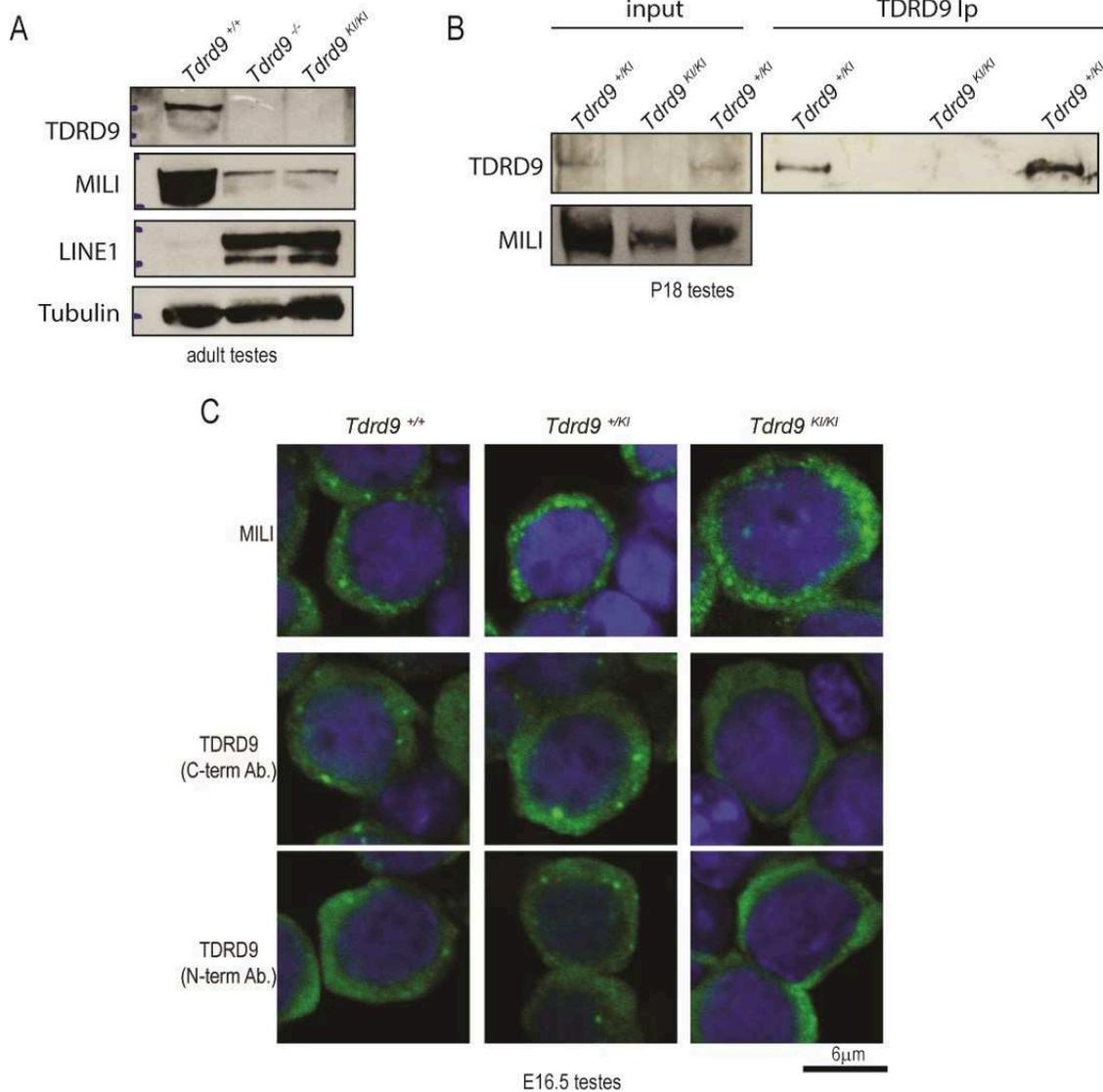


Figure 4.8 Detection of TDRD9^{DQ} in mouse mutant testes. (A) WB on adult mouse testis of the indicated phenotype. Reduction of MILI is a consequence of germ cell loss. TDRD9 protein is not detected, LINE1-Orf1p is overexpressed, Tubulin is used as loading control (B) WB on TDRD9 IP from P18 testes from different phenotypes, TDRD9 is not detected in knock-in mutant testes. MILI is used as loading control. (C) IF of TDRD9 in prenatal testes (E16.5), signal for TDRD9 is comparable to the wt control, but enriched in foci.

4.2.8 Impact on piRNA biogenesis in the *Tdrd9* catalytic-dead mutant mice

Next, we wanted to examine whether the loss of catalytic activity of TDRD9 has any impact on piRNA levels. This part of the project was collaboration with Raman Pandey and David Homolka, two post-docs in my lab. Given the early arrest of

4 - Functional characterization of TDRD9 in the piRNA pathway

spermatogenesis in the mutants, we examined the perinatal germline of new-born pups (P0). PIWI-proteins MILI and MIWI2 were immunoprecipitated from testis lysates and their associated piRNAs were revealed by 5'-end labelling (Figure 4.9A). As expected, RNAs associated with MIWI2 are slightly longer than those found in MILI, with a general size range of 26-28 nt (Aravin et al., 2008). When compared to the control (*Tdrd9^{+KI}*), I did not observe any dramatic difference in the abundance of piRNAs in the knock-in or knock-out mutants. To gain better understanding of the sequence composition of the piRNAs associated with MILI and MIWI2, small RNAs were sequenced; typically yielding 15-21 million small RNA reads per library. Within each library 29-40% of the reads could be mapped uniquely on the mouse genome (see materials and methods, [Chapter 2.3.6](#)). For further analysis only perfect matches reads of a defined length of 24-30nt were considered.

For all the three genotypes analysed (*Tdrd9^{+KI}*, *Tdrd9^{KI/KI}* and *Tdrd9^{-/-}*), the Gaussian distribution of MILI-loaded piRNAs peaks at 26nt whereas MIWI2 piRNAs are 28nt long on average. These small RNAs are bona fide piRNAs since in MILI libraries 86 to 87% of the associated sequences start with uridine (1-U bias) (Figure 4.9B). A cleavage product of a piRNA with a 1U bias can generate a secondary piRNA that is loaded into MIWI2 or MILI (Aravin et al., 2008; De Fazio et al., 2011). As a result of the 10nt overlap with the primary piRNA, the secondary piRNA has a bias for Adenosine at position 10 from its 5' end (10A bias). In the control library (*Tdrd9^{+KI}*) 23% of MILI piRNAs and 40% of Miwi2 piRNAs have a 10A bias, suggesting that they are generated by secondary biogenesis. This signature is also maintained in the libraries generated from mutant mice (*Tdrd9^{KI/KI}* and *Tdrd9^{-/-}*), where similar percentage of 10A bias are observed. This suggests that TDRD9 does not affect secondary piRNA biogenesis (Figure 4.9B).

Next, the distribution among different categories of sequences was analysed (Figure 4.9C). As expected, in the control library obtained from *Tdrd9^{+KI}* testes, piRNAs loaded into MILI are enriched in sense and antisense sequences derived from repeats, whereas MIWI2 piRNAs are mostly enriched in repeats derived from antisense-strands, reflecting their prevalent loading via secondary pathway initiated by sense-oriented primary MILI-piRNAs. Compared to the control, in the knock-in and knock-out mutants,

4 - Functional characterization of TDRD9 in the piRNA pathway

MILI and MIWI2 are both enriched in sense-oriented repeats piRNAs, whereas MILI is also enriched in anti-sense derived piRNAs (Figure 4.9C). This increase is compensated by a drop mainly in piRNAs derived from unannotated sequences (indicated as 'none' in Figure 4.9C).

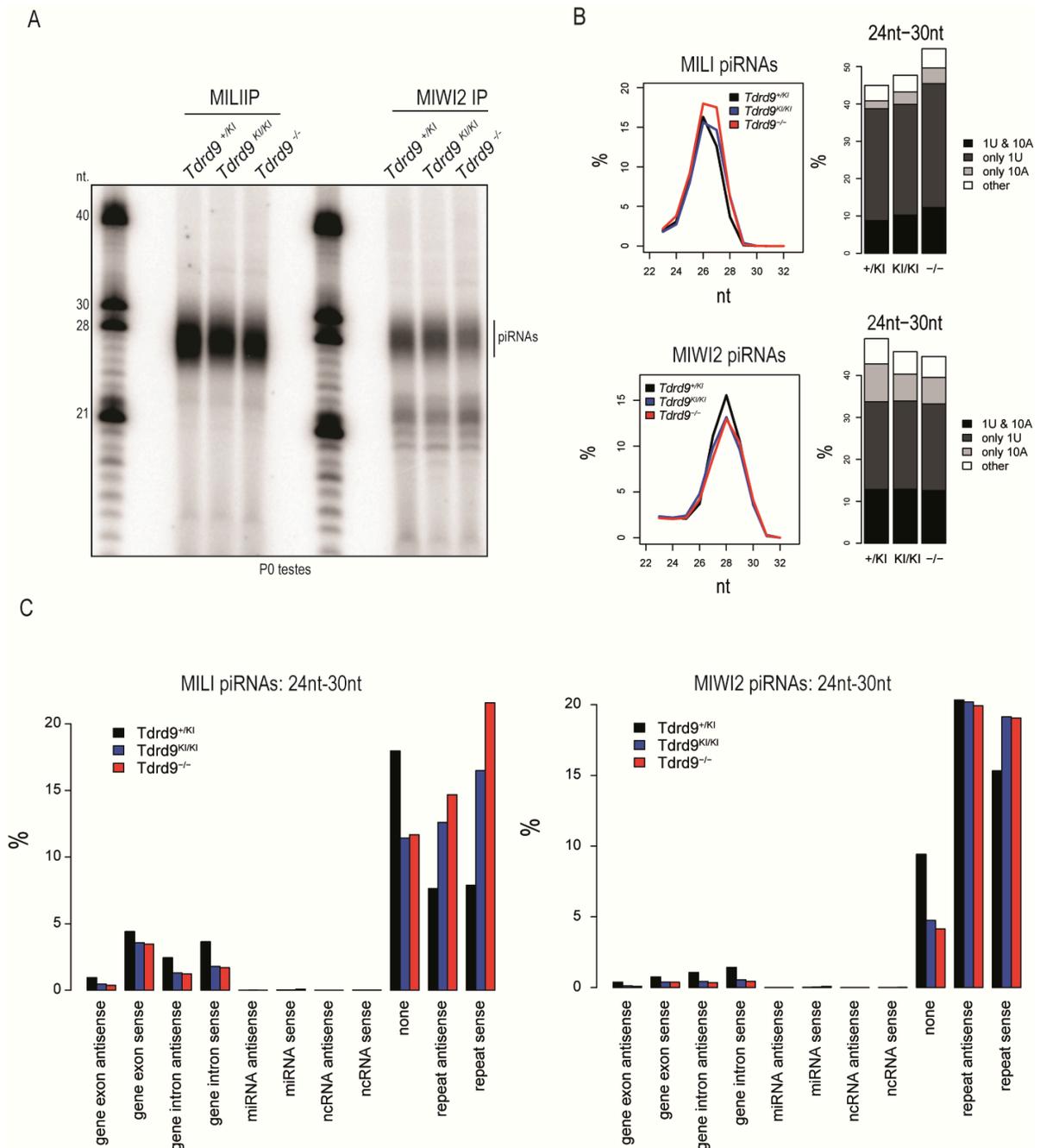


Figure 4.9 MILI and MIWI2 are loaded with piRNA in *Tdrd9* mutants. **(A)** Denaturing UREA PAGE of radiolabelled small RNAs extracted from MILI and MIWI2 IP from P0 *Tdrd9*^{+KI}, *Tdrd9*^{KI/KI} and *Tdrd9*^{-/-} testes **(B)** Length distribution and nucleotide bias of piRNAs plotted as abundance of unique reads. **(C)** Class distribution of reads in MILI- and MIWI2-bound 24-30nt

4 - Functional characterization of TDRD9 in the piRNA pathway

small RNAs. Values are in number of reads as percentage of the total number of reads in each library.

Analysis of the nature of these repeats-derived piRNAs revealed that both MILI and MIWI2 are enriched in sequences derived from LINE elements. Particularly MILI is enriched with piRNA derived from young LINE1 (L1) retrotransposons like L1-mdA, L1-mdT, L1-mdF2 (red dots in Figure 4.10A). LTR elements are also slightly up-regulated but mostly on the anti-sense strand bias (green dots). MIWI2 piRNAs are slightly enriched in transposon sequences derived from LINES and LTR elements, which can be an indirect effect of the increased loading of these elements into MILI (Figure 4.10A).

We showed before that L1 promoted are de-methylated in *Tdrd9* mutants (Figure 4.7). Thus we wondered if L1-derived piRNAs had the ping-pong signature which is typical of secondary biogenesis. We mapped the 5'-to-5' distance ("Ping-Pong" signature) of piRNAs assigned to MILI (sense strand) with piRNAs loaded into MIWI2 (antisense strand) (Figure 4.10C). A discrete peak at 9 nucleotides for all the three genotypes indicates that MILI and MIWI2 are engaged in ping-pong amplification. This reveals that secondary biogenesis of L1-derived piRNAs is correctly performed, as already suggested by the presence of the 10A bias in MIWI2 piRNAs (Figure 4.9B).

Curiously, we observed a strong decrease of SINE elements-derived piRNAs in MIWI2 (Figure 4.10D). The drop was more severe for anti-sense derived piRNAs and was observed also for MILI but with a minor extent (Figure 4.10D). This drop was even more striking when single subfamilies of SINE elements were analysed, as observed when we mapped piRNAs on a consensus sequence for B1Mus1 SINE (Figure 4.10E). The reduction is specific for MIWI2-loaded piRNAs as MILI-piRNAs map to the consensus at similar levels in the control and in the mutant.

4 - Functional characterization of TDRD9 in the piRNA pathway

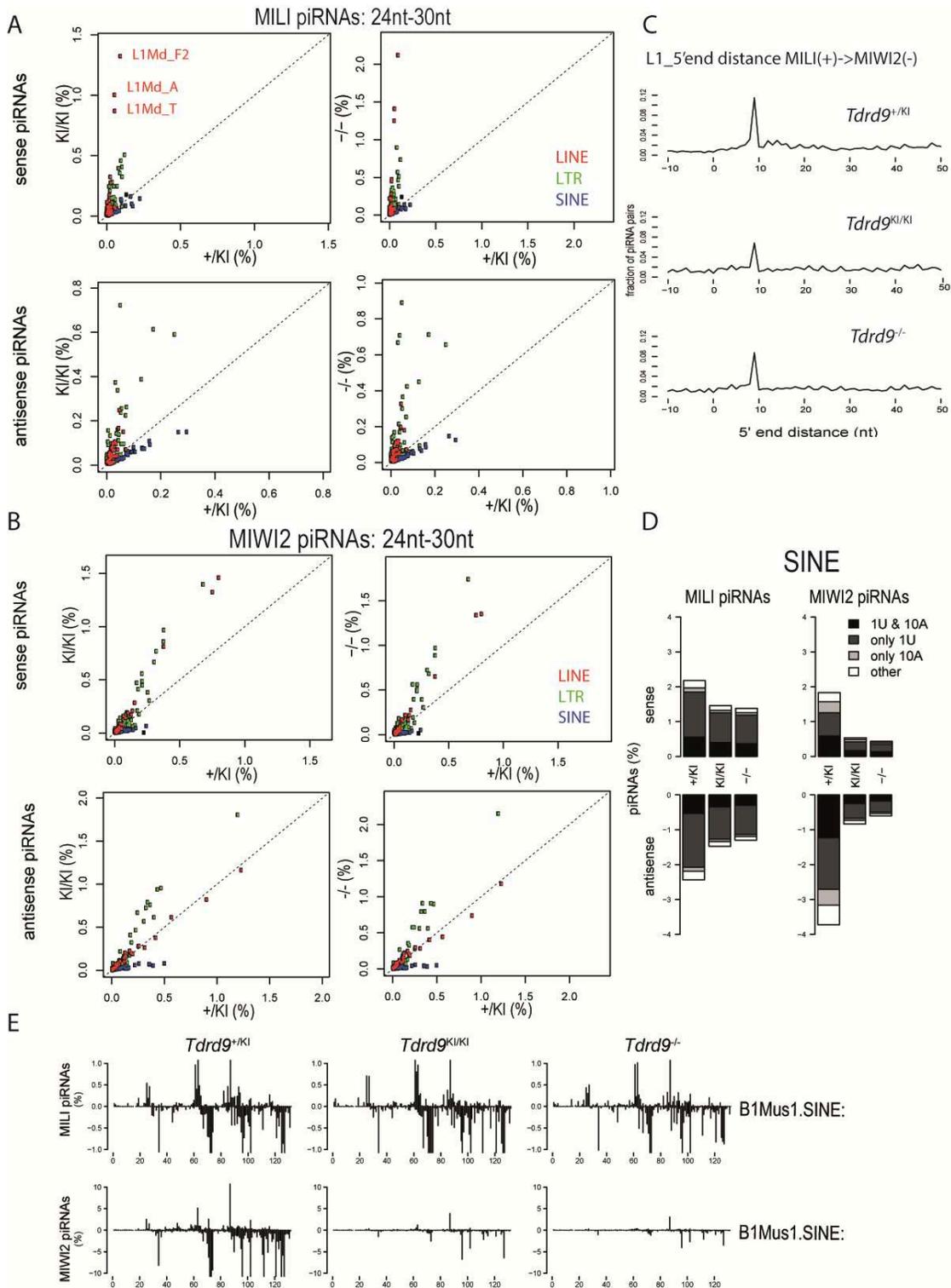


Figure 4.10 analysis of transposon-derived piRNAs. (A) Mapping of MILI- and (B) MIWI2-piRNAs to different transposon families. piRNA reads mapping on different transposon families (LINE, LTR and SINE) from *Tdrd9*^{-/-} and *Tdrd9*^{KI/KI} are plotted against the control reads (*Tdrd9*^{+/KI}) Values are in % of total library. (C) Ping-pong signature on L1 consensus: .the

4 - Functional characterization of TDRD9 in the piRNA pathway

distance between the 5' end of MILI sense piRNAs and the 5' end of MIWI2 antisense piRNAs is calculated. **(D)** SINE elements-derived piRNAs in MIWI2 IPs are strongly reduced for *Tdrd9*^{KI/KI} and *Tdrd*^{-/-} compared to the control *Tdrd9*^{+K1}. **(E)** Mapping of piRNA reads on B1Mus1 SINE element consensus.

4.3 Discussions and perspectives

In the germ line, piRNAs and PIWIs preserve the integrity of the genome against the uncontrolled activation of transposons. Since it was first identified in *Drosophila*, Spindle-E has been an enigmatic protein with an evident function in transposon control. Mutations in *spnE* gene cause transposons activation, female sterility and a drop in germ cells piRNA population, suggesting a role in piRNA biogenesis (Malone et al., 2009; Ott et al., 2014).

A similar role in piRNA biogenesis was also proposed for the orthologue mouse protein TDRD9, the loss of which phenotypically resembles the knock-out of *Miwi2*, *Vasa* and *Tdrd12*, in terms of block in spermatogenesis and transposon activation (Carmell et al., 2007; Kuramochi-Miyagawa et al., 2010; Pandey et al., 2013; Shoji et al., 2009). TDRD9 was proposed to act downstream of MIWI2, in establishing *de-novo* DNA methylation at transposon genomic loci because MIWI2 is correctly loaded with piRNAs in *Tdrd9* deficient mice. Furthermore, TDRD9 was reported to partially localise in the nucleus of pre-natal prospermatogonia, however, another report suggests it to be cytoplasmic (Aravin et al., 2009; Shoji et al., 2009).

To clarify its localisation during spermatogenesis, I produced two antibodies against TDRD9. In pre-natal and peri-natal testes sections, I repeatedly observed a clear cytoplasmic staining (Figure 4.2E, 4.8C) that contrasts with the previous nuclear localisation reported by other groups (De Fazio et al., 2011; Shoji et al., 2009). Our data therefore exclude the possible role of TDRD9 in transposon DNA methylation, at least in the developmental stages analysed in this study.

I expressed TDRD9 *in vitro* and showed that it is an active ATPase. We designed a point-mutation that abolishes such activity (DEVH to DQVH) and generated a knock-in mouse that expresses the same mutant protein *in vivo*, in the attempt to lock the protein in a functional transient state. Resulting knock-in males are infertile with a block in spermatogenesis at pre-pachytene stage of meiosis, proving that the ATPase activity of

4 - Functional characterization of TDRD9 in the piRNA pathway

TDRD9 is essential for its function. The block is caused most probably by overexpression of transposon elements like LINE-1, which fails to be methylated during *de-novo* DNA methylation. Unfortunately, the severity of the phenotype prevented us to perform biochemical experiments on the mutant protein, since TDRD9^{DQ} is detected only in pre-natal testis, a stage where very few germ cells are expressed.

We deep-sequenced piRNAs bound to MILI and MIWI2 in P0 testes, collected from knock-in and knock-out *Tdrd9* mutants. We observed that prenatal PIWI-proteins are properly loaded with piRNAs, clearly indicating that TDRD9 does not control piRNA biogenesis or loading into PIWIs, as it has already been proposed (Shoji et al., 2009). Differences in sequence composition of repeat-associated piRNAs were similarly observed in the absence of the protein (*Tdrd9*^{-/-}) or in the presence of the mutated protein (*Tdrd9*^{KI/KI}). MILI piRNAs are strongly enriched in sequences derived from LTR and non-LTR transposons like LINE-1 and IAP. Surprisingly, we observed a severe drop in MIWI2-specific piRNAs derived from SINE elements but the biological meaning of this observation needs to be further investigated.

Overall we demonstrated that the ATPase activity of TDRD9 is essential for its function in spermatogenesis, however, the precise role of TDRD9 helicase activity in transposon silencing remains to be investigated. Based on our data, we can propose two potential functions for TDRD9.

Since *Tdrd9* mutation neither affects piRNA-loading of PIWI proteins, nor their localisation in different cellular compartments, it is possible that TDRD9 affects transposon regulation at a different level, perhaps participating in a different parallel non-redundant transposon silencing pathway. The reduction in SINE elements-derived piRNAs could be an indirect consequence of *Tdrd9* mutation. The lack of its activity would generate overexpression of LTR and non-LTR transposons with a consequent enrichment of transposon transcripts in the cytoplasm. This would result in a strong response from the piRNA pathway with a bust in Line and IAP-derived piRNA population. Within MIWI2 piRNAs, an increase in this class of repeats could result in an indirect drop of SINE-element derived piRNAs.

Another possibility is that TDRD9 functions in licensing primary piRNA precursors to enter the piRNA pathway, sorting different species of transposons, and

4 - Functional characterization of TDRD9 in the piRNA pathway

loading them into MIWI2. The lack of its catalytic activity does not block the loading capacity of the primary machinery but dramatically changes the source of transcripts to be loaded. Consequently, SINE-elements derived transcripts are not loaded correctly into MIWI2. This could cause a variation in DNA methylation patterns on certain genomic loci, which has deleterious effects for the cell. SINE-elements derived piRNAs may be necessary to drive MIWI2 to silence specific genes required for a correct meiosis. Such a hypothesis would fit with previous studies where TDRD9 was found to interact with MIWI2 but not with MILI in piP-bodies (Aravin et al., 2009; Shoji et al., 2009). This would also be in line with biochemical studies performed on the orthologue protein Spindle-E from *Bombyx mori*. For instance, Siomi and co-workers proposed that Spindle-E mediates the loading of certain piRNAs into Siwi, and that its catalytic-dead mutant fails in doing so, as detected by Northern blot on Siwi-bound small RNAs (Nishida et al., 2015).

Nevertheless, a consistent discrepancy between the insect and mammalian system remains: deletion of Spindle-E generates a complete loss of piRNA population in *Drosophila* whereas deletion of *Tdrd9* in mouse leaves MILI and MIWI2 piRNA almost unchanged in prenatal testes.

Further experiments are needed to understand which of these two models proposed is the most consistent. We are currently performing total RNA-seq analysis from P0 mutant testes to check if the decrease in piRNAs from specific genomic loci correlates with changes in expression patterns of genes positioned in the same regions. How misloading of SINE element-derived piRNAs into MIWI2 would affect LINE DNA methylation remains a mystery.

Concluding remarks

The piRNA pathway silences transposable elements, ensuring genome integrity and fertility. Similar to miRNAs and siRNAs, piRNAs associate with members of the Argonaute family of proteins, specifically with members of the PIWI clade. However, fundamental questions regarding their biogenesis and function remain unsolved, as the molecular mechanisms underlying these pathways have not yet been elucidated. Genetic studies have implicated different factors to be involved in piRNA biogenesis but the exact role of many of them remains so far unknown. In this thesis I investigated on two RNA helicases involved in the pathway: Vasa and TDRD9.

RNA helicases consume ATP to dynamically interact with RNA and are pervasive in multiprotein complexes throughout gene expression. Here, we report a mutation (DQAD) within the DEAD box of the germline helicase Vasa, which prevents the release of ATP hydrolysis products, restraining it in a closed conformation. This traps Vasa with its *in vivo* RNA targets and transient PIWI-containing biogenesis complexes. We show that RNA clamping by Vasa dynamically assembles an ATP-gated biogenesis machinery on transposon mRNAs to generate new secondary piRNAs. Consistently, mice bearing the *Mouse Vasa Homologue* (*Mvh*) DQAD mutation are infertile.

Another putative RNA helicase/ATPase involved in piRNA biogenesis is TDRD9. Although it is essential to guarantee mouse male fertility, its function is unknown. I showed that TDRD9 is an active ATPase and that a point mutation in its DExH-box (DEVH to DQVH) abolishes such activity. Knock-in mice that carry the DQVH mutation are male-sterile with a block in the early meiotic stages of spermatogenesis, proving that the ATPase activity of TDRD9 is essential for its function. The block is probably a consequence of uncontrolled DNA damages generated by de-repressed transposon which fail to be silenced during *de-novo* DNA methylation. Although TDRD9 is important for transposon silencing, it is likely not via a role in

Concluding remarks

piRNA biogenesis since PIWIs are correctly loaded with transposons derived piRNAs in the knock-in mouse mutant.

Taken together these data show how the same mutation can have different effects on RNA helicases from two different families. Glu to Gln mutation in the DEAD box of Vasa blocks the turnover of the enzyme, which fails to release ATP hydrolysis products and remains bound to its RNA targets. This has a dominant negative effect *in vivo* with heterozygote mice for the DQAD mutation in *Mvh* showing male-sterility. The same mutation in the DEVH-box of TDRD9 produces a catalytically-dead enzyme, unable to hydrolyse ATP even at a basal level. We do not know how this mutation affects the RNA binding activity of the protein but heterozygote mice for the DQVH mutation in *Tdrd9* are fertile, indicating absence of a dominant phenotype.

Overall I used biochemical, structural and genetic approaches to uncover the function of Vasa and Tdrd9, and I believe that my work consistently increased the knowledge on these two enzymes in the field.

List of references

- Ambros, V. (2003). MicroRNA Pathways in Flies and Worms. *Cell* 113, 673-676.
- Anand, A., and Kai, T. (2012). The tudor domain protein kumo is required to assemble the nuage and to generate germline piRNAs in *Drosophila*. *The EMBO journal* 31, 870-882.
- Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T., *et al.* (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442, 203-207.
- Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Current biology* : 11, 1017-1027.
- 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. *Molecular cell* 31, 785-799.
- Aravin, A.A., van der Heijden, G.W., Castaneda, J., Vagin, V.V., Hannon, G.J., and Bortvin, A. (2009). Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. *PLoS genetics* 5, e1000764.
- Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281-297.
- Baulcombe, D. (2004). RNA silencing in plants. *Nature* 431, 356-363.
- Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., and Izaurralde, E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes & Development* 20, 1885-1898.
- Berninger, P., Jaskiewicz, L., Khorshid, M., and Zavolan, M. (2011). Conserved generation of short products at piRNA loci. *BMC genomics* 12, 46.
- Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363-366.
- Bieniossek, C., Imasaki, T., Takagi, Y., and Berger, I. (2012). MultiBac: expanding the research toolbox for multiprotein complexes. *Trends in biochemical sciences* 37, 49-57.

References

- Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* *126*, 713-725.
- Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* *431*, 96-99.
- Braun, J.E., Truffault, V., Boland, A., Huntzinger, E., Chang, C.T., Haas, G., Weichenrieder, O., Coles, M., and Izaurralde, E. (2012). A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5' exonucleolytic degradation. *Nature structural & molecular biology* *19*, 1324-1331.
- 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.
- Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* *322*, 1387-1392.
- Buttner, K., Nehring, S., and Hopfner, K.P. (2007). Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nature structural & molecular biology* *14*, 647-652.
- Cai, X., Hagedorn, C.H., and 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., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Developmental Cell* *12*, 503-514.
- Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* *436*, 740-744.
- Chuma, S., and Nakano, T. (2013). piRNA and spermatogenesis in mice. *Philosophical transactions of the Royal Society of London Series B, Biological Sciences* *368*, 1609.
- Cogoni, C., and Macino, G. (1999). Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* *399*, 166-169.
- Cora, E., Pandey, R.R., Xiol, J., Taylor, J., Sachidanandam, R., McCarthy, A.A., and Pillai, R.S. (2014). The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs. *RNA* *20*, 773-781.

References

- Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes & Development* 12, 3715-3727.
- Czech, B., and Hannon, G.J. (2011). Small RNA sorting: matchmaking for Argonautes. *Nature reviews Genetics* 12, 19-31.
- Davis, I.W., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Research* 32, W615-619.
- De Fazio, S., Bartonicek, N., Di Giacomo, M., Abreu-Goodger, C., Sankar, A., Funaya, C., Antony, C., Moreira, P.N., Enright, A.J., and O'Carroll, D. (2011). The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* 480, 259-263.
- Deng, W., and Lin, H. (2002). miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Developmental Cell* 2, 819-830.
- Di Giacomo, M., Comazzetto, S., Saini, H., De Fazio, S., Carrieri, C., Morgan, M., Vasiliauskaite, L., Benes, V., Enright, A., and O'Carroll, D. (2013). Multiple Epigenetic Mechanisms and the piRNA Pathway Enforce LINE1 Silencing during Adult Spermatogenesis. *Molecular Cell*, Volume 50 50, 601-608.
- Djuranovic, S., Nahvi, A., and Green, R. (2012). miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* 336, 237-240.
- Dupeux, F., Rower, M., Seroul, G., Blot, D., and Marquez, J.A. (2011). A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta crystallographica Section D, Biological crystallography* 67, 915-919.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta crystallographica* 60, 2126-2132.
- Ephrussi, A., Dickinson, L.K., and Lehmann, R. (1991). Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell* 66, 37-50.
- Ephrussi, A., and Lehmann, R. (1992). Induction of germ cell formation by oskar. *Nature* 358, 387-392.
- Eulalio, A., Tritschler, F., Buttner, R., Weichenrieder, O., Izaurralde, E., and Truffault, V. (2009). The RRM domain in GW182 proteins contributes to miRNA-mediated gene silencing. *Nucleic Acids Research* 37, 2974-2983.

References

- Fairman-Williams, M.E., Guenther, U.P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. *Current opinion in structural biology* *20*, 313-324.
- Fei Liu, A.P., Eckhard Jankowsky (2008). ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *PNAS* *51*, 20209–20214.
- Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature reviews Genetics* *9*, 102-114.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* *391*, 806-811.
- 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.
- Frost, R.J., Hamra, F.K., Richardson, J.A., Qi, X., Bassel-Duby, R., and Olson, E.N. (2010). MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 11847-11852.
- Fukaya, T., and Tomari, Y. (2012). MicroRNAs mediate gene silencing via multiple different pathways in *Drosophila*. *Molecular Cell* *48*, 825-836.
- Gavis, E.R., Lunsford, L., Bergsten, S.E., and Lehmann, R. (1996). A conserved 90 nucleotide element mediates translational repression of nanos RNA. *Development* *122*, 2791-2800.
- Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. *Nature reviews Genetics* *10*, 94-108.
- Gillespie, D.E., and Berg, C.A. (1995). Homeless is required for RNA localization in *Drosophila* oogenesis and encodes a new member of the DE-H family of RNA-dependent ATPases. *Genes & Development* *9*, 2495-2508.
- Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* *312*, 75-79.
- 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.
- Goodier, J.L., and Kazazian, H.H., Jr. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* *135*, 23-35.

References

- Goriaux, C., Desset, S., Renaud, Y., Vaury, C., and Brasset, E. (2014). Transcriptional properties and splicing of the flamenco piRNA cluster. *EMBO reports* *15*, 411-418.
- Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature* *432*, 235-240.
- Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* *455*, 1193-1197.
- Grivna, S.T., Beyret, E., Wang, Z., and Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. *Genes & Development* *20*, 1709-1714.
- Gu, W., Lee, H.C., Chaves, D., Youngman, E.M., Pazour, G.J., Conte, D., Jr., and Mello, C.C. (2012). CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as *C. elegans* piRNA precursors. *Cell* *151*, 1488-1500.
- 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.
- Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* *404*, 293-296.
- Han, B.W., Wang, W., Li, C., Weng, Z., and Zamore, P.D. (2015). Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science* *348*, 817-821.
- Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F.S., and Brennecke, J. (2013). The genetic makeup of the *Drosophila* piRNA pathway. *Molecular Cell* *50*, 762-777.
- Handler, D., Olivieri, D., Novatchkova, M., Gruber, F.S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R., and Brennecke, J. (2011). A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *The EMBO journal* *30*, 3977-3993.
- Hilbert, M., Kebbel, F., Gubaev, A., and Klostermeier, D. (2011). eIF4G stimulates the activity of the DEAD box protein eIF4A by a conformational guidance mechanism. *Nucleic Acids Research* *39*, 2260-2270.
- Homolka, D., Pandey, R.R., and Pillai, R.S. (2015). RNA features within precursors instruct directional and processive primary piRNA production. *Cell Reports* *12*(3) 418-428

References

- Honda, S., Kirino, Y., Maragkakis, M., Alexiou, P., Ohtaki, A., Murali, R., Mourelatos, Z., and Kirino, Y. (2013). Mitochondrial protein BmPAPI modulates the length of mature piRNAs. *RNA* *19*, 1405-1418.
- 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. *Current biology : CB* *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.
- Huang, C.F., and Zhu, J.K. (2014). RNA Splicing Factors and RNA-Directed DNA Methylation. *Biology* *3*, 243-254.
- Humphreys, D.T., Westman, B.J., Martin, D.I., and Preiss, T. (2005). MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *PNAS* *102*, 16961-16966.
- Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature reviews Genetics* *12*, 99-110.
- Illmensee, K., and Mahowald, A.P. (1974). Transplantation of posterior polar plasm in *Drosophila*. Induction of germ cells at the anterior pole of the egg. *Proceedings of the National Academy of Sciences of the United States of America* *71*, 1016-1020.
- Irene Briguglio, S.P., Paola Corona, and Antonio Carta (2011). Inhibition of RNA Helicases of ssRNA+ Virus Belonging to Flaviviridae, Coronaviridae and Picornaviridae Families. *International Journal of Medicinal Chemistry* *2011*.
- Jazdowska-Zagrodzinska, B. (1966). Experimental studies on the role of 'polar granules' in the segregation of pole cells in *Drosophila melanogaster*. *Journal of embryology and experimental morphology* *16*, 391-399.
- Jinek, M., and Doudna, J.A. (2009). A three-dimensional view of the molecular machinery of RNA interference. *Nature* *457*, 405-412.
- Joshua-Tor, L., and Hannon, G.J. (2011). Ancestral roles of small RNAs: an Ago-centric perspective. *Cold Spring Harbor perspectives in biology* *3*, a003772.
- Kabsch, W. (2010). Xds. *Acta crystallographica Section D, Biological crystallography* *66*, 125-132.

References

- Kamminga, L.M., Luteijn, M.J., den Broeder, M.J., Redl, S., Kaaij, L.J.T., Roovers, E.F., Ladurner, P., Berezikov, E., and Ketting, R.F. (2010). Hen1 is required for oocyte development and piRNA stability in zebrafish. *Embo Journal* *29*, 3688-3700.
- Kawaoka, S., Hayashi, N., Suzuki, Y., Abe, H., Sugano, S., Tomari, Y., Shimada, T., and Katsuma, S. (2009). The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. *RNA* *15*, 1258-1264.
- Kawaoka, S., Izumi, N., Katsuma, S., and Tomari, Y. (2011). 3' end formation of PIWI-interacting RNAs in vitro. *Molecular Cell* *43*, 1015-1022.
- Kennerdell, J.R., and Carthew, R.W. (1998). Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell* *95*, 1017-1026.
- Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes & Development* *15*, 2654-2659.
- Kirino, Y., Kim, N., de Planell-Saguer, M., Khandros, E., Chiorean, S., Klein, P.S., Rigoutsos, I., Jongens, T.A., and Mourelatos, Z. (2009). Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. *Nature Cell Biology* *11*, 652-658.
- Kirino, Y., and Mourelatos, Z. (2007). The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *RNA* *13*, 1397-1401.
- Kirino, Y., Vourekas, A., Sayed, N., de Lima Alves, F., Thomson, T., Lasko, P., Rappsilber, J., Jongens, T.A., and Mourelatos, Z. (2010). Arginine methylation of Aubergine mediates Tudor binding and germ plasm localization. *RNA* *16*, 70-78.
- Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A., *et al.* (2009). The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* *138*, 1137-1149.
- Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* *42*, D68-73.
- Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., *et al.* (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* *131*, 839-849.
- Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Takamatsu, K., Chuma, S., Kojima-Kita, K., Shiromoto, Y., Asada, N., Toyoda, A., Fujiyama, A., *et al.* (2010). MVH in piRNA processing and gene silencing of retrotransposons. *Genes & Development* *24*, 887-892.

References

- 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 & Development* 22, 908-917.
- Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Current biology* : 14, 2162-2167.
- Lasko, P.F., and Ashburner, M. (1988). The product of the *Drosophila* gene vasa is very similar to eukaryotic initiation factor-4A. *Nature* 335, 611-617.
- Le Thomas, A., Rogers, A.K., Webster, A., Marinov, G.K., Liao, S.E., Perkins, E.M., Hur, J.K., Aravin, A.A., and Toth, K.F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes & Development* 27, 390-399.
- Lee, R.C., Feinbaum, R.L., and 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., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. (2004a). MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal* 23, 4051-4060.
- Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004b). Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* 117, 69-81.
- Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15-20.
- Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., *et al.* (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* 137, 509-521.
- Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915-926.
- Li, X.Z., Roy, C.K., Dong, X., Bolcun-Filas, E., Wang, J., Han, B.W., Xu, J., Moore, M.J., Schimenti, J.C., Weng, Z., *et al.* (2013). An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Molecular Cell* 50, 67-81.
- Liang, L., Diehl-Jones, W., and 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.

References

- Lim, A.K., and Kai, T. (2007). Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America* *104*, 6714-6719.
- Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* *433*, 769-773.
- Lin, H., and Spradling, A.C. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development* *124*, 2463-2476.
- Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. *Nature reviews Molecular cell biology* *12*, 505-516.
- Liu, F., Putnam, A., and Jankowsky, E. (2008). ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc Natl Acad Sci U S A* *105*, 20209-20214.
- 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.
- Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* *301*, 1921-1925.
- Mahowald, A.P. (1968). Polar granules of *Drosophila*. II. Ultrastructural changes during early embryogenesis. *The Journal of experimental zoology* *167*, 237-261.
- Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* *137*, 522-535.
- Marchini, F.K., de Godoy, L.M., Rampazzo, R.C., Pavoni, D.P., Probst, C.M., Gnad, F., Mann, M., and Krieger, M.A. (2011). Profiling the *Trypanosoma cruzi* phosphoproteome. *PloS one* *6*, e25381.
- Mathioudakis, N., Palencia, A., Kadlec, J., Round, A., Tripsianes, K., Sattler, M., Pillai, R.S., and Cusack, S. (2012). The multiple Tudor domain-containing protein TDRD1 is a molecular scaffold for mouse Piwi proteins and piRNA biogenesis factors. *RNA* *18*, 2056-2072.
- Mathonnet, G., Fabian, M.R., Svitkin, Y.V., Parsyan, A., Huck, L., Murata, T., Biffo, S., Merrick, W.C., Darzynkiewicz, E., Pillai, R.S., *et al.* (2007). MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science* *317*, 1764-1767.

References

- McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R.J. (2005). Likelihood-enhanced fast translation functions. *Acta crystallographica Section D, Biological crystallography* *61*, 458-464.
- Meijer, H.A., Kong, Y.W., Lu, W.T., Wilczynska, A., Spriggs, R.V., Robinson, S.W., Godfrey, J.D., Willis, A.E., and Bushell, M. (2013). Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. *Science* *340*, 82-85.
- Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular Cell* *15*, 185-197.
- Moazed, D. (2009). Small RNAs in transcriptional gene silencing and genome defence. *Nature* *457*, 413-420.
- Mohn, F., Handler, D., and Brennecke, J. (2015). Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science* *348*, 812-817.
- Mohn, F., Sienski, G., Handler, D., and Brennecke, J. (2014). The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell* *157*, 1364-1379.
- Murchison, E.P., Stein, P., Xuan, Z., Pan, H., Zhang, M.Q., Schultz, R.M., and Hannon, G.J. (2007). Critical roles for Dicer in the female germline. *Genes & Development* *21*, 682-693.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica* *53*, 240-255.
- Nakanishi, K., Weinberg, D.E., Bartel, D.P., and Patel, D.J. (2012). Structure of yeast Argonaute with guide RNA. *Nature* *486*, 368-374.
- Nishida, K.M., Iwasaki, Y.W., Murota, Y., Nagao, A., Mannen, T., Kato, Y., Siomi, H., and Siomi, M.C. (2015). Respective functions of two distinct Siwi complexes assembled during PIWI-interacting RNA biogenesis in *Bombyx* germ cells. *Cell reports* *10*, 193-203.
- Nishida, K.M., Okada, T.N., Kawamura, T., Mituyama, T., Kawamura, Y., Inagaki, S., Huang, H., Chen, D., Kodama, T., Siomi, H., *et al.* (2009). Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in *Drosophila* germlines. *The EMBO journal* *28*, 3820-3831.
- Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* *130*, 89-100.
- Okano, M., Bell, D., Haber, D., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian Development. *Cell* *99*, 247-257.

References

- Olivieri, D., Sykora, M.M., Sachidanandam, R., Mechtler, K., and Brennecke, J. (2010). An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *The EMBO journal* *29*, 3301-3317.
- Orelle, C., Dalmas, O., Gros, P., Di Pietro, A., and Jault, J.M. (2003). The conserved glutamate residue adjacent to the Walker-B motif is the catalytic base for ATP hydrolysis in the ATP-binding cassette transporter BmrA. *The Journal of biological chemistry* *278*, 47002-47008.
- Ott, K.M., Nguyen, T., and Navarro, C. (2014). The DExH box helicase domain of spindle-E is necessary for retrotransposon silencing and axial patterning during *Drosophila* oogenesis. *G3* *4*, 2247-2257.
- Ozgur, S., Buchwald, G., Falk, S., Chakrabarti, S., Prabu, J.R., and Conti, E. (2015). The conformational plasticity of eukaryotic RNA-dependent ATPases. *The FEBS journal* *282*, 850-863.
- Pak, J., and Fire, A. (2007). Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* *315*, 241-244.
- Pandey, R.R., Tokuzawa, Y., Yang, Z., Hayashi, E., Ichisaka, T., Kajita, S., Asano, Y., Kunieda, T., Sachidanandam, R., Chuma, S., *et al.* (2013). Tudor domain containing 12 (TDRD12) is essential for secondary PIWI interacting RNA biogenesis in mice. *Proceedings of the National Academy of Sciences of the United States of America* *110*, 16492-16497.
- Pane, A., Wehr, K., and Schupbach, T. (2007). zucchini and squash encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Developmental Cell* *12*, 851-862.
- Patil, V.S., and Kai, T. (2010). Repression of retroelements in *Drosophila* germline via piRNA pathway by the Tudor domain protein Tejas. *Current biology : CB* *20*, 724-730.
- Pause, A., and Sonenberg, N. (1992). Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *The EMBO journal* *11*, 2643-2654.
- Pezic, D., Manakov, S.A., Sachidanandam, R., and Aravin, A.A. (2014). piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes & Development* *28*, 1410-1428.
- 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.
- Pillai, R.S., and Chuma, S. (2012). piRNAs and their involvement in male germline development in mice. *Dev Growth Differ* *54*, 78-92.

References

- Reuter, M., Berninger, P., Chuma, S., Shah, H., Hosokawa, M., Funaya, C., Antony, C., Sachidanandam, R., and Pillai, R.S. (2011). Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* *480*, 264-267.
- 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. *Nature structural & molecular biology* *16*, 639-646.
- Rivas, F.V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J., and Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. *Nature structural & molecular biology* *12*, 340-349.
- Rozhkov, N.V., Hammell, M., and Hannon, G.J. (2013). Multiple roles for Piwi in silencing *Drosophila* transposons. *Genes & Development* *27*, 400-412.
- Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (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., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature* *448*, 83-86.
- Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K.M., Siomi, H., and Siomi, M.C. (2010). Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes & Development* *24*, 2493-2498.
- Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M.C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes & Development* *20*, 2214-2222.
- 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 & Development* *21*, 1603-1608.
- Sanchez Alvarado, A., and Newmark, P.A. (1999). Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Proceedings of the National Academy of Sciences of the United States of America* *96*, 5049-5054.
- Sarot, E., Payen-Groschene, G., Bucheton, A., and 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.

References

- Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A., and Gvozdev, V. (2006). Telomere elongation is under the control of the RNAi-based mechanism in the *Drosophila* germline. *Genes & Development* *20*, 345-354.
- Saxe, J.P., Chen, M., Zhao, H., and Lin, H. (2013). Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. *The EMBO journal* *32*, 1869-1885.
- Schirle, N.T., and MacRae, I.J. (2012). The crystal structure of human Argonaute2. *Science* *336*, 1037-1040.
- Schirle, N.T., Sheu-Gruttadauria, J., and MacRae, I.J. (2014). Structural basis for microRNA targeting. *Science* *346*, 608-613.
- Schubach, T., and Wieschaus, E. (1986). Germline autonomy of maternal-effect mutations altering the embryonic body pattern of *Drosophila*. *Dev Biol* *113*, 443-448.
- Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell* *115*, 199-208.
- Sengoku T, N.O., Nakamura A, Kobayashi S, Yokoyama S. (2006). Structural basis for RNA unwinding by the DEAD-Box protein *Drosophila* VASA. *Cell*, vol.125 (287-300).
- Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., and Yokoyama, S. (2006). Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. *Cell* *125*, 287-300.
- Shoji, M., Tanaka, T., Hosokawa, M., Reuter, M., Stark, A., Kato, Y., Kondoh, G., Okawa, K., Chujo, T., Suzuki, T., *et al.* (2009). The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Developmental cell* *17*, 775-787.
- Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nature reviews Molecular cell biology* *12*, 246-258.
- Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nature reviews Genetics* *8*, 272-285.
- Snee, M.J., and Macdonald, P.M. (2004). Live imaging of nuage and polar granules: evidence against a precursor-product relationship and a novel role for Oskar in stabilization of polar granule components. *Journal of cell science* *117*, 2109-2120.
- 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.

References

- Soper, S.F., van der Heijden, G.W., Hardiman, T.C., Goodheart, M., Martin, S.L., de Boer, P., and Bortvin, A. (2008). Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Developmental Cell* *15*, 285-297.
- Sugiyama, T., Cam, H., Verdel, A., Moazed, D., and Grewal, S.I. (2005). RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *PNAS* *102*, 152-157.
- Suh, N., Baehner, L., Moltzahn, F., Melton, C., Shenoy, A., Chen, J., and Blelloch, R. (2010). MicroRNA function is globally suppressed in mouse oocytes and early embryos. *Current biology : CB* *20*, 271-277.
- Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., *et al.* (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* *453*, 534-538.
- Tanaka, T., Hosokawa, M., Vagin, V.V., Reuter, M., Hayashi, E., Mochizuki, A.L., Kitamura, K., Yamanaka, H., Kondoh, G., Okawa, K., *et al.* (2011). Tudor domain containing 7 (Tdrd7) is essential for dynamic ribonucleoprotein (RNP) remodeling of chromatoid bodies during spermatogenesis. *PNAS* *108*, 10579-10584.
- Tang, F., Kaneda, M., O'Carroll, D., Hajkova, P., Barton, S.C., Sun, Y.A., Lee, C., Tarakhovsky, A., Lao, K., and Surani, M.A. (2007). Maternal microRNAs are essential for mouse zygotic development. *Genes & Development* *21*, 644-648.
- Thermann, R., and Hentze, M.W. (2007). Drosophila miR2 induces pseudo-polysomes and inhibits translation initiation. *Nature* *447*, 875-878.
- Tinker, R., Silver, D., and Montell, D.J. (1998). Requirement for the vasa RNA helicase in gurken mRNA localization. *Development Biology* *199*, 1-10.
- Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P.D. (2004). A protein sensor for siRNA asymmetry. *Science* *306*, 1377-1380.
- 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.
- Vagin, V.V., Wohlschlegel, J., Qu, J., Jonsson, Z., Huang, X., Chuma, S., Girard, A., Sachidanandam, R., Hannon, G.J., and Aravin, A.A. (2009). Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes & Development* *23*, 1749-1762.

References

- Vagin, V.V., Yu, Y., Jankowska, A., Luo, Y., Wasik, K.A., Malone, C.D., Harrison, E., Rosebrock, A., Wakimoto, B.T., Fagegaltier, D., *et al.* (2013). Minotaur is critical for primary piRNA biogenesis. *RNA* *19*, 1064-1077.
- 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.
- Voinnet, O. (2008). Use, tolerance and avoidance of amplified RNA silencing by plants. *Trends in plant science* *13*, 317-328.
- Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* *297*, 1833-1837.
- von Moeller, P.B., Conti, E. (2009). The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner. *NSMB* *16*, 247-254.
- Vourekas, A., Zheng, K., Fu, Q., Maragkakis, M., Alexiou, P., Ma, J., Pillai, R.S., Mourelatos, Z., and Wang, P.J. (2015). The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. *Genes & Development* *29*, 617-629.
- Walbott, H., Mouffok, S., Capeyrou, R., Lebaron, S., Humbert, O., van Tilbeurgh, H., Henry, Y., and Leulliot, N. (2010). Prp43p contains a processive helicase structural architecture with a specific regulatory domain. *The EMBO journal* *29*, 2194-2204.
- Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO journal* *1*, 945-951.
- Wang, C., Dickinson, L.K., and Lehmann, R. (1994). Genetics of nanos localization in *Drosophila*. *Developmental Dynamics* *199*, 103-115.
- Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., *et al.* (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* *453*, 539-543.
- Waterhouse, P.M., Graham, M.W., and Wang, M.B. (1998). Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *PNAS* *95*, 13959-13964.
- Weir, J.R., Bonneau, F., Hentschel, J., and Conti, E. (2010). Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *PNAS* *107*, 12139-12144.

References

- Wightman, B., Ha, I., and 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.
- Wilczynska, A., and Bushell, M. (2015). The complexity of miRNA-mediated repression. *Cell death and differentiation* *22*, 22-33.
- Xiol, J., Cora, E., Kogelgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z., Berninger, P., Palencia, A., *et al.* (2012). A role for Fkbp6 and the chaperone machinery in piRNA amplification and transposon silencing. *Molecular Cell* *47*, 970-979.
- Xiol, J., Spinelli, P., Laussmann, M.A., Homolka, D., Yang, Z., Cora, E., Coute, Y., Conn, S., Kadlec, J., Sachidanandam, R., *et al.* (2014). RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts. *Cell* *157*, 1698-1711.
- Yabuta, Y., Ohta, H., Abe, T., Kurimoto, K., Chuma, S., and Saitou, M. (2011). TDRD5 is required for retrotransposon silencing, chromatoid body assembly, and spermiogenesis in mice. *The Journal of cell biology* *192*, 781-795.
- Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Development* *17*, 3011-3016.
- Yu, B., Yang, Z.Y., Li, J.J., Minakhina, S., Yang, M.C., Padgett, R.W., Steward, R., and Chen, X.M. (2005). Methylation as a crucial step in plant microRNA biogenesis. *Science* *307*, 932-935.
- Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi. *Cell* *101*, 25-33.
- Zamparini, A.L., Davis, M.Y., Malone, C.D., Vieira, E., Zavadil, J., Sachidanandam, R., Hannon, G.J., and Lehmann, R. (2011). Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in *Drosophila*. *Development* *138*, 4039-4050.
- Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B.S., Schultz, N., Vreven, T., Meignin, C., Davis, I., Zamore, P.D., *et al.* (2012). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell* *151*, 871-884.
- Zhang, Z., Koppetsch, B.S., Wang, J., Tipping, C., Weng, Z., Theurkauf, W.E., and Zamore, P.D. (2014a). Antisense piRNA amplification, but not piRNA production or nuage assembly, requires the Tudor-domain protein Qin. *The EMBO journal* *33*, 536-539.
- Zhang, Z., Wang, J., Schultz, N., Zhang, F., Parhad, S.S., Tu, S., Vreven, T., Zamore, P.D., Weng, Z., and Theurkauf, W.E. (2014b). The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* *157*, 1353-1363.

References

Zheng, K., Xiol, J., Reuter, M., Eckardt, S., Leu, N.A., McLaughlin, K.J., Stark, A., Sachidanandam, R., Pillai, R.S., and Wang, P.J. (2010). Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. *PNAS* *107*, 11841-11846.

Acknowledgments

Let me first thank my supervisor, Ramesh Pillai, for giving me the opportunity to work in this wonderful lab, with these amazing scientists. Let me thank him for letting me do my mistakes, for the endless discussions, for teaching me the fundamental concept of ‘good science’.

I would like to thank Alena and Marie-Laure for taking their time to read my thesis, if you managed to arrive here it means you read it all, well done! I hope you found it interesting and forgive me for eventual headaches it may have caused you. Thanks to my TAC members Dónal O’Carroll, André Verdel and Stephen Cusack, for pushing me to improve myself at every seminar.

I would like to thank all the people who made an active contribution to this work. Starting with Jordi, who initiated the Vasa project and guided my first steps into real science. Thanks to Stephen Cusack and Ian Kadlec, who solved the Vasa structure. Thanks to Mikhail Grigoriev who performed the molecular modeling and identified the histidine switch on Vasa. A big thank you to Zhaolin (for the IP of MVH), Raman (for the IP of MILI and MIWI2 in P0 *Tdrd9* mutants) and David (for the bioinformatics analysis of *Tdrd9* libraries).

Stellar thanks to Miriam, Sara, Magda, Jorge, Silvano and Maria, for reading my thesis and for your constructive comments, thanks a lot for your help. *Permettez-moi de remercier Leila pour 'aider' moi avec la traduction en français du résumé: merci beaucoup Leila!*

I would like to thank the other members of the lab who helped me with the day-to-day work, thank you Kuan-Ming, and thank you Elisa and Michael for teaching me so many things. Thank you to all EMBL Grenoble, it was a fantastic experience. Thank you HTX and EEF facilities. Thank you Joanna from the Mass spectrometry facility in Heidelberg.

A special thanks to my family, who supported me unconditionally for all these years, thank you for being there when I needed you more. Thank you Etienne for being a friend.

An additional thank you to my friends Berri, Lele, Zone, and all the pirates, for having rescued me from a painful summer with your boat, thank you for being a safe and endless presence in my life.

Appendix

RNA Clamping by Vasa Assembles a piRNA Amplifier Complex on Transposon Transcripts

Jordi Xiol,^{1,2,7} Pietro Spinelli,^{1,2} Maïke A. Laussmann,^{1,2,3} David Homolka,^{1,2} Zhaolin Yang,^{1,2} Elisa Cora,^{1,2} Yann Coute,⁴ Simon Conn,^{1,2} Jan Kadlec,^{1,2} Ravi Sachidanandam,⁵ Marko Kaksonen,³ Stephen Cusack,^{1,2} Anne Ephrussi,⁶ and Ramesh S. Pillai^{1,2,*}

¹European Molecular Biology Laboratory, Grenoble Outstation

²Unit for Virus Host-Cell Interactions

University Grenoble Alpes-EMBL-CNRS, 71 avenue des Martyrs, 38042, France

³Cell Biology and Biophysics Unit, EMBL, 69117 Heidelberg, Germany

⁴Laboratoire Biologie à Grande Echelle, IRTSV, CEA, 38054 Grenoble, France

⁵Department of Oncological Sciences, Icahn School of Medicine at Sinai, New York, NY 10029, USA

⁶Developmental Biology Unit, EMBL, 69117 Heidelberg, Germany

⁷Present address: Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA

*Correspondence: pillai@embl.fr

<http://dx.doi.org/10.1016/j.cell.2014.05.018>

SUMMARY

Germline-specific Piwi-interacting RNAs (piRNAs) protect animal genomes against transposons and are essential for fertility. piRNAs targeting active transposons are amplified by the ping-pong cycle, which couples Piwi endonucleolytic slicing of target RNAs to biogenesis of new piRNAs. Here, we describe the identification of a transient Amplifier complex that mediates biogenesis of secondary piRNAs in insect cells. Amplifier is nucleated by the DEAD box RNA helicase Vasa and contains the two Piwi proteins participating in the ping-pong loop, the Tudor protein Qin/Kumo and antisense piRNA guides. These components assemble on the surface of Vasa's helicase domain, which functions as an RNA clamp to anchor Amplifier onto transposon transcripts. We show that ATP-dependent RNP remodeling by Vasa facilitates transfer of 5' sliced piRNA precursors between ping-pong partners, and loss of this activity causes sterility in *Drosophila*. Our results reveal the molecular basis for the small RNA amplification that confers adaptive immunity against transposons.

INTRODUCTION

Mobile genomic elements called transposons constitute a large part of eukaryotic genomes and present a formidable threat to genome integrity due to their ability to cause mutations or ectopic recombination events (Kazazian, 2004). Animal gonads employ the dedicated PIWI clade Argonautes and their associated 24–30 nucleotide (nt) Piwi-interacting RNAs (piRNAs) to suppress transposon activity. piRNAs guide Piwi proteins to their

transposon targets, which are identified by sequence complementarity between the piRNA and the target. Some Piwi proteins function as small RNA-guided endonucleases (slicers), whereas nuclear Piwi complexes may also mediate silencing by altering chromatin or DNA methylation status of target genomic loci. The importance of this pathway is highlighted by the fact that Piwi mutants display infertility due to defects in germline development, presumably as a result of genome instability caused by transposon activation (Bagijn et al., 2012; Ghildiyal and Zamore, 2009; Luteijn and Ketting, 2013).

Guidance information for Piwi proteins is genetically encoded into a few hundred genomic regions called piRNA clusters (Brennecke et al., 2007). These genomic intervals serve as a registry of all transposons hosted in the genome by containing a high density of transposon fragment insertions. Clusters are transcribed into a long single-stranded piRNA precursor (up to 100 kb long) and are then exported to the cytoplasm by RNA export factors like Uap56 (Zhang et al., 2012) for maturation into tens of thousands of piRNAs via a poorly understood primary processing pathway (Brennecke et al., 2007). Processing is believed to take place in cytoplasmic perinuclear granules called nuage that concentrate most known piRNA processing factors (Lim and Kai, 2007). By providing germ cells with a complex mixture of sequences to target all mobile elements present in the genome, these primary piRNAs act as the genetically encoded component of an RNA-based innate immune system (Aravin et al., 2007).

Germ cells have an additional and more elegant piRNA biogenesis pathway where target silencing itself triggers generation of new piRNAs. Sequence analyses of small RNAs associating with the *Drosophila* Piwi proteins (Aubergine and Ago3) formed the basis for such a proposal, and these ideas are incorporated into the so-called ping-pong model (Brennecke et al., 2007; Gunawardane et al., 2007). It is proposed that antisense primary piRNAs guide Aubergine (Aub) for cytoplasmic endonucleolytic

cleavage (slicing) of target transposon RNAs. One of the resulting cleavage fragments is then utilized to generate a new secondary piRNA that associates with Ago3, with the cleavage site marking its 5' end. Given their origin from transposon transcripts, Ago3-bound piRNAs are of sense orientation. These, in turn, guide Ago3 slicer activity to complementary cluster transcripts, with a cleavage fragment maturing exactly as the same Aub-bound antisense piRNA that initiated the transposon cleavage. In such a scenario, Ago3 slicing serves to prioritize production of specific antisense piRNAs from piRNA cluster transcripts (Li et al., 2009). The system is built such that abundant transposons will produce more Ago3-loaded piRNAs, resulting in even greater production of antisense piRNAs in Aub. Such a feedforward amplification loop represents the adaptive arm of the RNA-based immune system that monitors transposon activity and selectively amplifies those piRNAs that target the most active transposons. The versatility of the ping-pong cycle is evident from the fact that maternally deposited piRNAs can also serve as inputs, thereby allowing the embryo to rapidly ramp up immunity against paternally inherited transposons to avert infertility caused by hybrid dysgenesis (Brennecke et al., 2008; Grentzinger et al., 2012). It can also promote emergence of new piRNA loci by targeting genomic regions in *trans* (de Vanssay et al., 2012).

We currently do not understand how such a complex small RNA amplification pathway works. RNA slicing by Piwi proteins and that occurring during RNA interference (RNAi) (Meister, 2013) are biochemically indistinguishable. However, RNAi leads to complete degradation of the target. Germ cells must therefore possess an undefined machinery that safely transfers the Piwi-generated slicer cleavage fragments from one ping-pong Piwi partner to the other for maturation as secondary piRNAs. Genetic and biochemical analyses have uncovered a number of factors that are essential for secondary piRNA biogenesis (Czech et al., 2013; Malone et al., 2009), but how they contribute to what is likely to be a multistep processing pathway is currently unclear. RNA helicases are prime candidates for bringing about dynamic associations and remodeling of ribonucleoprotein (RNP) complexes during a variety of RNA processing reactions (Linder and Jankowsky, 2011), and they are among the list of known piRNA biogenesis factors (Figure S1A available online). Among these, we focused our attention on the DEAD box RNA helicase Vasa, as it is the only factor with a conserved role in secondary biogenesis in both flies and mice (Kuramochi-Miyagawa et al., 2010; Malone et al., 2009). *vasa* was originally identified as a maternal-effect gene that is required for development of the female germline in *Drosophila* (Lasko and Ashburner, 1988; Schupbach and Wieschaus, 1986). Like other members of the DEAD box family, Vasa is demonstrated to have ATPase, RNA binding, and RNA unwinding activities (Liang et al., 1994; Sen-goku et al., 2006), but its molecular role in the piRNA pathway is not known.

Here, we explored the *in vivo* function of Vasa using the insect ovarian cell culture model *Bombyx mori* BmN4 (Kawaoka et al., 2009). RNA helicases like Vasa are ATP-gated RNA-binding proteins that engage in dynamic associations. We used a point mutation (DQAD) within the DEAD box of Vasa to prevent release of ATP hydrolysis products, locking the protein onto its bound RNA targets and reducing its *in vivo* dynamics. Surprisingly, this also

trapped a piRNA amplification complex that provides detailed insight into how endonucleolytic cleavage during transposon silencing is linked to generation of a new secondary piRNA. This transiently assembled complex, which we named Amplifier, includes Vasa, the two ping-pong Piwi partners, and the tudor protein Qin/Kumo. Biochemical and structural data reveal that RNA clamping on transposon transcripts by Vasa creates a binding platform for Amplifier components at the surface of Vasa's helicase domain, which is in its ATP-bound, closed conformation. ATP hydrolysis by Vasa opens the helicase domains, resulting in Amplifier disassembly and safe handover of sliced 5' processed precursors between Piwi proteins for further maturation into piRNAs. Genetic experiments in flies underscore the importance of this ATP-controlled process for fertility. Our results provide detailed mechanistic insight into how transposon cleavage fragments are utilized for generating new piRNAs that preserve genome integrity in the germline.

RESULTS

Comprehensive Transcriptome of the BmN4 Cell Line

The *Bombyx mori* (silkworm) ovary-derived BmN4 cell line is the only cell culture model for piRNA research that has an active ping-pong cycle in operation. These cells express two Piwi proteins—Siwi (Aub ortholog in *Bombyx*) and Ago3—that associate with piRNAs bearing sequence signatures consistent with secondary piRNA biogenesis (Kawaoka et al., 2009). To facilitate studies with this cell line, we cataloged the entire set of piRNA pathway factors in BmN4 using *de novo* assembly of transcriptome sequencing data (Figure S1A). When compared to *Drosophila*, several of the factors essential for both primary and secondary biogenesis (including Vasa) are conserved. However, notable absences include that of the effector protein Squash and nuclear factors like Rhino, Cutoff, and Deadlock, which are needed for piRNA biogenesis from dual-strand (ping-pong) clusters (Chen et al., 2007; Czech et al., 2013; Klattenhoff et al., 2009; Pane et al., 2007) (Figure S1A). The absence of the latter is likely due to the euchromatic nature of piRNA clusters in BmN4 (Kawaoka et al., 2013), precluding the need for these factors, as they serve to access *Drosophila* piRNA clusters that are embedded in constitutive heterochromatin (Brennecke et al., 2007; Klattenhoff et al., 2009). We also did not find any evidence for the existence of a nuclear Piwi pathway in these cells. Overall, these findings reveal that a unique selection of biogenesis factors fuels the ping-pong cycle in BmN4 cells.

A DEAD Box Mutation Reduces *In Vivo* Dynamics of Vasa

Our initial attempt at understanding *Bombyx mori* Vasa's role in the piRNA pathway consisted of characterizing isolated complexes from transfected BmN4 cells. However, these failed to reveal any interaction with core piRNA pathway components (see below). DEAD box RNA helicases like Vasa are known to mediate dynamic associations as part of multiprotein assemblies *in vivo* (Linder and Jankowsky, 2011), which might be hard to detect in biochemical experiments due to their transient nature. This prompted us to examine the dynamics of the proteins involved in the ping-pong cycle by live-cell imaging experiments rather than by looking at a steady-state situation.

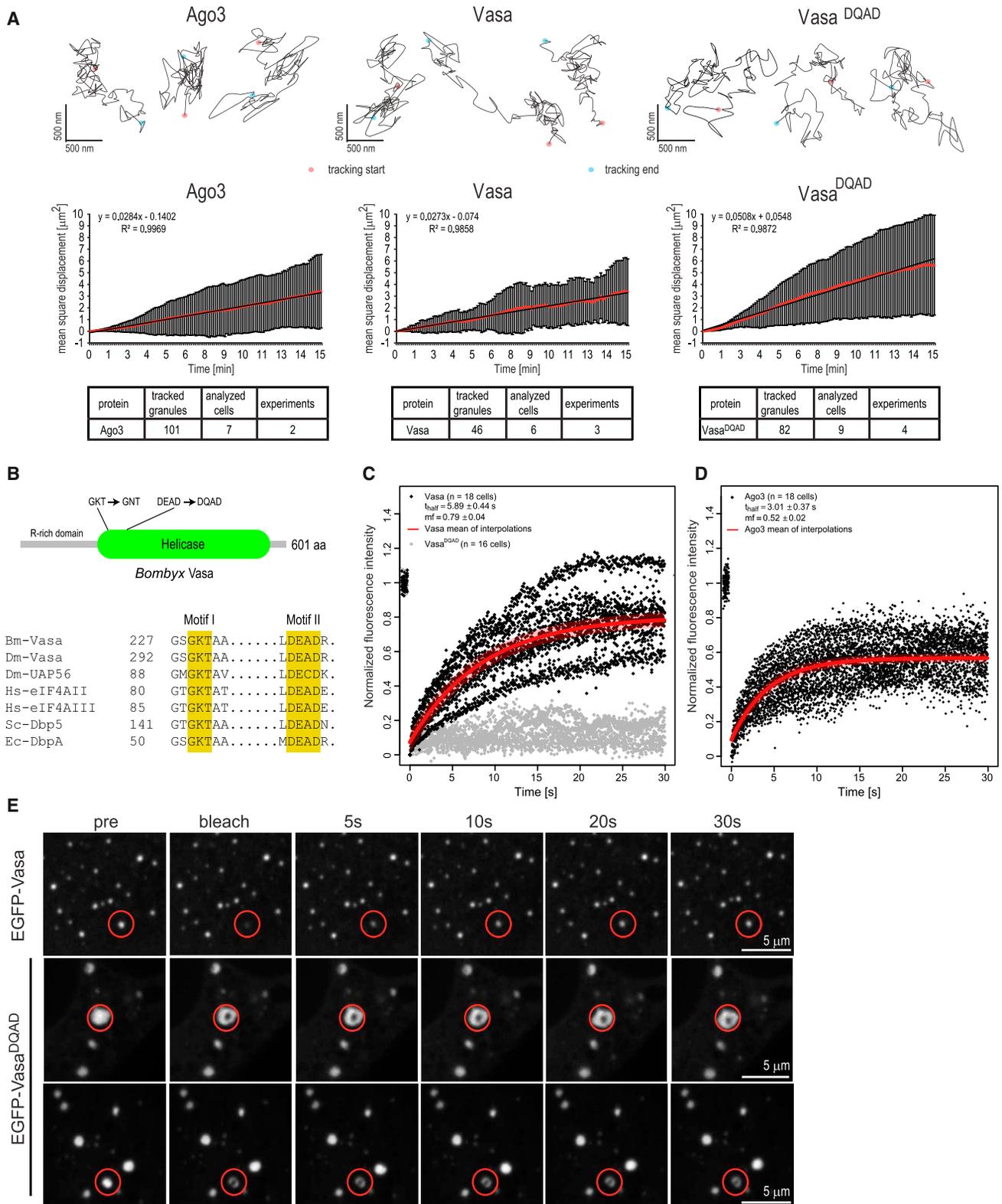


Figure 1. DQAD Mutation in the DEAD Box of Vasa Reduces Its In Vivo Dynamics

(A) BmN4 cells expressing fluorescently tagged Vasa, Vasa^{DQAD}, or Ago3 display abundant cytoplasmic granules, and these were tracked by live-cell imaging. Start and end of tracking is highlighted with a red and blue dot, respectively. Analysis indicates a random, nonlinear movement of these granules, which is unaffected by the DQAD mutation in Vasa.

(legend continued on next page)

Transiently expressed EGFP-Vasa and mCherry-Ago3 colocalize in numerous perinuclear cytoplasmic granules in BmN4 cells. These are analogous to the electron-dense granules called nuage observed in *Drosophila* ovarian nurse cells, which are credited to be piRNA processing centers (Liang et al., 1994; Lim and Kai, 2007). Cell tracking experiments now reveal that, far from being static structures, these display random, nonlinear movements (Figure 1A and Movie S1). More frequently, granules meet to fuse briefly before dissociating, whereas occasionally smaller granules fuse into larger structures. To evaluate the dynamics of the proteins within the nuage, we carried out fluorescence recovery after photobleaching (FRAP) experiments (Movie S2). The measured halftime of fluorescence recovery rates for EGFP-Vasa (5.89 ± 0.44 s) (Figures 1C and 1E) and EGFP-Ago3 (3.01 ± 0.37 s) (Figures 1D, S1B, and S1C) suggest a rapid transit of the proteins through the granules. Thus, any interactions mediated by them within the nuage should be short lived.

Because most activities of RNA helicases are linked to their consumption of ATP (Linder and Jankowsky, 2011), we examined whether abolishing ATPase activity might interfere with the dynamics of Vasa. The DEAD box family is characterized by the presence of a number of sequence motifs within the helicase module (Caruthers and McKay, 2002), prominent among them being the motif II (Asp-Glu-Ala-Asp; D-E-A-D). A single amino acid mutation within this motif (DEAD → DQAD) (Figure 1B) is shown to abolish ATPase activity in other family members (Pause and Sonenberg, 1992). We created a similar mutation within Vasa and expressed its EGFP-tagged version in BmN4 cells. A first indication of an impact of the mutation was the accumulation of EGFP-Vasa^{DQAD} in granules that are slightly larger than those formed by EGFP-Vasa (Figures 1E, 2F, and 2G). Nevertheless, tracking experiments did not reveal any impact of the mutation on the random mobility of the granules (Figure 1A). Strikingly, FRAP experiments showed that fluorescence in Vasa^{DQAD} granules failed to recover, irrespective of the size of the granule examined (Figures 1C and 1E). This is likely to be a consequence of the protein becoming trapped in the granule, preventing further ingress from the surrounding cytoplasm (Movies S3–S5). Thus, an emerging picture is one of highly dynamic events orchestrated within the nuage, which we managed to dampen as a result of the DQAD mutation within Vasa.

Vasa Is Part of a Transient piRNA Amplifier Complex

The retention of Vasa^{DQAD} in enlarged nuage prompted us to ask whether the point mutant might be accumulating in certain complexes in vivo. To this end, we isolated HA-Vasa or HA-Vasa^{DQAD} complexes and identified components by mass spectrometry. The protein arginine methyltransferase 5 (Prmt5 or Capsuleen in *Drosophila*) is common to both complexes, and this is explained by the presence of 20 arginine residues at the N termi-

nus of *Bombyx* Vasa (Figure 2A). Such arginine residues in *Drosophila* Vasa are shown to be targets of Prmt5 for symmetrical dimethyl arginine (sDMA) modifications (Kirino et al., 2010).

Interestingly, HA- and FLAG-Vasa^{DQAD} complexes contained a number of additional factors not present in wild-type Vasa immunoprecipitations (Figures 2A and S2A). In particular, the two most abundant hits among Vasa^{DQAD} interactors are the endogenous ping-pong Piwi partners Siwi (the *Bombyx* ortholog of Aub) and Ago3, which we confirmed by western blot analysis (Figure 2B). Furthermore, tandem immunopurification of FLAG-Vasa^{DQAD} followed by anti-Siwi immunoprecipitation reveals that all three proteins are part of the same entity (Figure 2C). Because the DQAD mutation in Vasa is predicted to prevent ATP hydrolysis, association with Piwi proteins is likely to be favored when Vasa is in its ATP-bound state. To confirm this, we created a mutation in the motif I (GKT → GNT) of the helicase domain that is predicted to prevent ATP binding (Figure 1B). Supporting the requirement for bound ATP, the Vasa^{GNT} mutant did not reveal any association with endogenous Piwi proteins (Figure 2B). The Tudor domain-containing protein Qin/Kumo, a factor essential for the ping-pong cycle (Anand and Kai, 2012; Zhang et al., 2011), was also identified in the Vasa^{DQAD} purifications (Figure 2A). We confirmed this result by western analysis using rabbit polyclonal antibodies to the endogenous Qin/Kumo expressed in BmN4 cells (Figures 2D and S2B). Because all of the validated components are involved in piRNA amplification, we called this Vasa-Siwi-Ago3-Qin/Kumo complex Amplifier.

Arginine methylation of Piwi proteins is shown to act as an affinity enhancer for mediating protein-protein interactions (Mathioudakis et al., 2012), so we wondered whether Vasa methylation plays any role in Amplifier assembly. Pointing to the absence of such a role, when we mutated all of the N-terminal arginines (to lysines; R → K) in Vasa^{DQAD}, the interaction with the Amplifier components was still maintained, as examined by mass spectrometry and western blot analysis (Figures 2A and 2E). In contrast, recovery of Prmt5 was abolished in both *Bombyx* Vasa^{R-K} and Vasa^{DQAD(R-K)} purifications (Figure 2A). We propose that the Amplifier complex is assembled on wild-type Vasa, but its transient nature prevents its detection (Figure S2C). However, the DQAD mutation in Vasa freezes this complex by preventing its disassembly, thereby allowing its identification.

Amplifier Assembly Takes Place within the Nuage

In BmN4 cells, Vasa and Ago3 are nuage residents, while the vast majority of Siwi is present diffused in the cytoplasm (Xiol et al., 2012). To examine where in the cell they might transiently associate, cells expressing HA-tagged Vasa proteins were counterstained for endogenous Amplifier components using specific

(B) Sequence alignment of DEAD box helicases showing critical residues for ATP binding (motif I) and ATP hydrolysis (motif II). Point mutations within *Bombyx* Vasa used in this study are indicated in the cartoon.

(C) Recovery of fluorescence after photo bleaching (FRAP) of EGFP-Vasa or EGFP-Vasa^{DQAD} granules is indicated as a function of time (in seconds, s).

(D) Recovery times for EGFP-Ago3.

(E) Time series snapshots (time in seconds, s) focusing on the granules in the cytoplasm from FRAP experiments conducted with BmN4 cells expressing indicated fluorescently labeled proteins. The bleached granule is indicated with a red circle. Note that Vasa^{DQAD} granules are generally larger but show a range of sizes, with the middle panel showing a granule where the laser photobleaches the center of the granule.

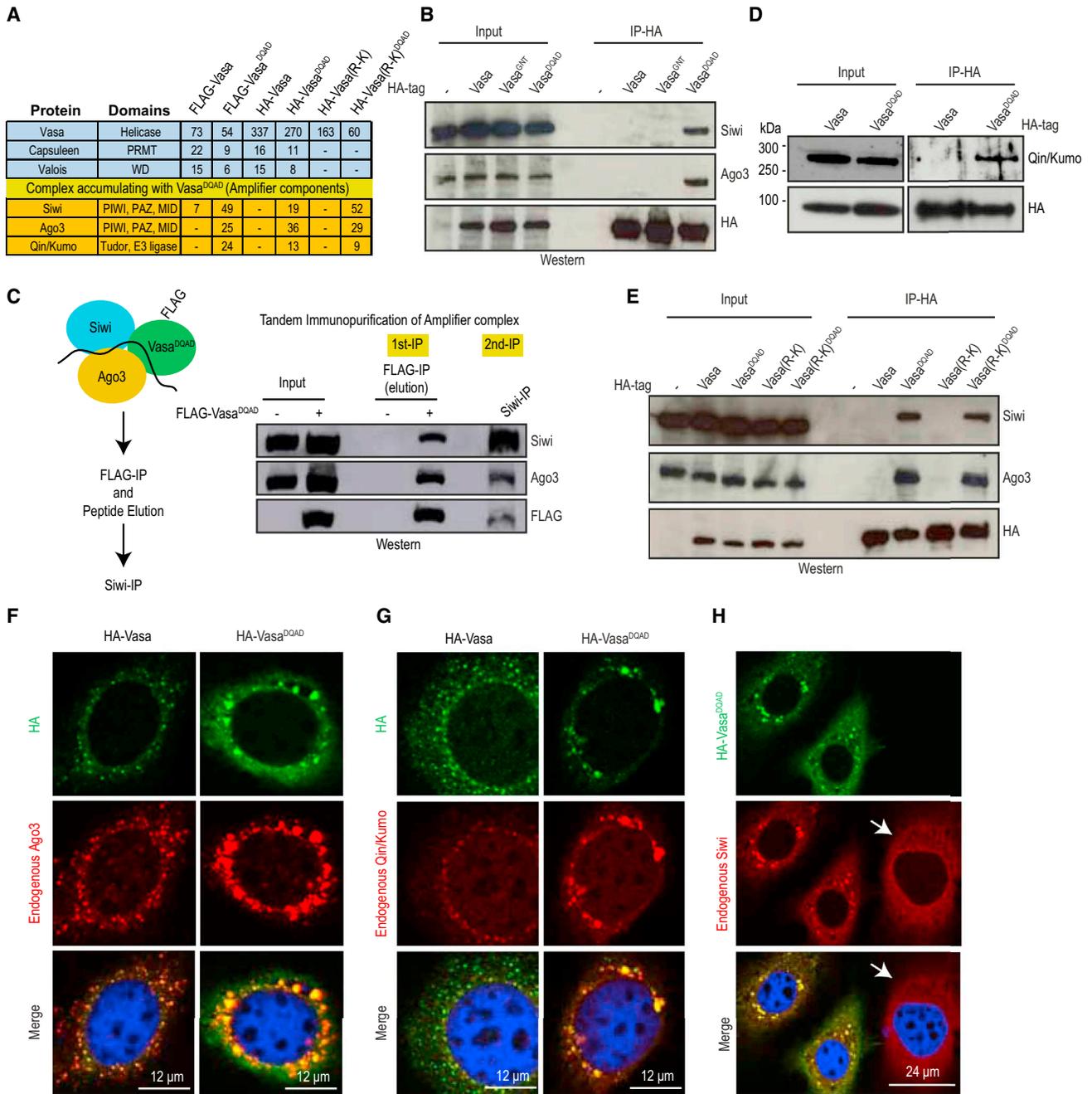


Figure 2. A Transient piRNA Amplifier Complex Containing Vasa Is Assembled in the Nuage

(A) Mass spectrometry analyses of indicated tagged protein complexes isolated from transiently transfected BmN4 cells. Note that Prmt5/Capsuleen is common to both Vasa and Vasa^{DQAD}, though the latter has a unique set of factors (see Figure S2A for a complete list), including ping-pong Piwi proteins and Qin/Kumo, which were validated by western blotting analyses. A ratio of spectral counts in the sample versus beads control is indicated.

(B) Immunoprecipitation and western blot analysis confirming specific association of endogenous Piwi proteins Siwi and Ago3 with HA-Vasa^{DQAD}.

(C) Tandem affinity purification showing presence of Vasa^{DQAD} and the ping-pong Piwi partners in the same complex.

(D) Western blot analysis showing specific association of endogenous Qin/Kumo with HA-Vasa^{DQAD}.

(E) Mutation of 20 N-terminal arginines to lysines (R-K) in Vasa^{DQAD} does not affect Amplifier assembly but abolishes recovery of Capsuleen (see A).

(F–H) Accumulation of Vasa^{DQAD} in large nuage in BmN4 cells that are also enriched in endogenous Amplifier components like Ago3 (F), Qin/Kumo (G), and Siwi (H). Note that, in untransfected cells lacking HA-Vasa^{DQAD} (white arrow in H), Siwi remains dispersed in the cytoplasm. Scale bars (in μm) are indicated.

antibodies. We find that, like endogenous Ago3 (Figure 2F), Qin/Kumo (Figure 2G) also accumulates in cytoplasmic granules that also contained HA-Vasa, whereas Siwi is spread out in the cytoplasm (Figure 2H). Interestingly, examination of the endogenous proteins in cells expressing HA-Vasa^{DQAD} shows that the enlarged nuages formed by Vasa^{DQAD} also contained a higher enrichment of Ago3 and Qin/Kumo (Figures 2F, 2G, and S2D). Furthermore, cytoplasmic Siwi was now concentrated in these perinuclear granules. The redistribution of Siwi from the cytoplasm to the granules is striking, given the almost complete absence of Siwi granules in untransfected or HA-Vasa-expressing cells (Figures 2H, S2E, and S2F). The above experiments suggest that all of the factors enter the nuage from the surrounding cytoplasm to assemble the Amplifier and then exit after execution of their functions (Figure S2G). Nevertheless, we cannot rule out the possibility that complex assembly may already take place in the cytoplasm, but its eventual disassembly (which is blocked by the DQAD mutation in Vasa) definitely takes place within the nuage.

Amplifier Contains Antisense piRNAs and Sense Transposon Targets

Given the presence of Piwi proteins in the Amplifier complex and to get an insight into its function within the nuage, we examined the presence of small RNAs in Vasa^{DQAD} complexes. To this end, we immunoprecipitated HA-tagged Vasa complexes from BmN4 cells and revealed the presence of small RNAs by 5' end labeling and 20% polyacrylamide gel electrophoresis (PAGE). Small RNAs consistent with the length of piRNAs (~30 nt) are observed in Vasa^{DQAD} purification, but not in Vasa and Vasa^{GNT} complexes (Figure 3A). As previously shown for Piwi proteins (Figures 2A and 2E), small RNA association with Vasa^{DQAD} is also independent of Vasa arginine methylation (Figure S3A). Surprisingly, another smeary RNA species of ~12 nt was also abundantly detected in Vasa^{DQAD} complexes (Figures 3A and 3B).

To uncover the identity of the short RNAs, we subjected them to deep sequencing (Figure 3B). The ~30-mer RNA species (duplicate libraries) have a read-length profile (peak at 27 nt) similar to that reported for *Bombyx* piRNAs (Figure S3B). They also display the characteristic nucleotide biases seen in piRNAs, such as preference for a uridine at the 5' end (1U-bias, ~85%) and presence of an adenosine at position 10 (A10-bias, ~40%) (Figure S3C). Based on comparison to available Siwi and Ago3 piRNA libraries (Xiol et al., 2012), ~75% of the reads in the Amplifier complex can be classified as piRNAs (hereafter referred to as Amplifier-piRNAs) (Figure 3C). Next, we wished to ascertain whether these piRNAs represent those bound to Siwi or Ago3, as this could help in determining the position of Amplifier within the ping-pong cycle. This task is complicated by the fact that the bulk of the reads is shared between Siwi and Ago3 libraries when only sequence identity is considered. However, when their relative abundance (see Extended Experimental Procedures) in a particular library is also accounted for, the reads can be sorted as belonging to either Siwi or Ago3. Based on this calculation, piRNAs from Siwi accounted for the majority in the Amplifier-piRNA pool (~75%) (Figure 3C). This is also evident when the reads are aligned on transposon consensus sequences. Amplifier-piRNA reads have a predominant antisense bias very much

like Siwi piRNAs, whereas those in Ago3 map to the sense strand (Figure 3D-E). Thus, Amplifier contains essentially Siwi-bound antisense piRNAs (Figure 3F).

Next, we prepared two independent deep-sequencing libraries for the abundantly present 12-mer species (median size of 12 nt) (Figure 3B). Mapping short reads like 12-mers can give multiple genomic hits, preventing unambiguous annotation. Therefore, we confined our search to transposon consensus sequences, allowing only perfect matches. Strikingly, the 12-mer reads predominantly mapped to the sense strand of the transposons (Figure 3E). This opposing polarity with Amplifier-piRNAs (that map to the antisense strand) is maintained over several *Bombyx* transposon sequences (Figure 3F). The short size of the 12-mers and their smeary nature suggests that these might be in vivo RNA footprints left by Vasa on bound transposon transcripts after degradation by cellular nucleases. In agreement with this, a recombinant preparation of Vasa^{DQAD} containing only the helicase core domain generates footprints (~6–10 nt) on a radioactively labeled RNA in vitro (Figure 3G). We conclude that the 12-mers are unlikely to be intermediates of the piRNA biogenesis process but might rather reflect the entire repertoire of RNA-binding sites of Vasa (Figures S3D–S3F and see discussion in Extended Experimental Procedures). To examine the presence of piRNA precursors within Vasa-bound transcripts, we extracted longer RNAs present in Amplifier complexes and prepared a strand-specific RNAseq library (fragment size of ~200 nt). Comparison to the total polyA+ transcriptome from BmN4 cells reveals an enrichment of transposon sequences (Figure 3H). Mapping to transposon consensus reveals these to be predominantly sense-oriented sequences (Figure 3I).

The ping-pong model predicts that the Piwi protein with antisense piRNAs (Siwi in BmN4 cells) facilitates biogenesis of sense-oriented Ago3-bound piRNAs from transposon transcripts (Brennecke et al., 2007). We found that Vasa gathers both Siwi and Ago3 in the nuage to assemble Amplifier, which contains mainly Siwi-bound antisense piRNAs and their target sense transposon transcripts. Considered together, the above results strongly suggest that Amplifier is the biochemical platform for generation of Ago3-bound secondary piRNAs from sense transposon precursors. Consistent with a role for RNA in complex formation, retention of both Siwi and Ago3 in the complex is sensitive to RNase treatment. However, Siwi was the most affected (Figures 3J and S3G), hinting to a requirement for piRNA-mediated base-pairing to the target for its inclusion in the complex.

Amplifier Complex Contains 5' Processed piRNA Intermediates

There could be two potential roles for Vasa within the Amplifier complex. Vasa might regulate slicer activity of Siwi so that it occurs only after the two ping-pong partners are brought together. This should ensure safe transfer of the sliced precursor fragment, i.e., the secondary piRNA intermediate, from Siwi to Ago3. Alternatively, slicing by Siwi may have already taken place, but transfer of the intermediate to Ago3 requires Vasa to complete ATP hydrolysis.

Because the repeat-rich nature of endogenous piRNAs prevents unambiguous determination of intermediates, we created an artificial secondary piRNA precursor. We identified 14 highly

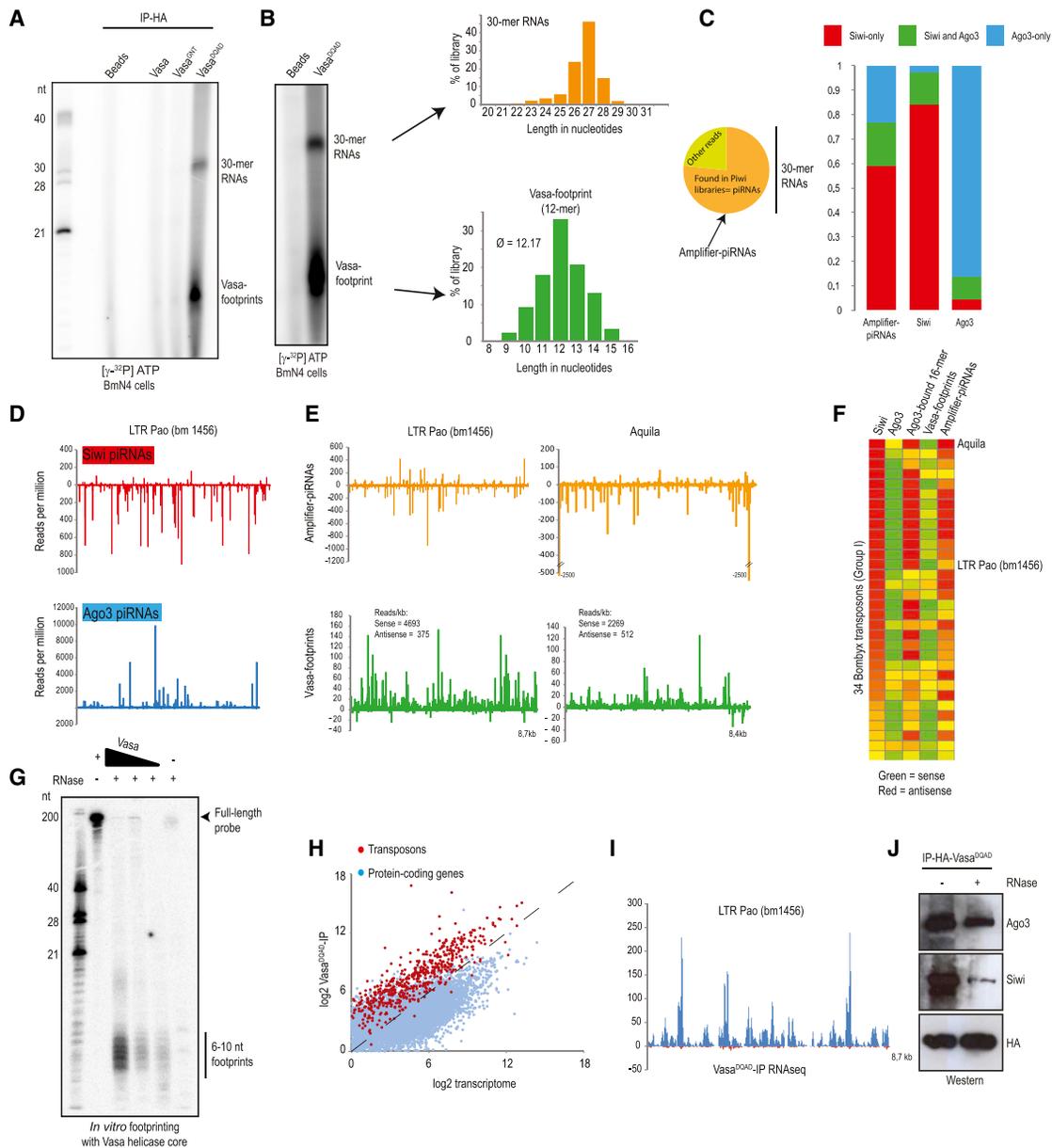


Figure 3. Amplifier Contains Siwi-Bound Antisense piRNAs and Sense Transposon Transcripts

(A) Association of Vasa^{DQAD} with short RNAs in BmN4 cells. The two prominent small RNA species—~30-mer RNAs and ~12-mer Vasa footprints—are identified.

(B) Read-length distribution of indicated RNAs.

(C) Overlap of ~30-mer RNAs with known piRNA libraries identifies ~75% of the reads as piRNAs (Amplifier-piRNAs). Sorting of piRNAs as those associating with the two Piwi proteins using sequence identity and enrichment filters (see [Extended Experimental Procedures](#)). The majority of the Amplifier-piRNAs are those associated with Siwi.

(D) Mapping of reads from indicated Piwi libraries on a transposon consensus sequence. Note the opposing polarity of Siwi- and Ago3-associated reads.

(E) Mapping of reads on two transposon consensus sequences. Note the opposing polarity of Amplifier-piRNAs and Vasa footprints.

(F) Heatmap showing strand bias of reads from Siwi, Ago3, Vasa-footprints, and Amplifier-piRNAs on 34 *Bombyx* transposon consensus sequences. Ago3-bound 16-mer refers to byproducts originating from Siwi piRNA precursors and has same strand orientation as Siwi piRNAs (Xiol et al., 2012).

(G) In vitro footprint generated by recombinant core helicase domain of Vasa^{DQAD} on a radioactively labeled RNA probe.

(H) Enrichment of transposon sequences within Amplifier (Vasa^{DQAD}-IP) complexes when compared to the general PolyA+ transcriptome of BmN4 cells.

(I) The Amplifier-bound (Vasa^{DQAD}-IP) transposon reads are mainly mapping to the sense strand of the indicated transposon consensus sequence.

(J) Presence of Piwi proteins in Amplifier (Vasa^{DQAD}) complexes is sensitive to RNase treatment. Note that Siwi is more sensitive to the treatment than Ago3. See also [Figure S3G](#).

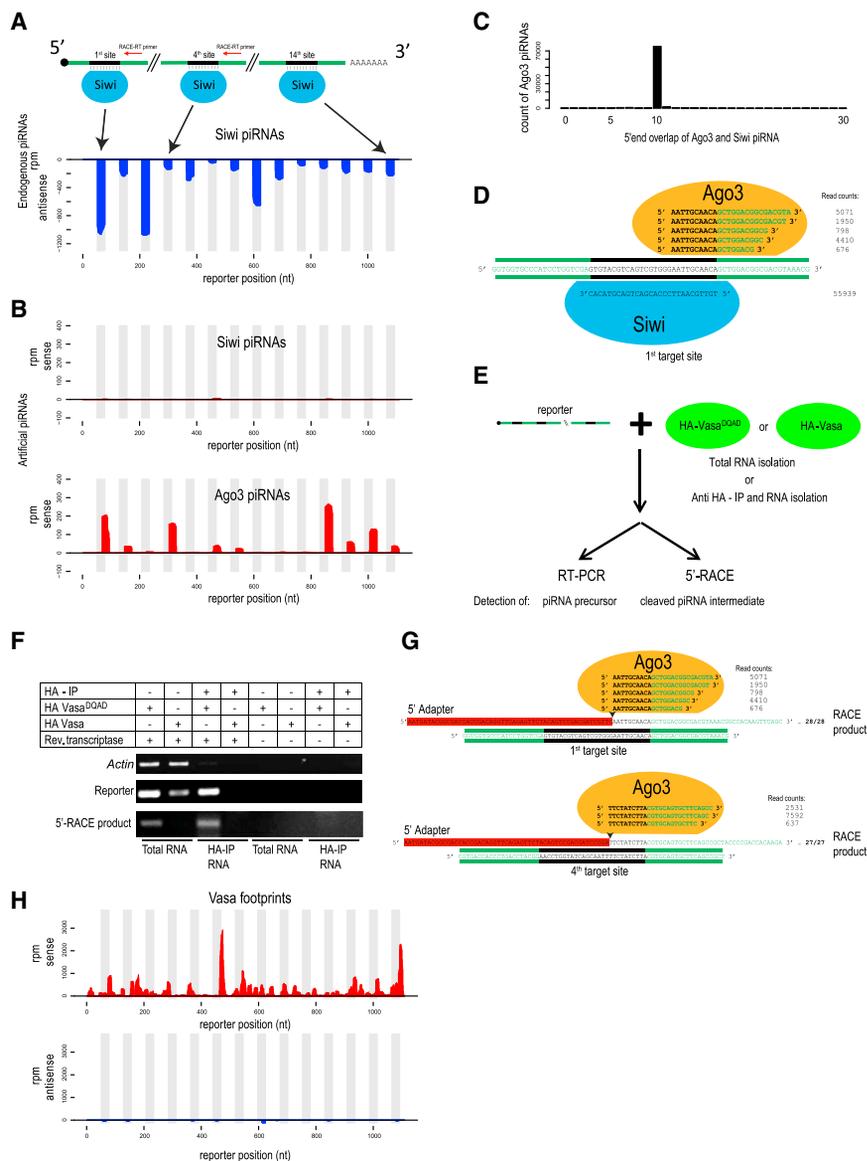


Figure 4. Amplifier Contains 5' Processed Secondary piRNA Intermediates

(A) Design of the artificial secondary piRNA precursor. The noncoding green fluorescent protein (GFP) sequence has perfectly complementary binding sites for 14 piRNAs abundantly present in endogenous Siwi complexes in BmN4 cells. Details of three such targeting sites and placement of reverse-transcription primers used for RACE experiments (RACE-RT primer) are shown. Shown below is a density plot (reads per million, rpm) of Siwi-bound piRNAs complementary to the artificial precursor. The shaded regions indicate the Siwi-piRNA-binding sites on the reporter.

(B) Endogenous Piwi proteins (Siwi and Ago3) were isolated from BmN4 cells expressing the reporter and presence of reporter-derived sequences (sense; shown in red) examined. Reporter-derived piRNAs are specifically sorted into endogenous Ago3 complexes, but not Siwi.

(C) The reporter-derived Ago3-bound piRNAs have their 5' ends overlapping exactly by 10 nt with that of Siwi-bound antisense piRNAs, indicating their origin from Siwi slicing of the reporter.

(D) Reporter-derived piRNAs are truly artificial, as they are chimeric with 10 nt of *Bombyx* repeat sequence (in black) and ~16 nt of GFP sequence (in green).

(E) Experimental setup to assay for presence of reporter RNA (by RT-PCR) or of piRNA intermediate (by 5'-RACE) in Vasa complexes or total cellular RNAs.

(F) Detection of reporter RNA or piRNA intermediate (5'-RACE product). Control reactions carried out without added reverse transcriptase is also shown.

(G) Sanger sequencing of 5' RACE products, revealing their 5' ends (marked with an arrowhead) to be exactly the same as those of mature reporter-derived Ago3-bound piRNAs (for example, in 28 of 28 sequenced clones for first target site). The sequence highlighted in red is the 5' adaptor used for cloning RACE products. The relevant portion of the reporter sequence is shown below.

(H) Mapping of Vasa footprints on reporter.

abundant Siwi piRNAs and placed perfectly complementary binding sites for these (spaced by ~50 nt) in the context of a noncoding (all ATGs removed) green fluorescent protein (GFP) sequence (Figure 4A) and expressed it in BmN4 cells. We then monitored the presence of reporter-derived piRNAs in endogenous Siwi and Ago3 complexes by deep sequencing (two independent libraries for each protein) (Figure S4A).

Mapping of reads (perfect match only) to the reporter sequence indicates the expected presence of antisense piRNAs in Siwi, and these are complementary to the inserted binding sequences (Figure 4A). Such antisense reads are also found in Ago3 libraries (Figure S4B) but with fewer read counts, again highlighting the difficulty of accurately assigning endogenous repeat piRNAs to distinct ping-pong partners. Significantly, reporter-derived (sense) piRNA sequences are found exclusively in the Ago3 libraries (Figure 4B). Such reads are nonrandomly

distributed on the sense strand and map at specific sites across from the complementary Siwi piRNAs. Confirming their origin from Siwi-guided slicing of the reporter, the 5' ends of sense reporter reads in Ago3 display an overlap of 10 nt with that of the complementary Siwi piRNAs (Figure 4C). The sense reads thus generated are truly artificial secondary piRNAs, as it is a chimera between *Bombyx* repeat sequence (1–10 nt) and the GFP backbone (11–26 nt) (Figure 4D). Thus, the reporter is a genuine precursor for Siwi-guided biogenesis of secondary piRNAs associating with Ago3. To examine the presence of the reporter within the Amplifier complex, we coexpressed the reporter and HA-tagged Vasa proteins in BmN4 cells. After anti-HA purifications, we examined the presence of the reporter sequence by RT-PCR analysis (Figure 4E). The reporter was easily detected in HA-Vasa^{DQAD} complexes, but not in HA-Vasa purifications, again reflecting the dynamic nature of the process (Figure 4F).

Next, we examined the presence of secondary piRNA processing intermediates within the Amplifier complex. Slicer cleavage by Argonautes like Siwi is expected to generate two fragments: one with a 5' monophosphate and another carrying a 3' hydroxyl (OH) group at the site of cleavage (Meister, 2013). The former (piRNA intermediate) is utilized to generate a secondary piRNA, whereas the latter is discarded. Taking advantage of the presence of a 5' phosphate on the piRNA intermediate fragment, we attempted 5' RACE experiments to detect these (Figures 4A and S4C). This revealed the specific presence of RACE products only in RNA isolated from the Vasa^{DQAD} complex (Figure 4F). Significantly, Sanger sequencing revealed that 100% of the 5' RACE products (28 out of 28 sequenced clones for first target site) contain 5' ends that are generated by Siwi slicer cleavage and match precisely to that of mature Ago3-bound reporter-derived piRNAs (Figure 4G). This strongly suggests that Amplifier contains piRNA intermediates with an already defined 5' end that are awaiting further 3' processing after loading into Ago3. Identical results were obtained in RACE experiments that were carried out for detection of another intermediate (at target site 4) (Figure 4G) in independent transfection experiments. Interestingly, 5' processed piRNA intermediates (RACE products) are also detected in total RNA from cells expressing HA-Vasa^{DQAD}, but not from those expressing HA-Vasa (Figure 4F), even though they are a necessary intermediate for producing mature secondary piRNAs found in Ago3 (Figure 4B). This is likely explained by the fact that, in the latter situation, piRNA intermediates are transferred to Ago3 and are rapidly processed, whereas in the case of cells expressing HA-Vasa^{DQAD}, these are trapped in the Amplifier complex, allowing its efficient detection even in total cellular RNAs. Finally, mapping of the Vasa footprints on the reporter sequence reveals peaks that cluster around the ping-pong sites (Figures 4H and S4D). This indicates that Vasa physically binds to the secondary piRNA intermediate for safely transferring it to Ago3 for maturation. Thus, although target slicing by Siwi has already taken place in the snapshot of Amplifier that we have frozen, handover of the cleavage fragment to Ago3 and maturation of the fragment as a new secondary piRNA fail to occur in the absence of ATP hydrolysis/disassembly of the complex by Vasa.

The DQAD Mutation in Vasa Prevents Release of ATP Hydrolysis Products

Our cell culture experiments point to an important role for Vasa's ATPase cycle in regulating Amplifier assembly, as we were able to trap this complex only by using a mutation that is predicted to abolish its ATPase activity. We therefore took an in vitro approach to gain molecular insight into this process. We prepared recombinant versions of *Bombyx* Vasa containing only the helicase core domain (135–564 aa) (Figure S5A). As expected from studies of other DEAD box helicases (Pause and Sonenberg, 1992), the Vasa^{GNT} mutant failed to bind ATP, whereas both Vasa and Vasa^{DQAD} showed binding in an UV crosslinking assay (Figure 5A). We note that Vasa^{DQAD} consistently gave stronger signals in the ATP-binding assays. Vasa and Vasa^{DQAD} also displayed ATP-stimulated binding activity to a poly-U RNA (Figure S5B). The DQAD mutation in eIF4A, a prototypic member of the DEAD box helicase family, is reported to abolish its

ATPase activity (Pause and Sonenberg, 1992), so we were surprised to find that Vasa^{DQAD} still retained catalytic activity, as assayed in our low-substrate conditions (Figure 5B). We conclude that Vasa^{DQAD} behaves very much like the wild-type protein in the in vitro experiments, leaving us with no explanation for the observed in vivo effects. This prompted us to take a closer look at the helicase core by X-ray crystallography (Figure S5C and Extended Experimental Procedures).

We were able to determine the structure of *Bombyx* Vasa^{DQAD} helicase core domain in complex with a 6 nt RNA and ATP at 2.1 Å resolution. The structure of the wild-type version of the *Drosophila* ortholog had been previously determined in the presence of a nonhydrolyzable ATP analog AMPPNP (Sengoku et al., 2006), likely reflecting the need to freeze the protein in an ATP-bound state in order to perform crystallographic studies. The BmVasa^{DQAD}-RNA-ATP structure revealed the two RecA-like domains in a closed conformation, with the ATP-binding site created in the cleft between them. The characteristically bent RNA contacts both the RecA-like domains and is held in a position opposite to the ATP-binding pocket (Figure 5C). The overall architecture is similar to that found in wild-type *Drosophila* Vasa (Figure S5D) (Sengoku et al., 2006), and our second *Bombyx* Vasa^{DQAD} structure with AMPPNP shows that the catalytic water is correctly positioned prior to ATP hydrolysis (Figures S5E and S5F).

Surprisingly, in the structure obtained from crystals containing added ATP, the ligand was hydrolyzed and the products (ADP and γ -phosphate [Pi]) were retained in the protein (Figures 5C and S5G). ATP hydrolysis products are normally released from ATPases and are not observable within RNA helicases, as they promote recycling of the enzyme (Liu et al., 2008). Our structure now provides the basis for retention of the hydrolysis products within Vasa^{DQAD}. First, the mutation (E339Q) converts the negatively charged side chain of glutamate (E) to a neutral charge in glutamine (Q), removing a repulsive force that would normally be used to eject the negatively charged γ -phosphate. Second, the free phosphate is held in place by hydrogen bonding with a number of residues, with the mutated residue Q339 in Vasa^{DQAD} providing an additional interaction that is not available in the native context (Sengoku et al., 2006) (Figure 5C). Failure to release hydrolysis products can prevent turnover of the enzyme (Liu et al., 2008), perhaps explaining the increased crosslinking of ATP to the protein (Figure 5A) and reduced ATPase activity under high substrate concentrations (Figure S5H). Thus, we conclude that Vasa^{DQAD} is a product-release-trap mutant.

ATP-Gated RNA Clamping by Vasa Allows Amplifier Assembly on Its Helicase Core Domain

Next, we asked how failure to release ATP hydrolysis products by Vasa contributes to the observed in vivo effects in BmN4 cells. DEAD box helicases are ATP-gated RNA-binding proteins; ATP binding promotes the closed conformation of the two RecA-like domains, and this translates into a tight binding to the RNA, whereas ATP hydrolysis releases this grip (Linder and Jankowsky, 2011). The fact that RNA intermediates of the ping-pong cycle are detected in Amplifier (Figure 4F) prompted us to look at the dynamics of RNA binding by Vasa. We used surface plasmon resonance (SPR) experiments to monitor the dynamics of Vasa affinity for single-stranded RNA in the presence of ATP

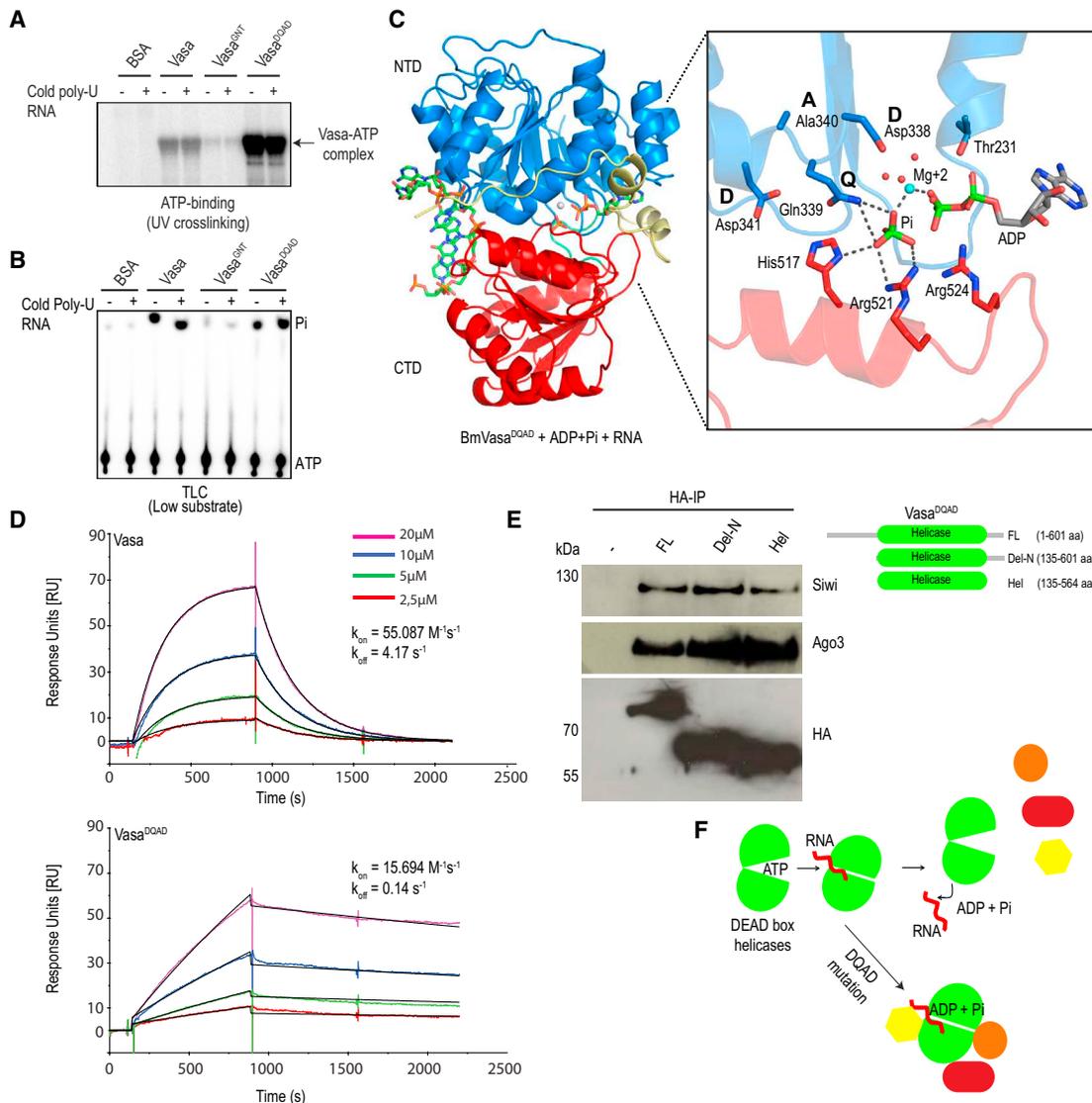


Figure 5. The ATP-Bound Closed Conformation of Vasa's Core Helicase Domain Provides a Platform for Amplifier Assembly

(A) UV crosslinking assay to detect ATP binding with recombinant *Bombyx* Vasa helicase domains carrying the indicated mutations. Bovine serum albumin (BSA) serves as a negative control. Reactions were carried out in the presence (+) or absence (–) of unlabeled (cold) poly-U RNA.

(B) ATP hydrolysis assays to reveal generation of inorganic phosphate (Pi). Low substrate; indicates that reactions contained only radioactive ³²P-γ-ATP, without added cold ATP. See also Figure S5H.

(C) Crystal structure of *Bombyx mori* (Bm) Vasa^{DQAD}-RNA-ADP-P_i complex. The two recA-like domains (blue and red) of the helicase module are represented as ribbons, whereas RNA and ATP are indicated as ball-and-sticks. ATP added during crystallization with BmVasa^{DQAD} becomes hydrolyzed into ADP and Pi. Interaction with glutamine (Q339) traps Pi in the protein, creating a product-release mutant. Magnesium ion (Mg²⁺) (blue ball) and water molecules (red balls) are indicated.

(D) Surface plasmon resonance (SPR) experiments indicate that both proteins have similar RNA-binding kinetics (k_{on}), but Vasa^{DQAD} shows drastically slower off rates (k_{off}). Interactions were examined at different protein concentrations (different line colors), and binding to RNA is reported in arbitrary response units (RU) as a function of time (in seconds, s).

(E) Deletion constructs of HA-Vasa^{DQAD} were transfected into BmN4 cells and association of endogenous Piwi proteins examined by western blotting.

(F) A model showing consequence of DQAD mutation for DEAD box proteins. The mutation traps helicases in the closed conformation, freezing bound RNA and protein complexes.

(Figure 5D). Both Vasa and Vasa^{DQAD} bound RNA with rather similar kinetics (K_{on}). However, the off rate for Vasa^{DQAD} ($K_{off} = 0.14 \text{ s}^{-1}$) was ~40-fold lower than that of Vasa wild-type ($K_{off} = 4.17 \text{ s}^{-1}$), indicating very tight binding to RNA. This indicates

that the inability to release ATP hydrolysis products and the consequential enforcement of the closed conformation of Vasa^{DQAD} traps it on bound RNAs, which we identified as sense piRNA precursors (Figure 3H). Interestingly, our biochemical

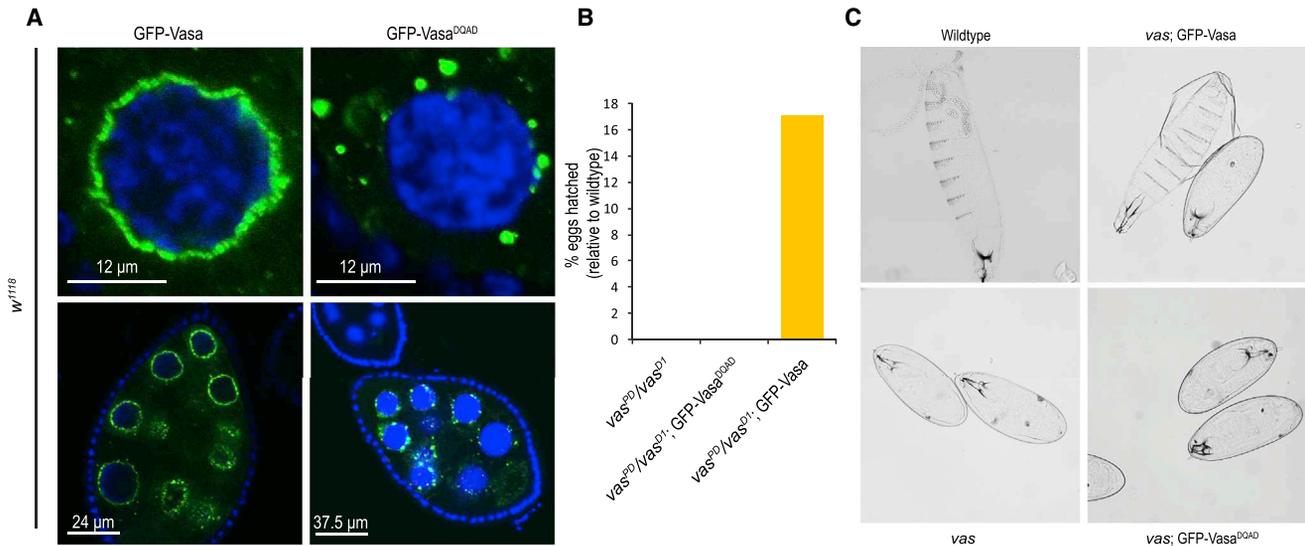


Figure 6. ATP Hydrolysis by Vasa Is Essential for Fertility in *Drosophila* Females

(A) Fluorescence detection of GFP fusion proteins in transgenic fly ovaries in the wild-type genetic background. GFP-DmVasa^{DQAD} accumulates in granules that are removed from the regular perinuclear ring-like nuage localization, potentially indicating an impact on the in vivo dynamics of the protein.

(B) Eggs laid by the *vas* mutant fail to hatch, and this is rescued by a GFP-Vas transgene, but not by the GFP-DmVasa^{DQAD} transgene. Number of eggs hatched is indicated as a percentage of that seen in the wild-type. The partial rescue (~16%) is explained by the fact that the transgenes are expressed from a nonnative promoter.

(C) Cuticle staining of such hatched embryos shows the segmentation pattern expected of normal development (in wild-type and *vas* mutants complemented with GFP-Vas).

experiments (Figures 2A–2E) in BmN4 cells suggest that such an ATP-bound and RNA-clamped state of Vasa also provides a binding platform for Amplifier assembly.

Vasa has a rather simple architecture with an N-terminal arginine-rich region (R-rich domain), followed by the canonical DEAD box helicase core domain (Figure 1B). To understand how Vasa can mediate Amplifier assembly, we prepared deletion constructs and examined their ability to support association with other factors in vivo. Transiently expressed deletion constructs were purified by anti-HA immunoprecipitations and probed with antibodies to detect endogenous Piwi proteins. Deletion of the R-rich domain (1–134 aa) did not affect association with Siwi and Ago3 (Figure 5E). This adds to our observation that N-terminal arginines in Vasa are not required for mediating complex formation (Figures 2A and 2E). Remarkably, a construct having only the helicase core domain was capable of associating with Siwi and Ago3 (Figure 5E), indicating that Amplifier is assembled on the surface of the helicase core of Vasa. We propose that the binding interface for Amplifier components is created only when the helicase domain is in a closed conformation and, consequently, only when Vasa is bound to its target RNAs (Figure 5F). Subsequent ATP hydrolysis would promote release of the RNA and would result in an open conformation of the helicase domain, disrupting the protein interaction surface and leading to complex disassembly.

ATP Hydrolysis by Vasa Is Essential for Fertility in the *Drosophila* Female Germline

Impaired piRNA pathway leads to infertility in all animal models, including in *Drosophila vasa* (*vas*) mutants (Malone et al., 2009).

Because the DQAD mutation within *Bombyx* Vasa prevents Amplifier disassembly, we wished to examine whether such a mutation within *Drosophila* Vasa might have a functional consequence for fly fertility. We expressed *Drosophila* Vasa transgenes (GFP-Vas or GFP-Vas^{DQAD}) in wild-type *Drosophila* ovaries (Figure S6A). Whereas GFP-Vas localized as a perinuclear ring to the nuage of nurse cells (Lasko and Ashburner, 1988), GFP-Vas^{DQAD} protein appeared to form distinct foci that are removed from the perinuclear location (Figure 6A). These are reminiscent of enlarged granules observed in BmN4 cells (Figures 1E, 2F, and 2G). Next, we expressed the transgenes in the *vas*^{P0}/*vas*^{D1} mutant background (these mutants lay eggs that fail to hatch) (Lasko and Ashburner, 1988). When GFP-Vas is expressed in the mutant background, it complements infertility (Figures 6B, S6A, and S6B). The rescued embryos showed normal segmentation patterns, very similar to those hatched from eggs laid by wild-type females (Figure 6C). Significantly, the GFP-Vasa^{DQAD} transgene failed to rescue the infertility phenotype of the mutant (Figures 6B and 6C and Extended Experimental Procedures). These experiments provide strong genetic evidence for the physiological effect of preventing disassembly of Vasa RNPs in the *Drosophila* germline.

DISCUSSION

Secondary piRNA biogenesis (the ping-pong cycle) links post-transcriptional silencing of transposons by Piwi endonucleases to piRNA biogenesis by converting one of the cleavage fragments into a new secondary piRNA (Brennecke et al., 2007; Gunawardane et al., 2007). This poses a unique problem, as

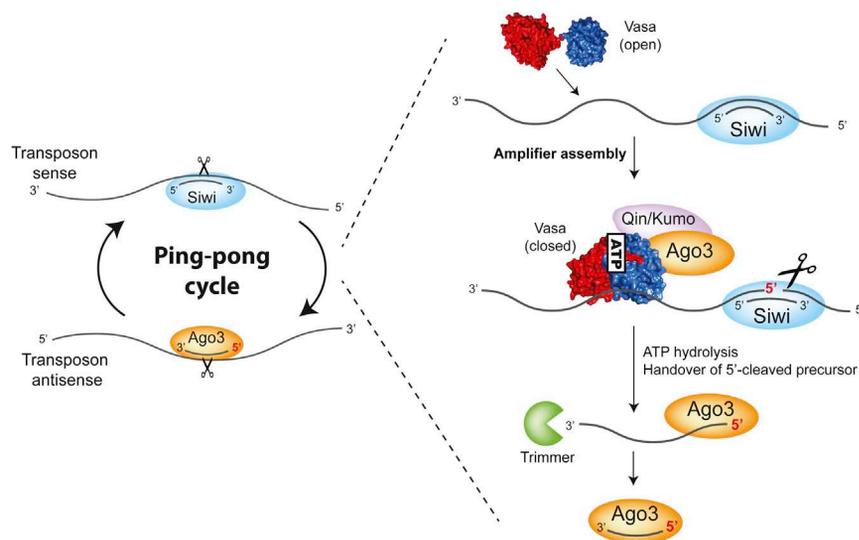


Figure 7. A Model for Vasa's Role in the Ping-Pong Cycle

Recognition of transposon transcripts by Siwi-bound antisense piRNAs is proposed to result in recruitment of Vasa. The open conformation of Vasa's helicase domain is modeled on the eIF4A structure (PDB: 1FUU). Vasa binds RNA in its ATP-bound form, taking up a closed conformation of its RNA helicase domain. This closed conformation also provides an assembly platform for Amplifier components (ping-pong Piwi partners and the Tudor domain protein Qin/Kumo). After Siwi slicing of the transposon transcript, the cleavage fragment carrying the 5' monophosphate (piRNA intermediate) is transferred to Ago3. ATP hydrolysis in Vasa triggers this exchange and eventual disassembly of the Amplifier. Subsequently, the piRNA intermediate is subject to 3' end maturation by a putative exonuclease (Trimmer) to mature as a new secondary piRNA. Loaded Ago3 then enters the feedforward step of the ping-pong cycle to generate more of the Siwi-bound piRNAs.

endonucleolytic cleavage of a target RNA by an Argonaute normally results in complete degradation of the two resulting cleavage fragments in all cell types. Germ cells must therefore possess an unknown machinery that safely delivers a cleavage fragment (piRNA intermediate) from one Piwi protein to another (ping-pong partners), but the molecular basis for this process is not known. Among the list of over 30 factors that are genetically linked to the piRNA pathway (Figure S1A), we chose to investigate Vasa, as it is the only factor that is demonstrated to have a conserved and exclusive role in secondary piRNA biogenesis in flies and mice (Kuramochi-Miyagawa et al., 2010; Malone et al., 2009). Our investigations reveal that Vasa functions as an RNA clamp to coordinate the assembly of a piRNA Amplifier complex on transposon transcripts. Within this protected environment, ATP hydrolysis by Vasa promotes safe handover of sliced piRNA intermediates between ping-pong partners.

Our results allow us to propose a step-by-step biochemical model for the Ping-pong cycle operating in BmN4 cells (Figure 7). Transposon silencing is initiated by antisense primary piRNAs loaded into Siwi. This is closely followed by deposition of Vasa on the transcript and by assembly of the Amplifier by the joining of Qin/Kumo and Ago3 within the nuage. Amplifier assembly is linked to a conformational change in the core helicase domain of Vasa as it transitions from an ATP-unbound open state to the ATP- and RNA-bound closed state, such that the latter provides an interaction platform for the different components. Interestingly, Qin is already reported to facilitate heterotypic ping-pong between Aub (Siwi ortholog in *Drosophila*) and Ago3 by promoting interactions between them in the fly germline (Anand and Kai, 2012; Zhang et al., 2011). Siwi initiates the ping-pong cycle by slicing the target RNA to generate two fragments. Following this, the cleavage fragment carrying the mature 5' end of a future secondary piRNA (piRNA intermediate) is transferred from Siwi to Ago3. We show that ATP hydrolysis by Vasa is critical for triggering this exchange and subsequent disassembly

of the complex. We propose that the role of Vasa in assembling Amplifier is to generate a protective environment in which the ping-pong partners are brought to physical proximity, such that the cleaved target is not accessible to nucleases and is instead loaded onto Ago3 upon release from Siwi. Once acquired by Ago3, the mature 3' end of the new piRNA is generated by an unknown 3'-5' exonuclease, tentatively called Trimmer (Kawaoka et al., 2011). Such sense-oriented secondary piRNAs then guide Ago3 to participate in the feedforward step of the ping-pong cycle (Brennecke et al., 2007).

Our results provide, for the first time, a biochemical snapshot of events that take place during the ping-pong cycle. However, a number of questions still remain. It is not clear how the closed conformation of Vasa's helicase core domain can promote Amplifier assembly. In this context, it is tempting to draw a parallel between Vasa's role and the function of the helicase domain of the related DEAD box helicase eIF4AIII in nucleating the formation of the exon junction complex (EJC) on spliced mRNAs (Le Hir et al., 2000). In its ATP-bound closed conformation, the eIF4AIII core domain clamps on spliced mRNAs to provide a binding surface for other EJC components Mago, Y14, and Barentsz (Andersen et al., 2006; Bono et al., 2006). A similar situation can be envisaged during Amplifier assembly. Also unknown is whether any control is exercised over Siwi slicer activity to prevent futile cleavages that do not result in secondary piRNA production. Because piRNA intermediate exchange and complex disassembly are linked to ATP hydrolysis by Vasa, this might also be subject to regulation. It is only logical that target slicing by Siwi is followed by ATP hydrolysis by Vasa. We can only speculate that perhaps allosteric interactions between Piwi proteins and Vasa might link slicing of the target to ATP hydrolysis in Vasa. Recent genome-wide genetic screens have identified a number of factors that are implicated in the germline piRNA pathway in flies (Czech et al., 2013). An examination of our top hits in the Vasa^{DQAD} proteomics analyses (Figure S2A) did not reveal any overlaps except for the ping-pong Piwi partners and Qin/Kumo, which we validated in our study.

The ping-pong model predicts a reciprocal relationship in which piRNA-guided Siwi facilitates biogenesis of Ago3-bound piRNAs and vice versa (Brennecke et al., 2007). This means that there are potentially two distinct complexes mediating piRNA amplification: one occupying the Siwi-to-Ago3 segment and another one mediating the Ago3-to-Siwi feedforward step of the amplification loop. Our data strongly support a role for Vasa in the former step of the ping-pong cycle, but we cannot entirely rule out its participation in the latter. It is also possible that such a role in the feedforward segment might be fulfilled by other RNA helicases like Spn-E, a factor that is essential for the ping-pong cycle in flies (Malone et al., 2009), but not in mice (Shoji et al., 2009), wherein a linear secondary piRNA biogenesis pathway operates (De Fazio et al., 2011).

The DEAD box family is the largest family of helicases, and they are found from bacteria to humans, with most of the 26 members in yeast being essential for viability (de la Cruz et al., 1999). The in vivo RNA targets and RNP complexes of most DEAD box helicases have not been identified due to their transient nature. Here, we used a DQAD mutation to trap dynamic in vivo associations of Vasa. Structures of several DEAD box helicases with bound ATP analogs and RNA are closely related, indicating that perhaps similar outcomes to those we describe here can be expected for other family members assembling dynamic complexes in various RNA processing pathways.

EXPERIMENTAL PROCEDURES

Clones and Constructs

For expression studies in BmN4 cells, the coding sequence for the full-length *Bombyx mori* Vasa (GenBank accession number: NM_001043882.1) or its deletion versions were cloned into the pBEMBL-HA (Xiol et al., 2012) or pBEMBL-Myc vectors. Point mutations in the BmVasa helicase domain were introduced by an overlap PCR strategy: ATP-binding mutant (K230N; GKT → GNT) or ATP hydrolysis product-trap mutant (E339Q; DEAD → DQAD). The R → K mutant was created by converting 20 arginine (R) residues to lysines (K) at the Vasa N terminus (within a region spanning 1–117 aa) using a chemically synthesized DNA fragment. For live-cell imaging studies, the coding sequence for BmVasa, Vasa^{DQAD}, or Ago3 was cloned downstream of an enhanced green fluorescent protein (EGFP) tag or the red fluorescent protein variant mCherry in the pBEMBL vector backbone.

Crystallization and Biochemical Procedures

The helicase domain of *Bombyx* Vasa (135–564 aa) carrying the E339Q (DEAD → DQAD) mutation was produced in *E. coli* as a His-tagged fusion. Purified protein was incubated with a 6-mer (UGACAU) RNA and ATP or the ATP analog (AMPPNP) at a molar ratio of 1:1.2:1.2 (protein:RNA:ATP/AMPPNP). A summary of data collection statistics is provided in the [Extended Experimental Procedures](#) as Table S1.

Small RNA Sequencing

RNA libraries were prepared from immunoprecipitated complexes and were deep sequenced on the Illumina Hi-Seq (50 cycles) or MiSeq platforms (32 cycles). For long RNAs present in Vasa^{DQAD} complexes, a strand-specific RNA-seq library was prepared and sequenced on an Illumina Hi-Seq platform (105 cycles). Reads were aligned to the *Bombyx* genome and were analyzed using custom pipelines.

Fly Experiments

Drosophila vas mutant stocks were: vas^{PD} (Schupbach and Wieschaus, 1986) and vas^{D1} (Lehmann and Nüsslein-Volhard, 1991). The vas-deficient background was achieved by the use of the transheterozygotes (vas^{PD}/vas^{D1}). Expression of GFP-tagged *Drosophila melanogaster* Vasa transgenes

(Johnstone and Lasko, 2004) in the fly germline was controlled from the upstream activating sequence (UAS) using GAL4 driven from a *nanos* promoter.

ACCESSION NUMBERS

The PDB accession codes for crystal structures are 4D26 and 4D25. Deep-sequencing data are deposited with Gene Expression Omnibus (GEO) (GSE57420).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, five movies, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.05.018>.

ACKNOWLEDGMENTS

We thank Paul Lasko, Hugo Bellen, and Beat Suter for fly stocks/reagents. We are grateful to Sanjay Ghosh, Andrea Picco, Enrico Zamattia, Sandra Müller, and Anna Cyrklaff for help with data analysis and experiments. We thank the following EMBL core facilities: Advanced Light Microscopy, Genomics, Protein Expression and Purification, and High-Throughput Crystallization. Crystallographic data collection was at European Synchrotron Radiation Facility (ESRF), Grenoble, France. This work was supported by grants from European Union (European Research Council; “persistence”) to R.S.P. We are grateful for fellowships from the EMBL Interdisciplinary Postdoc Programme (EIPOD) under Marie Curie COFUND Actions (M.A.L), EMBO (D.H), and Region Rhone Alp (E.C.). This work was supported by the EMBL.

Received: February 25, 2014

Revised: April 28, 2014

Accepted: May 15, 2014

Published: June 5, 2014

REFERENCES

- Anand, A., and Kai, T. (2012). The tudor domain protein kumo is required to assemble the nuage and to generate germline piRNAs in *Drosophila*. *EMBO J.* 31, 870–882.
- Andersen, C.B., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L., Pedersen, J.S., Séraphin, B., Le Hir, H., and Andersen, G.R. (2006). Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* 313, 1968–1972.
- 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.
- Bagijn, M.P., Goldstein, L.D., Sapetschnig, A., Weick, E.M., Bouasker, S., Lehrbach, N.J., Simard, M.J., and Miska, E.A. (2012). Function, targets, and evolution of *Caenorhabditis elegans* piRNAs. *Science* 337, 574–578.
- Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* 126, 713–725.
- 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.
- Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322, 1387–1392.
- Caruthers, J.M., and McKay, D.B. (2002). Helicase structure and mechanism. *Curr. Opin. Struct. Biol.* 12, 123–133.
- Chen, Y., Pane, A., and Schupbach, T. (2007). Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in *Drosophila*. *Curr. Biol.* 17, 637–642.

- Czech, B., Preall, J.B., McGinn, J., and Hannon, G.J. (2013). A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Mol. Cell* *50*, 749–761.
- De Fazio, S., Bartonicek, N., Di Giacomo, M., Abreu-Goodger, C., Sankar, A., Funaya, C., Antony, C., Moreira, P.N., Enright, A.J., and O'Carroll, D. (2011). The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* *480*, 259–263.
- de la Cruz, J., Kressler, D., and Linder, P. (1999). Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. *Trends Biochem. Sci.* *24*, 192–198.
- de Vanssay, A., Bouge, A.L., Boivin, A., Hermant, C., Teyssset, L., Delmarre, V., Antoniewski, C., and Ronsseray, S. (2012). Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. *Nature* *490*, 112–115.
- Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* *10*, 94–108.
- Greutzinger, T., Armenise, C., Brun, C., Mugat, B., Serrano, V., Pelisson, A., and Chambeyron, S. (2012). piRNA-mediated transgenerational inheritance of an acquired trait. *Genome Res.* *22*, 1877–1888.
- 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.
- Johnstone, O., and Lasko, P. (2004). Interaction with eIF5B is essential for Vasa function during development. *Development* *131*, 4167–4178.
- Kawaoka, S., Hayashi, N., Suzuki, Y., Abe, H., Sugano, S., Tomari, Y., Shimada, T., and Katsuma, S. (2009). The *Bombyx* ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. *RNA* *15*, 1258–1264.
- Kawaoka, S., Izumi, N., Katsuma, S., and Tomari, Y. (2011). 3' end formation of PIWI-interacting RNAs in vitro. *Mol. Cell* *43*, 1015–1022.
- Kawaoka, S., Hara, K., Shoji, K., Kobayashi, M., Shimada, T., Sugano, S., Tomari, Y., Suzuki, Y., and Katsuma, S. (2013). The comprehensive epigenome map of piRNA clusters. *Nucleic Acids Res.* *41*, 1581–1590.
- Kazazian, H.H., Jr. (2004). Mobile elements: drivers of genome evolution. *Science* *303*, 1626–1632.
- Kirino, Y., Vourekas, A., Kim, N., de Lima Alves, F., Rappsilber, J., Klein, P.S., Jongens, T.A., and Mourelatos, Z. (2010). Arginine methylation of vasa protein is conserved across phyla. *J. Biol. Chem.* *285*, 8148–8154.
- Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A., et al. (2009). The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* *138*, 1137–1149.
- Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Takamatsu, K., Chuma, S., Kojima-Kita, K., Shiromoto, Y., Asada, N., Toyoda, A., Fujiyama, A., et al. (2010). MVH in piRNA processing and gene silencing of retrotransposons. *Genes Dev.* *24*, 887–892.
- Lasko, P.F., and Ashburner, M. (1988). The product of the *Drosophila* gene vasa is very similar to eukaryotic initiation factor-4A. *Nature* *335*, 611–617.
- Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000). The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.* *19*, 6860–6869.
- Lehmann, R., and Nüsslein-Volhard, C. (1991). The maternal gene nanos has a central role in posterior pattern formation of the *Drosophila* embryo. *Development* *112*, 679–691.
- Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., et al. (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* *137*, 509–521.
- Liang, L., Diehl-Jones, W., and 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, A.K., and Kai, T. (2007). Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* *104*, 6714–6719.
- Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* *12*, 505–516.
- Liu, F., Putnam, A., and Jankowsky, E. (2008). ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc. Natl. Acad. Sci. USA* *105*, 20209–20214.
- Luteijn, M.J., and Ketting, R.F. (2013). PIWI-interacting RNAs: from generation to transgenerational epigenetics. *Nat. Rev. Genet.* *14*, 523–534.
- Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* *137*, 522–535.
- Mathioudakis, N., Palencia, A., Kadlec, J., Round, A., Tripsianes, K., Sattler, M., Pillai, R.S., and Cusack, S. (2012). The multiple Tudor domain-containing protein TDRD1 is a molecular scaffold for mouse Piwi proteins and piRNA biogenesis factors. *RNA* *18*, 2056–2072.
- Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.* *14*, 447–459.
- Pane, A., Wehr, K., and Schüpbach, T. (2007). zucchini and squash encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Dev. Cell* *12*, 851–862.
- Pause, A., and Sonenberg, N. (1992). Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J.* *11*, 2643–2654.
- Schupbach, T., and Wieschaus, E. (1986). Germline autonomy of maternal-effect mutations altering the embryonic body pattern of *Drosophila*. *Dev. Biol.* *113*, 443–448.
- Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., and Yokoyama, S. (2006). Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. *Cell* *125*, 287–300.
- Shoji, M., Tanaka, T., Hosokawa, M., Reuter, M., Stark, A., Kato, Y., Kondoh, G., Okawa, K., Chujo, T., Suzuki, T., et al. (2009). The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Dev. Cell* *17*, 775–787.
- Xiol, J., Cora, E., Kogelgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z., Berninger, P., Palencia, A., et al. (2012). A role for Fkbp6 and the chaperone machinery in piRNA amplification and transposon silencing. *Mol. Cell* *47*, 970–979.
- Zhang, Z., Xu, J., Koppetsch, B.S., Wang, J., Tipping, C., Ma, S., Weng, Z., Theurkauf, W.E., and Zamore, P.D. (2011). Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. *Mol. Cell* *44*, 572–584.
- Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B.S., Schultz, N., Vreven, T., Meignin, C., Davis, I., Zamore, P.D., et al. (2012). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell* *151*, 871–884.

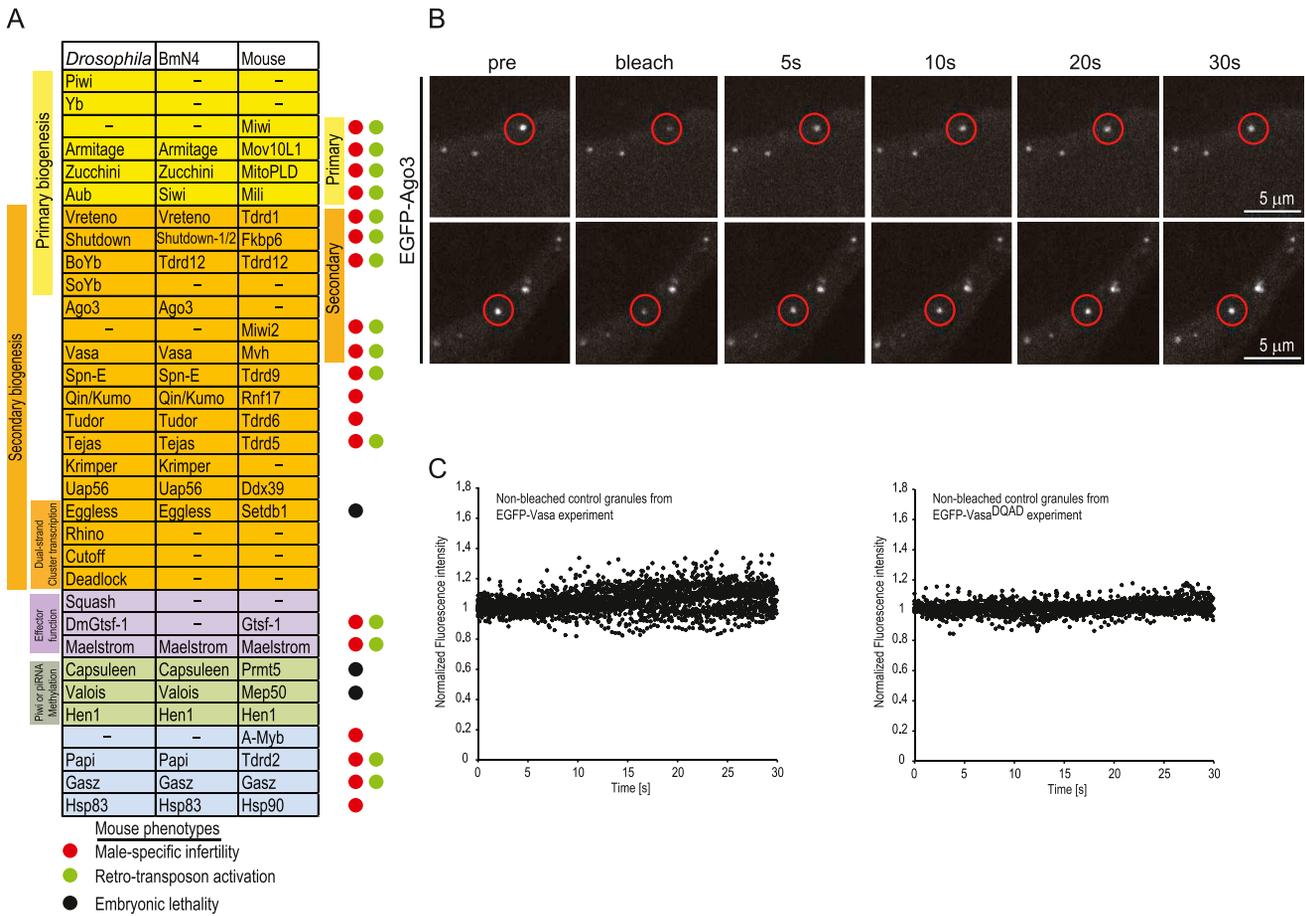


Figure S1. Components of the BmN4 piRNA Pathway and In Vivo Dynamics, Related to Figure 1

(A) A table showing factors genetically implicated in the piRNA pathway in either flies or mice. Those shown to impact either primary or secondary piRNA biogenesis in *Drosophila* or the mouse systems are indicated. Some factors in *Drosophila* are shown to affect both primary and secondary biogenesis in the ovarian germline. The orthologs of *Drosophila* genes detected in the BmN4 cell line transcriptome are shown. Importantly, fly Vasa or the Mouse Vasa Homolog (Mvh) is the only conserved factor with an exclusive role in secondary biogenesis. Studies in *Drosophila* allows grouping of some of the other factors into functional classes as indicated in boxes. Phenotypes of mouse mutants are indicated as filled circles: red, male-specific infertility, green, activation of retrotransposons and black, embryonic lethality.

(B) FRAP experiments reveal that EGFP-Ago3 dynamically enters and exits nuage in BmN4 cells. Snapshots of time-series [in seconds (s)] before and after bleaching are shown (for experiment in Figure 1D). The bleached granule is circled in red.

(C) Normalized fluorescence intensity of neighboring control non-bleached granules from the same experiment as shown in Figure 1C. The aim was to ensure that movement of the granules in and out of the focal plane did not account for fluorescence fluctuation seen in Figure 1C. Detailed explanation can be found in the [Extended Experimental Procedures](#). Prebleach phase was averaged and set to 1. Time (in seconds, s) of bleach stimulus is set to 0 s.

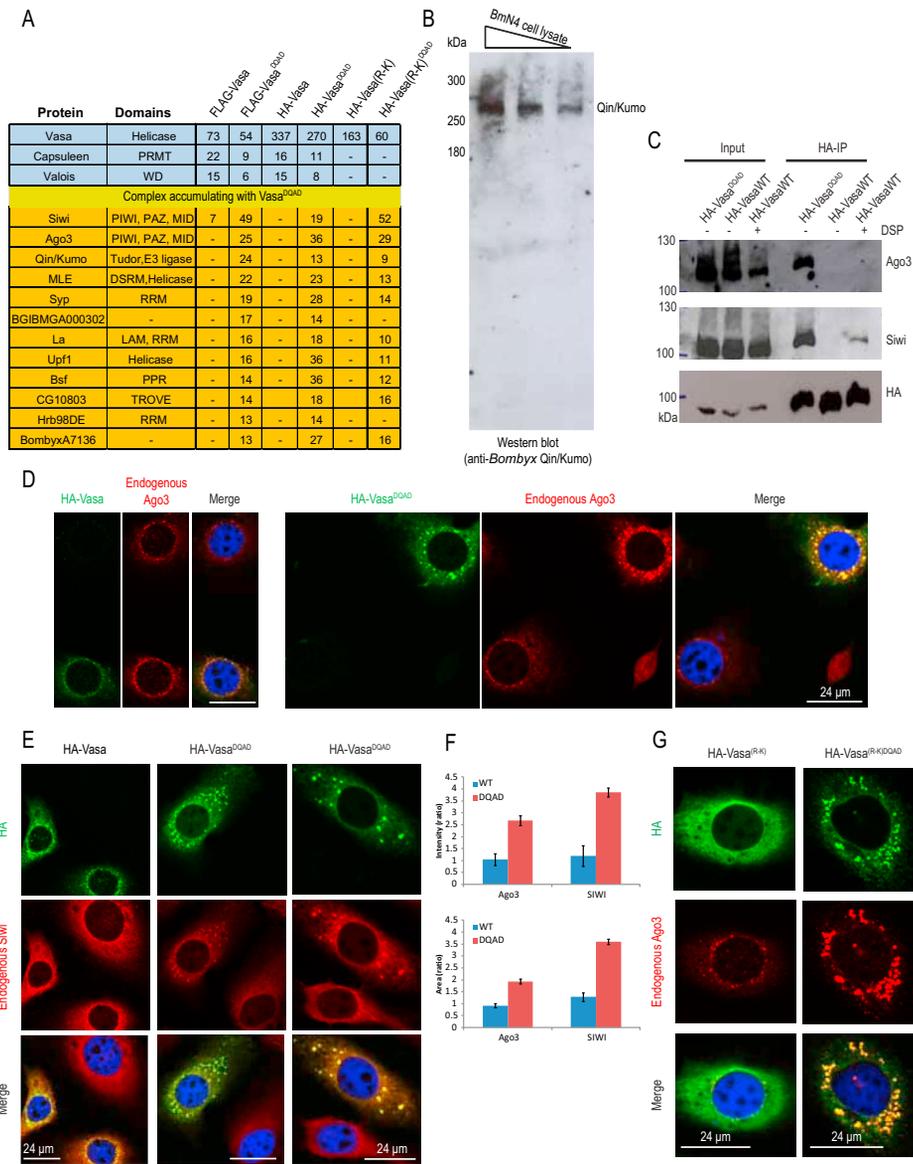


Figure S2. Vasa^{DQAD} Accumulates with Endogenous Piwi Proteins in Nuage, Related to Figure 2

(A) Proteomics of Vasa complexes isolated from transfected BmN4 cells. List of factors identified is shown (a shorter version is presented in Figure 2A). Enrichment value between positive and control samples are given for each protein and only enrichment values above 5 are indicated in the table.

(B) Detection of endogenous Qin/Kumo in BmN4 cell lysates using rabbit polyclonal antibodies.

(C) Co-precipitation of Ping-pong Piwi partners with HA-Vasa^{DQAD}, but not with HA-Vasa isolated from transfected BmN4 cells. However, treatment of cells with the in vivo crosslinking agent DSP allows detection of Siwi with HA-Vasa.

(D) Increased accumulation of endogenous Ago3 in large nuage formed by HA-Vasa^{DQAD} in BmN4 cells.

(E) Increased accumulation of endogenous Siwi in HA-Vasa^{DQAD} nuage. Note that in untransfected cells, Siwi is predominantly present diffused in the cytoplasm. The scale in all images of panels D-E is 24 μm.

(F) Quantification of signal intensity and surface area of Siwi- and Ago3-stained granules in BmN4 cells transfected with either HA-Vasa or HA-Vasa^{DQAD} expression constructs. The results from three independent experiments (20 cells each) are represented as a ratio between that obtained with HA-Vasa (WT) or HA-Vasa^{DQAD} (DQAD). The values for HA-Vasa (WT) were set to 1. Piwi granules increase in size and surface area in cells expressing HA-Vasa^{DQAD}. The error bars represent SE between the different samples from three independent experiments.

(G) Arginine methylation of Vasa is required for nuage localization. Mutation of all twenty N-terminal arginines to lysines (R-K) in Vasa resulted in HA-Vasa^{R-K} being completely diffused in the cytoplasm of transfected BmN4 cells. Endogenous Ago3 serves as a marker for the nuage. When an additional DQAD mutation is made, the HA-Vasa^{(R-K)DQAD} mutant becomes enriched in the nuage. This shows that the protein transits through the nuage and becomes trapped in the nuage in the presence of the DQAD mutation, very similar to the situation with HA-Vasa^{DQAD}. The conclusion from this experiment is that despite its steady-state localization in the nuage, Vasa also freely circulates in and out of the nuage, as demonstrated by FRAP experiments (Figure 1C).

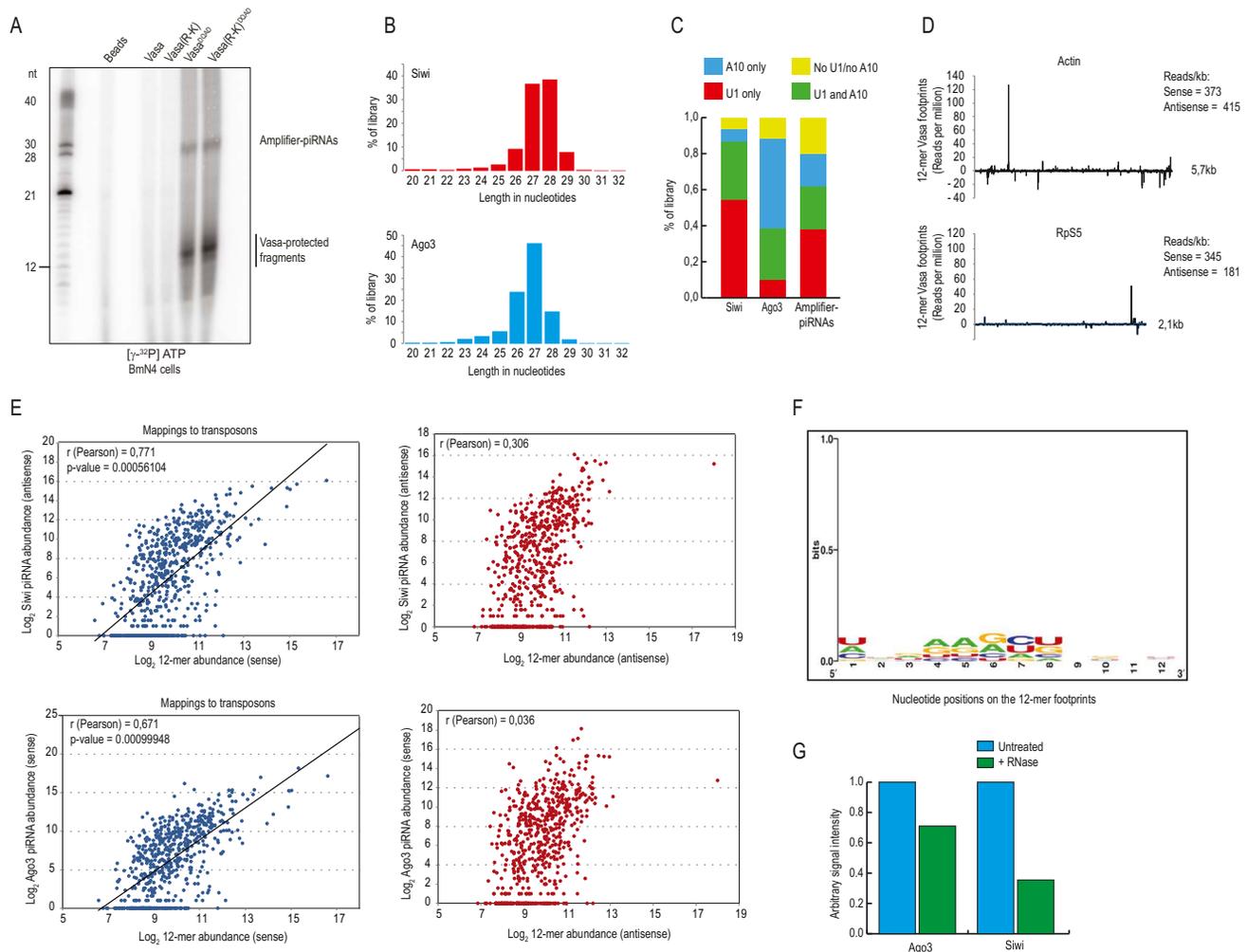


Figure S3. Amplifier Contains Antisense piRNAs and Sense Transposon Transcript Footprints, Related to Figure 3

(A) RNAs associated with HA-Vasa^{(R-K)DQAD} carrying DQAD and R → K mutations. R → K, refers to mutation of the twenty N-terminal arginines (R) in Vasa to lysines (K), and these are designed to prevent arginine methylation of the protein. The R-K mutation does not prevent association with RNAs, in vivo. The two prominent small RNA species are identified as Amplifier-piRNAs and Vasa-footprints.

(B) Read-length distribution of sequences from Siwi and Ago3 libraries.

(C) Nucleotide-bias of reads from indicated libraries. Amplifier-piRNAs are mainly those present in Siwi. Preference for a 5'U (U1) or an adenosine at 10th position (A10) are indicated.

(D) Vasa-footprint reads map poorly to abundant cellular mRNAs like *actin* and *rps5*, indicating that sense transposon transcripts are the major in vivo targets of Vasa.

(E) Scatter plots showing strong correlation for abundant Ago3 piRNAs and Vasa-footprint (12-mers) reads to map to the sense strand of transposon consensus sequences. Similar correlation is seen for Siwi piRNAs (on antisense strand) and Vasa-footprints (on sense strand) mapping to opposing strands. Scenarios (right panels) examining Vasa-footprint abundance from the antisense strand of transposon consensus do not show any correlation.

(F) Analysis of nucleotide bias within the 12-mer Vasa-footprint libraries reveals the absence of any, confirming the sequence-independent binding mode of DEAD box helicases.

(G) RNase-treatment of HA-Vasa^{DQAD} complexes and western analysis for presence of Siwi and Ago3 (experiment in Figure 3J). Signals were quantified and values obtained for untreated sample was set to 1. While retention of both Piwi proteins is sensitive to the treatment, Siwi is much more dependent on RNA for its retention in the complex.

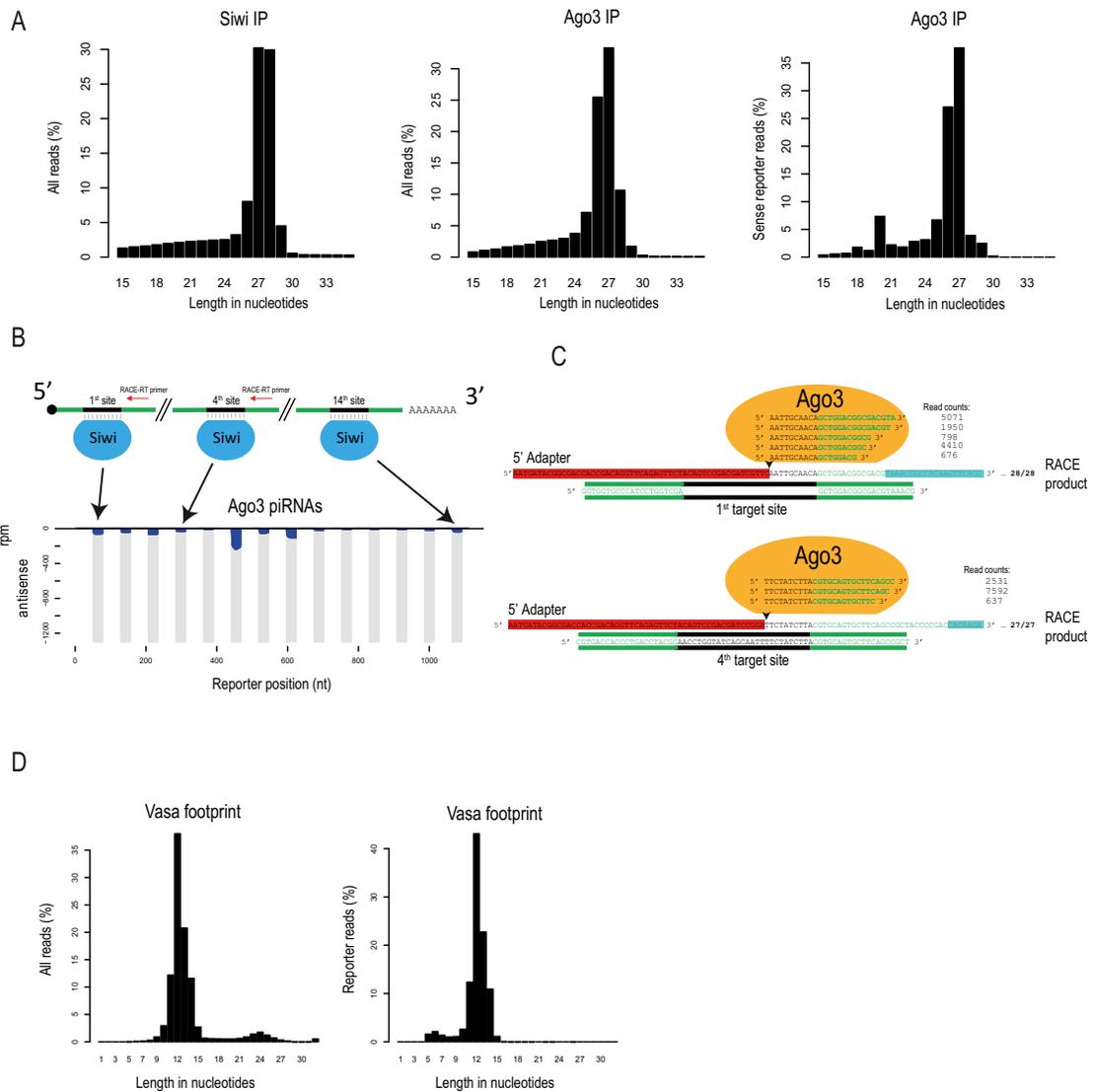


Figure S4. Design of an Artificial Secondary piRNA Precursor, Related to Figure 4

(A) Read-length distribution of sequences in indicated libraries. Note that the third panel shows length distribution of only reporter-derived artificial piRNAs present in Ago3 libraries.

(B) Design of an artificial piRNA precursor based on the green fluorescent protein (GFP) backbone that was rendered noncoding by removal of all ATGs. There are perfectly complementary binding sites for 14 individual piRNAs that were originally selected based on their abundance in sequenced Siwi piRNA libraries. The above density plot (rpm, reads per million) shows the abundance of the same piRNAs in Ago3 libraries prepared in this study (Table S2:DH03-04), highlighting the difficulties of sorting reads as specifically belonging to each Ping-Pong Piwi member. The scale is same as in Figure 4A, and it indicates that the reporter-targeting piRNAs are predominantly present in Siwi complexes, while reporter-derived piRNAs are associating with Ago3.

(C) Cartoon showing details of artificial piRNA generation from the reporter at two locations (sites 1 and 4). Arrowhead shows cleavage site on the reporter by antisense piRNA-guided Siwi. Red shows the 5' adaptor that was ligated to the RNAs in RACE experiments and in blue is the position of the 3' reverse-transcription (RT) primer.

(D) Read-length distribution of Vasa footprints from experiment with co-transfected HA-Vasa^{DQAD} and the artificial piRNA reporter (as analyzed in Figure 4H).

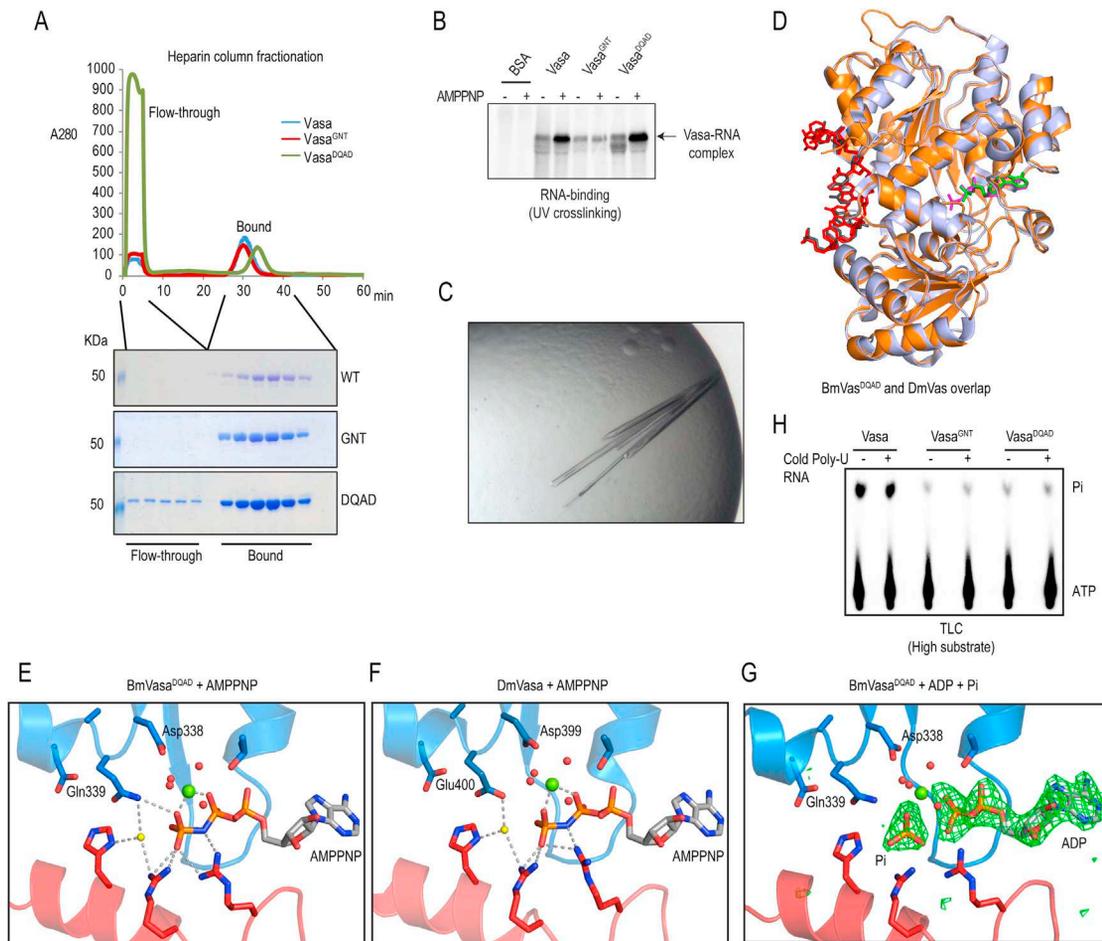


Figure S5. Crystallization of *Bombyx* Vasa^{DQAD}, Related to Figure 5

(A) *Bombyx* Vasa helicase domain with or without indicated point mutations were expressed and purified from *E. coli* as His-tag fusion recombinant proteins. Subsequent to His-tag removal, proteins were purified over a heparin column (retains positively charged nucleic-acid binding proteins). Flow-through and bound (and later eluted) fractions were monitored over time (in minutes, min) by absorption at A280 (above) and Coomassie gel (below). A small but significant part of Vasa^{DQAD} protein failed to bind the column and was found in the flow-through, suggesting potential association with bacterial nucleic acids. Nucleic-acid free proteins were used for further experiments.

(B) Binding of a radioactively labeled poly-U RNA to recombinant Vasa helicase domain proteins assayed with UV crosslinking under conditions with (+) or without (-) the non-hydrolyzable ATP analog AMPPNP.

(C) Crystals obtained with Vasa^{DQAD}-RNA-ADP-Pi.

(D) Overlay of crystal structures from BmVasa^{DQAD}-RNA-ADP-Pi (blue) and DmVasa-RNA-AMPPNP (orange).

(E) Magnified view of the ATP-binding pocket in the *Bombyx mori* Vasa^{DQAD}-RNA-AMPPNP structure (PDB:4D25).

(F) Magnified view of the ATP-binding pocket in the *Drosophila melanogaster* Vasa-RNA-AMPPNP structure (PDB:2DB3).

(G) Magnified view of the ATP-binding pocket in the *Bombyx mori* Vasa^{DQAD}-RNA-ADP-Pi structure (PDB:4D26). The actual electron density (in green) showing the presence of ATP hydrolysis products Pi and ADP within the protein.

(H) Thin layer chromatography experiment detecting ATP hydrolysis by recombinant Vasa helicase domain. High substrate; indicates that in addition to radioactive ³²P-γ-ATP, the reaction was supplemented with cold ATP. Note that under these conditions, only the wild-type Vasa protein shows hydrolysis activity, while Vasa^{DQAD} displays no detectable activity. This is explained by the fact that failure to release hydrolysis products makes the Vasa^{DQAD} protein a low-turnover enzyme. Pi, inorganic phosphate.

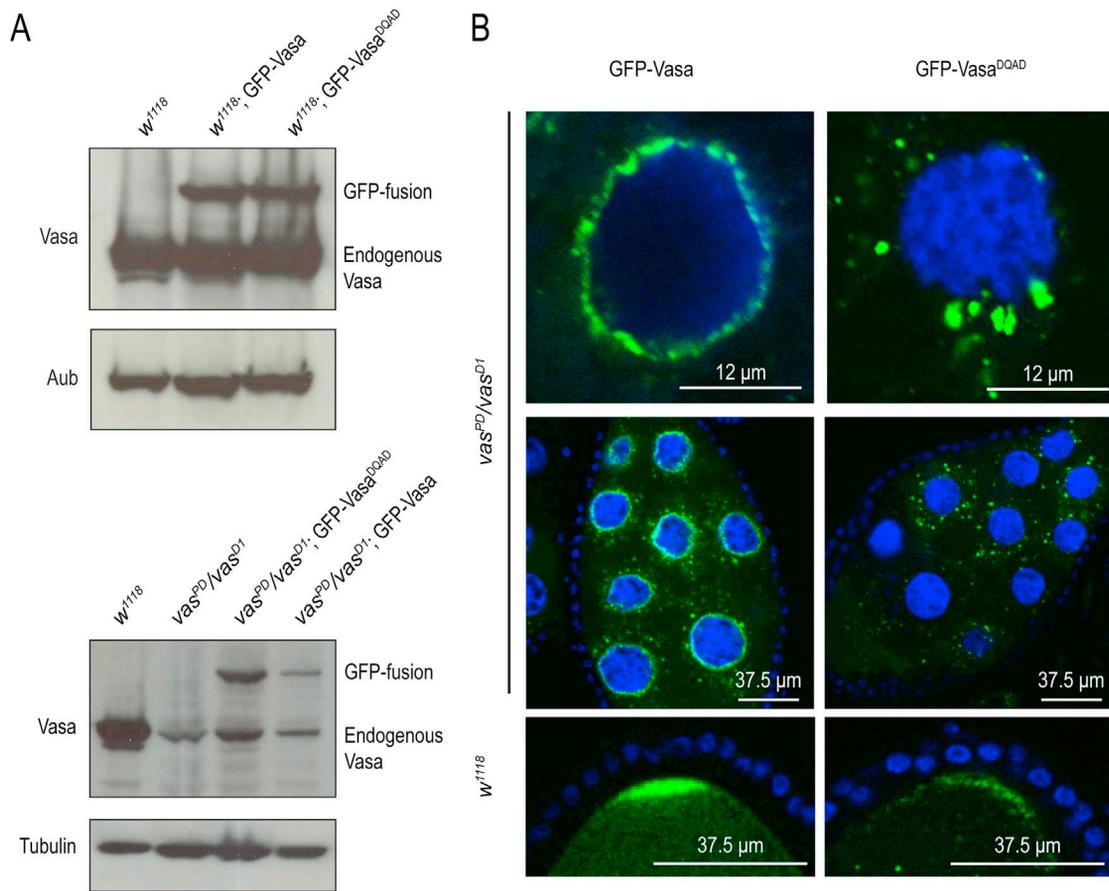


Figure S6. Infertility in *vas* Mutant *Drosophila* Females Is Not Rescued by a DmVasa^{DQAD} Transgene, Related to Figure 6

(A) (Top) Western blot analysis of ovaries from wild-type (w^{1118}) and transgenic flies expressing GFP-DmVasa or GFP-DmVasa^{DQAD}. The anti-DmVasa antibody recognizes both the endogenous and transgene Vasa. Aubergine (Aub) serves as loading control. Note that transgene expression levels are low compared to endogenous Vasa. (Bottom) Similar analysis as above but with *vasa* (*vas*) mutants. Tubulin is the loading control.

(B) Fluorescence detection of GFP fusion proteins in transgenic fly ovaries in the *vas* mutant genetic background. GFP-DmVasa^{DQAD} accumulates in nuages that are removed from the regular perinuclear ring-like localization, potentially indicating an impact on the in vivo dynamics of the protein. (Bottom) Localization of the GFP fusions to the pole plasm.