



UNIVERSITÉ DE STRASBOURG
École doctorale des sciences de la vie et de la santé

Institut de Biologie Moléculaire et Cellulaire
M3i : Modèles Insectes d'Immunité Innée

Thèse
présentée par
Loïc Talide
soutenue le
22 octobre 2019
pour obtenir le grade de
Docteur de l'Université de Strasbourg
spécialité
Aspects moléculaires et cellulaires de la biologie

**Détection des ARNs viraux par
Dicer-2 chez la drosophile**

Thèse dirigée par

Pr. Carine Meignin

Rapporteurs externes :

Dr. Martine Simonelig
Pr. Andréas Pichlmair

Autres membres du jury :

Dr. Todd Blevins
Dr. Peter Sarkies

“To the scientist there is the joy in pursuing truth which nearly counteracts the depressing revelations of truth.”

H. P. Lovecraft (1999). “The Call of Cthulhu and Other Weird Stories”



Figure 0: WordCloud gathering, I hope, every important person I met before or during my Ph.D and that I would like to thank.

Acknowledgements

It is hard to believe that after four long years spent in this lab, the last words that I will be writing are these ones. This Ph.D experience was surely the weirdest one of my life with a lot of ups and many more downs. Nevertheless, I've grown out of it and this could not have been possible without the support of many people.

First, I could not have started my Ph.D without the trust that my advisor, Carine Meignin put in me. You gave me the opportunity to join your newly formed team and guided me during this troubled period. Your scientific support was priceless and kept me on track of my project. For all of that, I am deeply thankful. I also would like to thank Jean-Luc Imler for all the advices and meetings, for involving me in other aspects of lab life such as paper reviewing and for always pushing me forward.

I would like to express my sincere gratitude to my Ph.D jury members, Martine Simonelig, Peter Sarkies, Todd Blevins and Andreas Pichlmair who accepted to evaluate my work both on paper and soon now, in live during my defense. I also greatly appreciated interacting with my mid-thesis jury members, Sébastien Pfeffer and Etienne Decroly and with João Marques who all helped me at crucial moments.

One of the most rewarding experiences that I got during my Ph.D was the opportunity to work in collaboration with the wonderful teams of Brenda Bass and Franck Martin. Having the chance to attend this RNA meeting 2019 with both of you was amazing and your scientific contribution to the different parts of my project invaluable. Special thanks to Fatima, Helen, Joe and Kyle who obtained or helped me obtain some of the data presented in this manuscript.

When it comes to everyday lab support, more names are to add to the list. First, I want to make a huge big up to Nelson and Roenick for basically teaching me bioinformatics and statistics from scratch. Then, I got lucky enough to mentor two great and dedicated interns, Louise and Lisa, who quickly became fully autonomous, trustworthy and did an amazing work. I also want to thank you Claire for your help that started during your master internship and continued until now. Good luck to you and Matthieu for this crazy yet amazing journey of a Ph.D.

I am so grateful to have good friends inside and outside the lab with whom I laughed, sang, spoke, drank, ate, played, headbanged, traveled... during all these years. It was just too difficult for me to cherry pick among the many people to give a shout out to. That is why I hope you will enjoy the very first and maybe most important figure of this manuscript: *Figure 0*. Special mention to you Léa, for always making me feel happy, for sharing my sometimes crazy or boring hobbies and for taking great care of me during the writing. I hope I will be able to give it back to you.

In the end, I could not have become the scientist I am today without my strongest supporters since literally day 1: my parents. Thank you so much for what you did for me.

Abbreviations

ABC: ATP binding cassette	miRNA: micro-RNA
Ago2: argonaute-2	MOI: multiplicity of infection
ATP: adenosine triphosphate	mRNA: messenger RNA
CARD: caspase activation and recruitment domain	NAT: natural antisens transcripts
cDNA: complementary DNA	ORF: open reading frame
CMCT: 1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate	PABP: poly A binding protein
CrPV: cricket paralysis virus	PAZ: Piwi Argonaut and Zwillie
Ct: threshold cycle	PFU: plaque-forming unit
DCV: drosophila C virus	pi: post infection
Diap I: death-associated inhibitor of apoptosis I	piRNA: piwi-interacting RNA
dm: <i>Drosophila melanogaster</i>	qPCR: quantitative PCR
DMS: dimethyl sulfate	RACE: rapid amplification of cDNA ends
DNA: deoxyribonucleic acid	RdRp: RNA-dependent RNA Polymerase
DRA: Duplex RNA activated ATPase	RFP: red fluorescent protein
dsRNA: double strand RNA	RIG-I: retinoic acid-inducible gene I
endo-siRNA: endogenous siRNA	RISC: RNA-induced silencing complex
FL: full length	RLR: Rig-I like receptor
GFP: green fluorescent protein	RNA: ribonucleic acid
GMR: glass multiple reporter	RNAi: RNA interference
GO: gene ontology	RNase: ribonuclease
HEL: helicase	RT-qPCR: reverse-transcription qPCR
HRP: horseradish peroxidase	RT: room temperature
HTS: high throughput sequencing	S2: Schneider 2
IFN: interferon	SF: superfamily
IGR: intergenic region	siRNA: small interfering RNA
IMD: immunodeficiency	SNBN: Stephanie Blandin (library index)
IRES: internal ribosome entry site	ssRNA: single strand RNA
ISG: interferon stimulated genes	STAT: signal transducers and activators of transcription
JAK: janus kinases	TRIS: tris(hydroxymethyl) Aminomethane
LCTD: Loïc Talide (library index)	U: unit
LGP2: laboratory of genetics and physiology 2	UTR: untranslated region
MAVS: mitochondrial antiviral signaling protein	VPg: virus protein genome-linked
MDA5: melanoma differentiation-associated protein 5	VSR: viral suppressors of RNAi
mRNA: messenger RNA	VSV: vesicular stomatitis virus
	w^{IR}: white inverted repeat
	WT: wild type

Table of contents

GENERAL INTRODUCTION	2
PREAMBLE	3
SENSING VIRAL INFECTIONS IN INSECTS: A DEARTH OF PATHWAY RECEPTORS	4
OUTRO	19
SEARCHING FOR <i>DROSOPHILA MELANOGASTER</i> DICER-2 ENTRY POINT ON DICISTROVIRUSES	20
INTRODUCTION	21
RESULTS	24
I. KINETICS OF siRNAs APPARITION IN S2 CELLS DURING DCV AND CrPV INFECTION	24
II. SMALL RNA HIGH THROUGHPUT SEQUENCING (HTS) OF DCV INFECTED ADULT FLIES	34
CONCLUSIONS – DISCUSSIONS – PERSPECTIVES	40
CHARACTERIZATION OF DCV 5' UTR REGION	49
INTRODUCTION	50
RESULTS	52
I. ANALYSIS OF THE DOMAIN I STRUCTURES OF DCV AND CrPV	52
II. CHARACTERIZATION OF DCV DOMAIN I SENSITIVITY TOWARD CLEAVAGE	55
III. RNAI-BASED SCREEN IN S2 CELLS TO IDENTIFY NEW ANTIVIRAL AND PROVIRAL PROTEINS	63
CONCLUSIONS – DISCUSSIONS – PERSPECTIVES	68
STUDY OF DICER-2 HELICASE DOMAIN INVOLVEMENT IN RNA SENSING AND PROCESSING	76
INTRODUCTION	77
RESULTS	81
I. IMPACT OF TWO DICER-2 HELICASE MUTATIONS ON THE ENDO-siRNA PATHWAY	81
II. IMPACT OF TWO DICER-2 HELICASE MUTATIONS ON THE ANTIVIRAL siRNA PATHWAY	90
CONCLUSIONS – DISCUSSIONS – PERSPECTIVES	94
CONCLUDING REMARKS	102
MATERIALS & METHODS	103
BIBLIOGRAPHY	117

Preamble

During the period of my Ph.D, I got the chance to write a book chapter in “**Insect Molecular Virology: Advances and Emerging Trends**”. This book was edited by eminent scholar and professor Bryony C. Bonning and published in June 2019. The result of this writing work is a global review of the insects’ antiviral defense mechanisms with a specific focus made on the innate immunity receptors that sense viral infections. As stated by the title of this manuscript, my Ph.D work mainly revolved around the detection of viral RNAs by Dicer-2 in *Drosophila melanogaster*.

Thus, due to the relevance of my book chapter and its linked bibliography in the frame of my Ph.D work, I decided to use it as a global introduction of my manuscript. Then, a small transition will highlight the main questions that guided the writing of the following three Chapters. Finally, deeper introduction of some specific aspects of the field will be done in the corresponding Chapters.

Sensing Viral Infections in Insects: A Dearth of Pathway Receptors

3

Loïc Talide¹, Jean-Luc Imler^{1,2} and Carine Meignin^{1,2*}

¹Université de Strasbourg, Centre National de la Recherche Scientifique, Insect Models of Innate Immunity (M3i; UPR9022), Strasbourg, France.

²Université de Strasbourg, Faculté des Sciences de la Vie, Strasbourg, France.

*Correspondence: c.meignin@ibmc-cnrs.unistra.fr

<https://doi.org/10.21775/9781912530083.03>

Abstract

Insects, the most diverse group of animals, can be infected by an extraordinary diversity of viruses. Among them, arthropod-borne viruses can be transmitted to humans, while bee and silkworm viruses cause important economic losses. Like all invertebrates, insects rely solely on innate immunity to counter viral infections. Protein-based mechanisms, involving restriction factors and evolutionarily conserved signalling pathways regulating transcription factors of the NF- κ B and STAT families, participate in the control of viral infections in insects. In addition, RNA-based responses play a major role in the silencing of viral RNAs. We review here our current state of knowledge on insect antiviral defence mechanisms, which include conserved as well as adaptive, insect-specific strategies. Identification of the innate immunity receptors that sense viral infection in insects remains a major challenge for the field.

Introduction

With more than 1 million known species, insects are the largest group of multicellular organisms, representing over 70% of animal species. Dating back to the Early Ordovician (about 480 million years ago, Ma), they were among the first animals to colonize terrestrial and freshwater ecosystems and have undergone major expansions, culminating in the spectacular diversification of holometabolous

insects (Hymenoptera, Diptera, Lepidoptera) during the Early Cretaceous (about 120 Mya) (Misof *et al.*, 2014). Insects can be credited with major adaptations, such as flight and establishment of social groups. Importantly, they have undoubtedly contributed to shape the planet's biota and actively interact with other multicellular eukaryotes such as plants and vertebrates. Like them, insects are exposed to a large panel of infectious microorganisms, which they control through their innate immune system (Hoffmann *et al.*, 1999).

Among infectious microbes, viruses represent a particular threat because they offer few intrinsic targets for inhibition by antiviral molecules. This is because they consist in their simplest form of a nucleic acid encapsidated in a protein shell, and hijack molecular machineries from host cells to complete their replication cycle. Therefore, viruses exert great selective pressure on their host to evolve resistance pathways. These, in turn, favour the adaptation of viruses to escape antiviral mechanisms. This arms race results in the diversification of both host-defence and virus escape mechanisms. As a result, it can be highly instructive to broaden the study of antiviral immunity to non-mammalian models. In light of their diversity, insects represent an interesting group of animals for this type of comparative study (Marques and Imler, 2016; Martins *et al.*, 2016).

Recent advances in high-throughput sequencing technologies have opened the way to the

characterization of the virome of insects (i.e. the genetic diversity of viruses in a biological sample; see also Chapter 1). In a landmark article, Yong-Zhen Zhang and colleagues analysed the transcriptome of more than 220 invertebrate species covering nine animal phyla and reported the identification of close to 1500 new viruses (Shi *et al.*, 2016). Thus, infection by one or several viruses is common in invertebrates. In addition, the genetic diversity of these viruses surpassed that described previously. Many newly identified viruses fell between families and genera from the current virus classification, filling major phylogenetic gaps and revealing that viruses form a continual spectrum of phylogenetic diversity (Shi *et al.*, 2016). A more detailed analysis focusing on 70 arthropod species and negative-sense RNA viruses, which include important pathogens causing a variety of diseases in humans (flu, rabies, encephalitis, haemorrhagic fever), led to the discovery of 112 new viruses (Li *et al.*, 2015). This study revealed that much of the diversity of negative-sense RNA viruses found in plants and vertebrate animals falls within the genetic diversity of viruses associated with arthropods (Dudas and Obbard, 2015; Li *et al.*, 2015). Of note, arthropods (and insects in particular) can live in large and dense populations, facilitating propagation and transmission of viruses. The close interaction between many insects and plants or vertebrate animals further support the hypothesis that negative-sense RNA viruses, including vertebrate-specific ones, are derived from arthropod dependent viruses (Li *et al.*, 2015).

There are several specific reasons to study virus–host interactions in insects. First, infection of insects can cause important economic losses (e.g. viral diseases of silkworms; contribution to colony-collapse in honey-bees) (Bradshaw *et al.*, 2016; Carrillo-Tripp *et al.*, 2016). Second, haematophagous insects such as *Aedes* or *Anopheles* mosquitoes can transmit viral diseases to mammalian hosts. These viruses, the arthropod-borne viruses or arboviruses, include dengue (DENV), yellow fever (YFV), and West-Nile virus (WNV) (Molina-Cruz *et al.*, 2016; Powers and Waterman, 2017). Third, microbial pathogens (e.g. baculoviruses) can be used as biological control agents against insect pests, which necessitates some knowledge of the host response to these microorganisms (Popham *et al.*, 2016). Fourth, insects such as the genetically

tractable model organism *Drosophila melanogaster* can be used to decipher evolutionarily conserved innate immune mechanisms.

Antiviral immunity in insects

NF- κ B and STAT dependent inducible responses

Innate immunity is the first line of defence that multicellular organisms deploy to limit pathogen infections. In vertebrates, the innate immune response also regulates the production of cytokines and co-stimulatory molecules, which shape the subsequent adaptive immune response (Hoffmann *et al.*, 1999). Studies on innate immunity in *Drosophila* initially focused on bacterial and fungal infections and revealed that the systemic production of antimicrobial peptides (AMPs) (humoral response) plays an important role in host defence (Steiner *et al.*, 1981; Imler and Bulet, 2005). Also, in the haemolymph, proteolytic cascades involving sequential activation of serine proteases participate in the clotting and melanisation responses to wounding (Binggeli *et al.*, 2014; Theopold *et al.*, 2014). In addition, cellular responses involving both circulating and sessile haemocytes participate in antimicrobial host defence in flies, in particular *via* phagocytosis of bacteria or virus infected cells by macrophage-like plasmatocytes and, in larvae, encapsulation of parasitic wasp eggs by lamellocytes (Gold and Brückner, 2015; Letourneau *et al.*, 2016; Weavers *et al.*, 2016). In the case of viruses, it is now well established that the cell intrinsic mechanism of RNA interference (RNAi) plays a central role in the control of viral infections in insects, as it does in plants and other invertebrates (see below) (Ding, 2010). In addition, inducible responses and restriction factors also contribute to resistance to viral infections (Mussabekova *et al.*, 2017).

Expression of AMPs is controlled by the evolutionarily conserved signalling pathways Toll and IMD (immune deficiency), which regulate the activity of transcription factors of the NF- κ B family (Hoffmann, 2003). These pathways are activated by pattern recognition receptors (PRRs) that sense components of the bacterial or fungal cell wall such as peptidoglycan in the case of bacteria or β -glucans in the case of fungi (Steiner, 2004; Royet *et al.*, 2011; Rao *et al.*, 2018). Based on the

transcriptomic signature of virus-infected insects and the phenotype of flies or mosquitoes with genes encoding important components of the pathway mutated or silenced, both of these pathways are proposed to also participate in antiviral immunity (Avadhanula *et al.*, 2009; Costa *et al.*, 2009; Ferreira *et al.*, 2014; McFarlane *et al.*, 2014; Paradkar *et al.*, 2014; Carissimo *et al.*, 2015; Lamiabile *et al.*, 2016a; Fig. 3.1). A third evolutionarily conserved pathway connected to inflammation in mammals, the Jak/STAT pathway, has also been proposed to play a role in insect antiviral immunity (Dostert *et al.*, 2005; Fragkoudis *et al.*, 2008; Souza-Neto *et al.*, 2009; Paradkar *et al.*, 2012; Kemp *et al.*, 2013; Barribeau *et al.*, 2015; Carissimo *et al.*, 2015; Merklings *et al.*, 2015a; Zhang *et al.*, 2016; Jupatanakul *et al.*, 2017; West and Silverman, 2018). This pathway is activated by cytokines of the Unpaired (Upd) family, which are upregulated by viral infection or stress in *Drosophila* (Jiang *et al.*, 2009; Kemp *et al.*, 2013; Gordon *et al.*, 2018; Fig. 3.1). In *Culex* mosquitoes and bumblebees, a single von Willebrand factor type C domain secreted factor, related to the *Drosophila* antiviral factor Vago (Deddouche *et al.*, 2008), appears to activate the Jak/STAT pathway and antiviral immunity (Paradkar *et al.*, 2012; Wang *et al.*, 2017). Two major questions in the field arise at this stage: (1) How are these pathways activated

by viruses? And (2) what are the antiviral effectors they regulate and how do they counter viruses?

Activation of the Toll and IMD pathways by viruses

The Toll and the IMD pathways are activated by PRRs of the peptidoglycan recognition protein (PGRP) and β -glucan recognition protein (GNBP) families, which sense components of the bacterial and fungal cell walls (reviewed in Ferrandon *et al.*, 2007; Royet *et al.*, 2011). Interestingly, both pathways can also be activated by virulence factors. In the case of the Toll pathway, the circulating zymogen Persephone senses protease activity independently of microbial patterns and triggers a proteolytic cascade that culminates in the processing of the cytokine Spaetzle to generate an active Toll ligand (Gottar *et al.*, 2006; El Chamy *et al.*, 2008; Issa *et al.*, 2018; Fig. 3.1). In a conceptually similar manner, deamidation by bacterial toxins of a critical glutamine residue in the Rho GTPase Rac2 can be sensed by IMD, resulting in activation of the pathway independently from PRRs (Boyer *et al.*, 2011).

For the moment, it remains unclear how viruses activate the Toll and IMD pathways, and whether this involves PRRs or other sensors. In mammals, PRRs belonging to different structural families

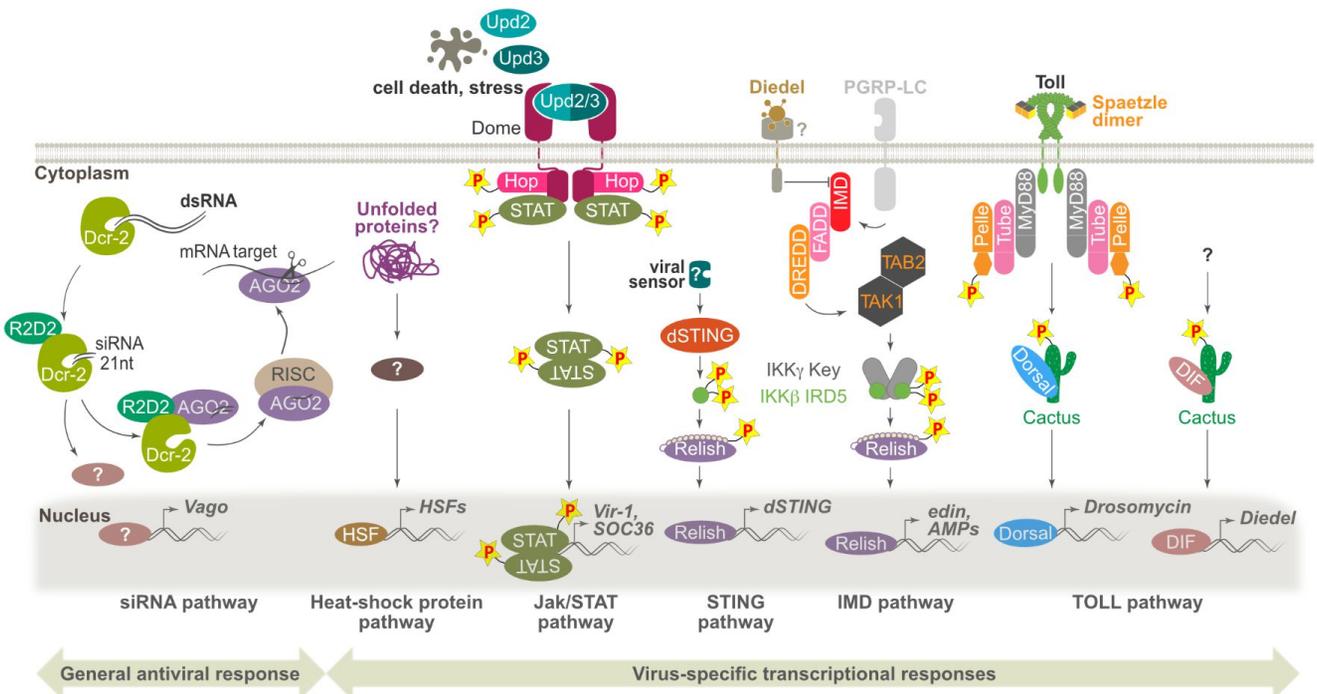


Figure 3.1 Antiviral transcriptional responses in *Drosophila melanogaster*. The signalling pathways activated during viral infection in *Drosophila* are illustrated.

[e.g. Toll like receptors (TLRs), RIG-I-like receptors (RLRs), nucleotide binding domain and leucine-rich repeat containing receptors (NLRs)] sense viral nucleic acids and trigger expression of the antiviral cytokines of the interferon family (Goubau *et al.*, 2013; Roers *et al.*, 2016). At this stage, the only identified receptor for viruses is Dicer-2, which interacts with viral double-stranded (ds)RNA. Interestingly, Dicer-2 has been proposed to serve a dual function in antiviral immunity, activating both antiviral RNAi and signalling leading to expression of antiviral molecules (e.g. Vago) (Ded-douche *et al.*, 2008; Paradkar *et al.*, 2012; Wang *et al.*, 2017; Asad *et al.*, 2018; Fig. 3.1). Viruses also affect the physiology of the cells, which can trigger a response. For example, infection by *Drosophila C virus* (DCV) induces a heat shock response, possibly in reaction to the accumulation of unfolded viral proteins in the cytosol of infected cells (Mer-kling *et al.*, 2015b). Viruses are also notorious for inhibiting cellular translation or altering cellular membranes, which may trigger a cellular response, such as apoptosis or autophagy (Shelly *et al.*, 2009; Nainu *et al.*, 2015; Lamiabile *et al.*, 2016b; Khong *et al.*, 2017).

Of note, it is becoming apparent that components of the canonical Toll or IMD pathways may participate in novel pathways, responding to different cues and activating distinct transcriptional programs. For example, two closely related NF- κ B proteins, Dorsal and DIF, are regulated by the Toll pathway in fruit flies (Tanji *et al.*, 2010). In spite of strong sequence conservation, DIF lacks important features of Dorsal-mediated pattern formation in the *Drosophila* embryo (Stein *et al.*, 1998). Conversely, DIF, but not Dorsal, is required to regulate expression of the antifungal peptide Drosomycin to resist fungal infections in adult flies, although both proteins are expressed (Lemaitre *et al.*, 1995; Rutschmann *et al.*, 2000a). Interestingly, in a context of oral infection of adult flies, it is Dorsal, rather than DIF, that is required for resistance to DCV (Ferreira *et al.*, 2014; Fig. 3.1). The other components of the Toll pathway, up to the cytokine Spaetzle, are also involved in resistance to DCV. An intriguing unresolved question pertains to the proteases acting upstream of Spaetzle in the context of DCV oral infection (Ferreira *et al.*, 2014). Interestingly, although not essential in the context of DCV infection, DIF is required for the strong

up-regulation of the cytokine Dieldel (see below) triggered by Sindbis virus (SINV) and Vesicular stomatitis virus (VSV). However, MyD88, a key signalling adapter protein of the Toll pathway, is dispensable for induction of Dieldel (Lamiabile *et al.*, 2016a). This suggests that, in the context of some viral infections, DIF can be activated in a MyD88-independent manner and, presumably, by a mechanism distinct from the canonical Toll pathway.

The IMD pathway can be activated by SINV in *Drosophila*, by a mechanism that remains unclear (Avadhanula *et al.*, 2009; Lamiabile *et al.*, 2016a). Some antimicrobial peptides regulated by this pathway have been proposed to participate in the control of this virus (Huang *et al.*, 2013). The relevance of this pathway in the context of viral infections is highlighted by the observation that members of different families of DNA viruses have hijacked a gene called *Dieldel*, which encodes an immunomodulatory cytokine down-regulating the IMD pathway and antagonizing apoptosis (Coste *et al.*, 2012; Lamiabile *et al.*, 2016a; Mlih *et al.*, 2018). This observation prompted a reanalysis of the contribution of the IMD pathway to antiviral immunity. Unexpectedly, this study revealed that two components of the pathway, the kinase IKK β and the NF- κ B transcription factor Relish, restrict replication of the dicistroviruses DCV and Cricket paralysis virus (CrPV) in *Drosophila* (Goto *et al.*, 2018). Strikingly, the other components of the canonical IMD pathway, including IKK γ , the regulatory subunit of the IKK complex, are not required for virus suppression. Further analysis revealed that in the context of viral infection, the kinase IKK β is activated by a different pathway, involving the *Drosophila* homologue of the mammalian gene STING (Stimulator of Interferon genes), a critical component of the cytosolic DNA sensing pathway in mammals (Fig. 3.1). The genes regulated by this alternative new pathway are different from the antibacterial peptides regulated by IMD, suggesting that Relish interacts with an additional transcription factor (Goto *et al.*, 2018). This would be conceptually similar to the cooperation of NF- κ B with IRF3 to regulate antiviral genes in mammals (Ikushima *et al.*, 2013). Also pointing to an involvement of components of the IMD pathway in antiviral immunity, in *Culex* mosquitoes, a Dicer-2-dependent pathway regulates

expression of the gene *Vago* upon activation of a TRAF factor and the homologue of Relish, REL2 (Paradkar *et al.*, 2014). In summary, it appears that NF- κ B pathways are more diverse than initially thought, and that these evolutionarily conserved transcription factors can be activated by alternative branches of the canonical Toll and IMD pathways initially characterized in the context of bacterial and fungal infections. Consistent with these findings, vankyrins form a family of I κ B-like molecules encoded by polydnviruses that can antagonize NF- κ B-dependent responses (Kroemer and Webb, 2005, 2006; Gueguen *et al.*, 2013).

Control of viruses by restriction factors

Constitutively expressed restriction factors also participate in the control of viruses in insects. Some of them are evolutionarily conserved and the functions of their mammalian homologues point to mechanisms for virus inhibition. For example, *Drosophila* ref(2)P, a restriction factor for Sigma virus (i.e. a Rhabdovirus that is a natural pathogen of *Drosophila*), is a homologue of human p62/sequestosome-1, pointing to possible involvement of autophagy in the control of this virus (Carré-Mlouka *et al.*, 2007; Ktistakis and Tooze, 2016). Interestingly, other restriction factors are not conserved and represent insect-specific adaptations. One example is the gene *CHKov1*, which encodes another restriction factor for Sigma virus (Magwire *et al.*, 2011). The function of *CHKov1* is unknown, but the protein contains a choline kinase domain, which is intriguing in light of the function of choline kinase as a host factor for hepatitis C virus (HCV) in human hepatocytes (Wong and Chen, 2016, 2017). The gene *pastrel* is another example of a non-conserved gene having a potent restricting activity on DCV and CrPV in *Drosophila* (Magwire *et al.*, 2012; Cao *et al.*, 2017). How this protein functions to block viral replication, and if its action involves interaction with viral RNA remains unknown.

In summary, it is clear that protein-based mechanisms, involving evolutionarily conserved genes and pathways, are involved in insect antiviral immunity. However, the receptors that sense viruses and trigger these responses are unknown. As a result, Dicer-2 remains the only well characterized sensor for viral infection in insect cells (Fig. 3.1).

The siRNA pathway of RNA interference: mechanism and regulation

RNAi pathways in insects

Fire, Mello and coworkers coined the term RNA interference, or 'RNAi', to describe the observation that dsRNA can block gene expression when introduced into *Caenorhabditis* nematodes (Fire *et al.*, 1998). This discovery was rapidly followed by biochemical characterization of RNAi in fruit flies using embryos and the S2 cell line (Hammond *et al.*, 2000; Zamore *et al.*, 2000; Bernstein *et al.*, 2001; Liu *et al.*, 2003). We currently know that different RNAi mechanisms play important regulatory roles in development, maintenance of genome stability, gene expression and antiviral defence. These RNA-based mechanisms involve proteins of the Argonaute family combined with small regulatory RNAs ranging from 20–30 nt length (Treiber *et al.*, 2019). Argonaute proteins associate with small RNAs that guide them towards target mRNAs leading to inhibition of their translation or direct cleavage catalysed by their RNase-H like domain (Song *et al.*, 2004; Ma *et al.*, 2005; Yuan *et al.*, 2005). In insects, three RNAi pathways have been well documented, involving small RNAs of 21–22 nt (small interfering RNAs or siRNAs), 22–23 nt (micro RNAs or miRNAs) or 24–30 nt (Piwi-interacting RNAs or piRNAs). Whereas siRNAs and miRNAs are produced from dsRNA precursors by RNaseIII proteins such as Dicers and Drosha, piRNAs are processed independently of these enzymes (see section on piwi pathway below for more detail). *Drosophila* genetics defined two well-separated pathways involving Dicer-1/AGO1 for miRNAs and Dicer-2/AGO2 for siRNAs (Lee *et al.*, 2004; Okamura *et al.*, 2004). A different clade of AGO proteins involving PIWI, AGO3 and Aubergine (Aub) in *Drosophila* regulates the production and activity of piRNAs (reviewed in Huang *et al.*, 2017). miRNAs are essentially produced from nuclear precursors and participate in tight regulation of gene expression during development or cellular homeostasis (e.g. Posadas and Carthew, 2014). Of note, some DNA viruses use virus-encoded or cellular miRNAs to regulate their own gene expression or to modulate host cell transcriptome (Müller and Imler, 2007; Hussain and Asgari, 2014; see also Chapter 4). However, control of viral infections in

insects mostly relies on siRNAs and, in some cases might also involve piRNAs.

Production of small non-coding RNAs by Dicer enzymes

The antiviral RNAi pathway, the siRNA pathway, is triggered by long dsRNA, which are processed in insects by Dicer-2 (Dcr-2). This cytoplasmic enzyme is composed of a N-terminal Duplex RNA activated ATPase (DRA) domain, a central double-stranded RNA-binding domain (dsRBD), a platform-PAZ domain, two RNaseIII domains and a C-terminal dsRBD (Fig. 3.2, left). Interestingly, Dicer enzymes share a common phylogenetic origin with the vertebrate RLRs, which sense viral nucleic acids in the cytosol and trigger synthesis of interferons. Although Dicer-2 and RLRs have different sizes and domain composition, they all contain a conserved DRA domain (Paro *et al.*, 2015). Our understanding of the contribution of the DRA domain of Dicer-2 to virus sensing remains limited and is largely based on *in vitro* studies.

In vitro, Dicer-2 is efficient at processing long dsRNAs in addition to shorter structured RNAs such as pre-miRNAs which are *bona fide* targets of

Dicer-1 (Cenik *et al.*, 2011). *In vivo*, it appears that the activity of Dicer enzymes is regulated by cofactors containing two or three dsRBDs, e.g. R2D2 and Loquacious (Loqs) (Fig. 3.2, right). Dicer-2 forms a stable heterodimer with R2D2, which restricts its cellular localization to cytoplasmic D2 bodies (Nishida *et al.*, 2013) and is mandatory for the efficient loading of siRNAs on AGO2 (Liu *et al.*, 2003; Lee *et al.*, 2004; Tomari *et al.*, 2004a). Based on *in vitro* experiments, a role for R2D2 in preventing processing of pre-miRNAs by Dicer-2 has also been proposed (Cenik *et al.*, 2011). However, this was not confirmed by *in vivo* experiments (Marques *et al.*, 2013). Furthermore, inorganic phosphate has been proposed as a factor restricting the panel of targets for Dicer-2 *in vitro*, precluding it from processing aberrant pre-miRNA targets and short dsRNAs (Fukunaga *et al.*, 2014). The other cofactor, Loqs has two main isoforms, Loqs-PB and Loqs-PD. Whereas Loqs-PB functions as a cofactor of Dicer-1 in the miRNA pathway (Förstemann *et al.*, 2005; Jiang *et al.*, 2005; Saito *et al.*, 2005), Loqs-PD enhances production of siRNAs by Dicer-2, especially from synthetic dsRNA and endogenously encoded dsRNAs derived from structured loci,

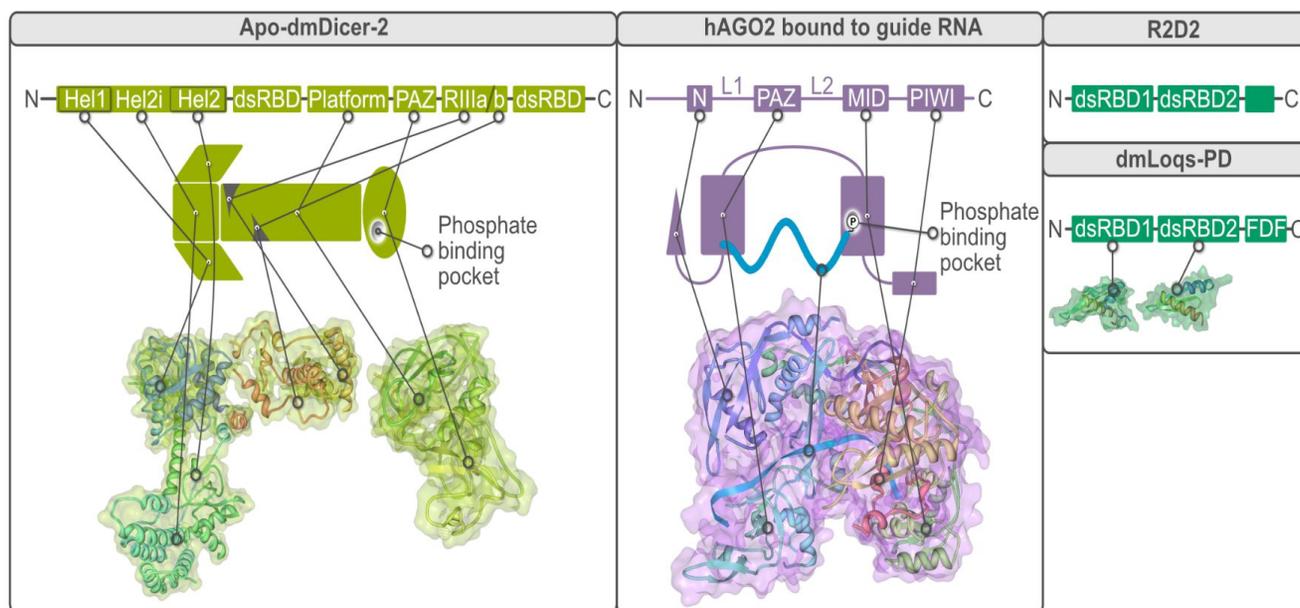


Figure 3.2 Structural organization of the canonical components of the siRNA pathway. The domain organization is shown at the top, above a schematic representation of the proteins and experimentally determined 3D structures. (Left) Apo-dmDicer-2 3D structure was obtained by Cryo-electron microscopy at a resolution of 7.1Å (Sinha *et al.*, 2018; PDB ID: 6BUA). (Centre) Guide RNA-loaded human AGO2 structure was resolved at 2.9Å using X-Ray crystallography (Schirle *et al.*, 2014, PDB ID: 4W5N). (Right) No 3D structures were obtained so far for Loqs-PD and R2D2 full proteins. However, the 3D structures of the two dsRBDs of dmLoqs, identical between all isoforms, were individually predicted using solution Nuclear Magnetic Resonance spectroscopy (Tants *et al.*, 2017; PDB IDs: 5NPG, 5NPA).

sense–antisense pairs and transposable elements (Zhou *et al.*, 2009; Marques *et al.*, 2010; Miyoshi *et al.*, 2010a).

Once produced, the siRNA duplex is loaded on the effector protein AGO2 in a highly coordinated but energetically unfavourable series of events. In a first step, the Dcr-2/R2D2 complex, which displays intrinsically low affinity for duplex siRNAs, associates with TAF11, a TATA-box binding protein Associated Factor. Colocalized with Dicer-2 and R2D2 in D2 bodies, TAF11 acts as a chaperone facilitating the tetramerization of the Dicer-2–R2D2 heterodimer and increasing affinity for siRNAs by tenfold (Liang *et al.*, 2015). R2D2 tends to preferentially bind the extremity of the siRNA duplex showing the strongest stability, creating an asymmetry in the complex (Tomari *et al.*, 2004b). This asymmetry determines the preferential loading of the strand featuring the least stable 5′ extremity in AGO2 to serve as guide siRNA (Fig. 3.3B and C).

Structure–function of AGO proteins

AGO proteins are composed of four globular domains named N, PAZ, MID and PIWI (Fig. 3.2, centre). They adopt a closed, flexible and unstable conformation not suitable to accept the siRNA duplex in their ‘Apo’ form. A chaperone machinery composed of the Hsp70 system (Hsp40 + Hsp70) and Hsp90 system (Hop, Hsp90 and p23) is required for the efficient loading of siRNA duplexes. Briefly, Hsp70 opens the structure of AGO2 while Hsp90 is required to extend the duration of this opened state to allow sufficient time for the recognition of the 5′ phosphate at the extremity of the guide strand and subsequent loading of the entire duplex (Miyoshi *et al.*, 2010b; Iwasaki *et al.*, 2010, 2015; Tsuboyama *et al.*, 2018). Both processes require ATP hydrolysis. The coordinated action of the heat shock proteins was proposed as the trigger for Dicer-2–R2D2 tetramer destabilization and transfer of the siRNA duplex (Fig. 3.3C). Intriguingly, Hsp proteins can be induced by viral infections in *Drosophila*. However, the siRNA pathway remains functional in flies mutant for the Heat Shock factor (Merkling *et al.*, 2015b). In spite of this progress, the exact mechanism of AGO2 loading remains unclear especially because a 3D structure of the protein in its ‘Apo’ form is lacking. The semi-closed AGO2 protein loaded with the siRNA

duplex is rigid and stable, and constitutes the pre-RNA Induced Silencing Complex (RISC). Studies with AGO proteins from eukaryotic (human) or prokaryotic (*Pyrococcus furiosus*) systems indicate that the strand of the duplex showing the less stable 5′ phosphate extremity is anchored to the phosphate binding pocket of the AGO2 MID domain while its 3′ extremity is bound to the hydrophobic cavity of the PAZ domain (Ma *et al.*, 2004; Song *et al.*, 2004; Fig. 3.3C). In the pre-RISC, the passenger strand occupies the same position as the future target RNAs (Kim *et al.*, 2007). The full-length crystal structure of human AGO2 complexed with RNA reveals large structural differences between Argonautes from different kingdoms of life, even if individual domains superimpose reasonably well (Schirle and MacRae, 2012).

Maturation and slicing

After loading, the pre-RISC is matured through two essential steps required for downstream RNA silencing activity. First, the passenger strand is discarded. This is achieved by the coordinated action of the N-terminal domain of AGO2, acting as a wedge to unwind the siRNA duplex (Kwak and Tomari, 2012) and the PIWI domain RNaseH-like catalytic core, which cleaves the passenger strand in two small RNAs of 9 and 12 nt (Kim *et al.*, 2007; Matranga *et al.*, 2005; Miyoshi *et al.*, 2005; Rand *et al.*, 2005). The Component 3 Promoter Of RISC (C3PO) complex then helps AGO2 to get rid of the unstable cleavage products (Liu *et al.*, 2009; Ye *et al.*, 2011; Mo *et al.*, 2018) (Fig. 3.3C).

A slicer-independent ejection model, relying on the thermal dynamics of the PAZ domain may also participate in discarding of the passenger strand (Gu *et al.*, 2012; Park and Shin, 2015; Nakanishi, 2016). The final maturation step resides in the 2′O methylation of the 3′ extremity guide of the siRNA by the Hen1 enzyme (Horwich *et al.*, 2007). This methylation step is crucial for the protection of the small guide RNA from 3′ uridylation and further 3′–5′ degradation (Li *et al.*, 2005). The mature RISC, programmed with a guide RNA, functions as a Mg²⁺-dependent, multiple-turnover enzyme that will recognize its mRNA target by perfect base-paired complementarity. The RISC will then slice the mRNA target and release the degradation product in an ATP assisted manner (Hutvagner and Zamore, 2002; Tang *et al.*, 2003; Haley and Zamore,

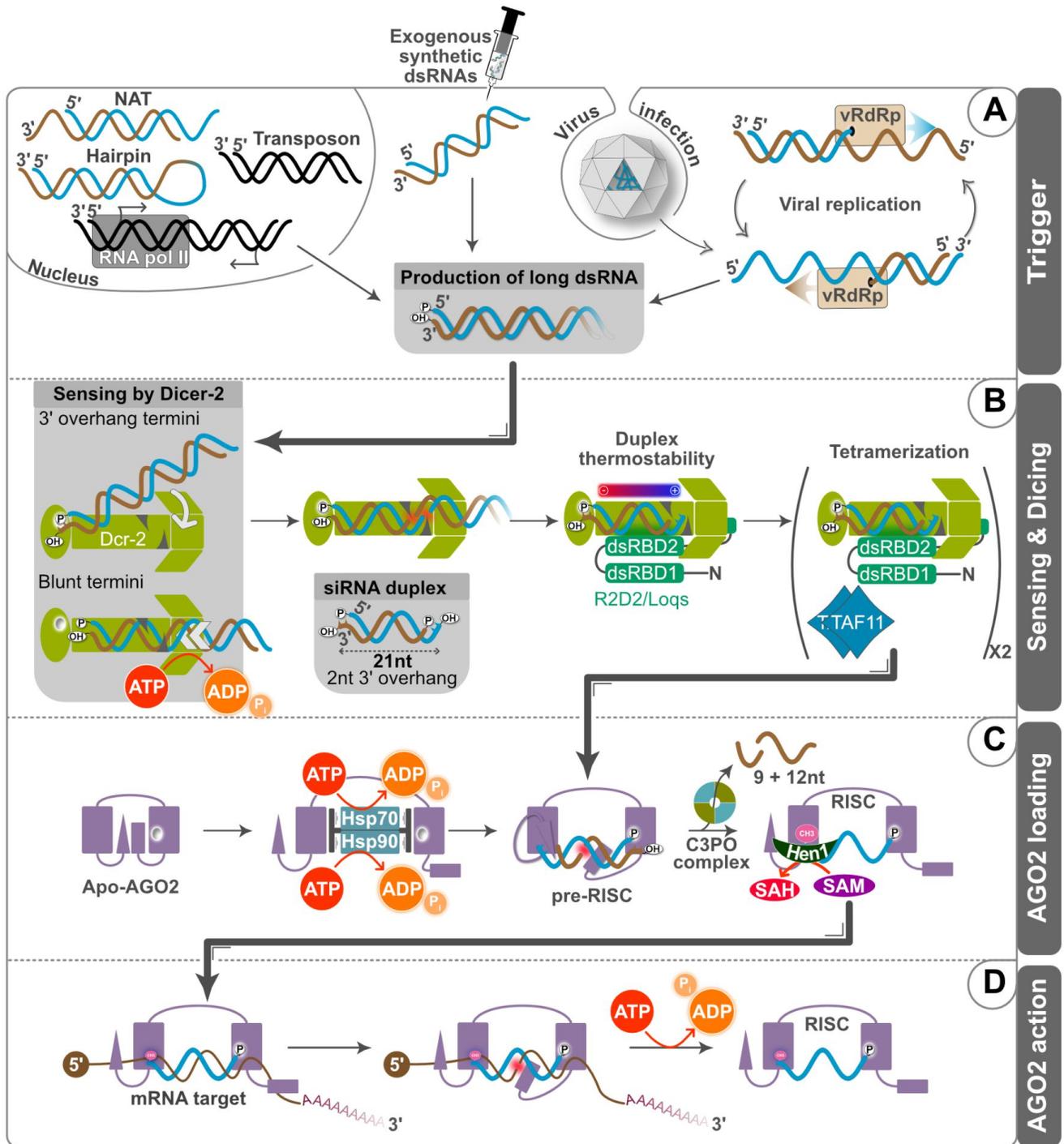


Figure 3.3 The siRNA pathway in *Drosophila melanogaster*. (A) The siRNA pathway can be activated by diverse dsRNA substrates encompassing endogenous transcription of Natural Antisense Transcripts (NATs), transposons, convergent transcripts or structured RNA (hairpins). Experimentally delivered synthetic dsRNAs and virus intermediates of replication represent an exogenous source of dsRNA molecules and triggers RNAi. Blue and brown strands represent RNAs while black strands represent DNA. (B) Endogenous and exogenous long dsRNAs are sensed and diced by Dicer-2. Because Dicer-2 preferentially accesses its substrates at extremities, dsRNA termini are crucial determinants of the mode of action of Dicer-2. Indeed, while blunt end dsRNA triggers an efficient, DRA domain and ATP-dependent processive activity of Dicer-2, a dsRNA molecule with 3' overhang termini promotes a slow, ATP-independent distributive activity. Generated siRNA duplexes are dsRNA molecules with a characteristic 2nt 3' overhang signature and are 21nt long. The relative thermostability of siRNA extremities determines the orientation of the RNA duplex in Dicer-2. (C) Loading of AGO2 relies on a highly coordinated series of events. The closed structure of Apo-AGO2 is opened by two ATP-dependent chaperone complexes (Hsp70 and Hsp90). The siRNA duplex is transferred from the Dicer-2–R2D2–TAF11 complex to the opened form of AGO2, forming the pre-RISC. Maturation of the RISC includes cleavage and removal of the passenger strand. The 3' extremity of the remaining guide strand is 2'-O-methylated. (D) Mature RISC sequence specifically targets and cleaves complementary mRNA targets.

2004; Schwarz *et al.*, 2004) (Fig. 3.3D). In the case of viruses, viral mRNAs seem to be preferentially targeted compared to the viral genome (Marques *et al.*, 2013).

Control of viruses by the siRNA pathway

Genetic evidence in *Drosophila*

The discovery that small RNAs are produced in plants after viral infection predates the characterization of RNAi pathways in flies (Hamilton and Baulcombe, 1999). Evidence that RNAi operates against viruses in *Drosophila* was initially provided in S2 cells infected with Flock House Virus (FHV) (Li *et al.*, 2002). Of note, null mutants for the three main components of the siRNA pathway in flies, namely Dicer-2, AGO2 or R2D2 are homozygous viable (Liu *et al.*, 2003; Lee *et al.*, 2004; Okamura *et al.*, 2004). As a result, adult mutant flies can be infected with viruses and both the survival rate and the viral load can easily be monitored. These experiments established that flies mutant for the siRNA pathway are susceptible to a variety of viruses with RNA or DNA genomes (e.g. Galiana-Arnoux *et al.*, 2006; van Rij *et al.*, 2006; Wang *et al.*, 2006; Mueller *et al.*, 2010; Bronkhorst *et al.*, 2012; Kemp *et al.*, 2013). Interestingly, Dicer-2, R2D2 and AGO2 are among the 3% fastest evolving of all *Drosophila* genes highlighting the high selection pressure on the RNAi pathway (Obbard *et al.*, 2006). Population genomic analysis in multiple invertebrates shows that RNAi genes display a greater rate of adaptive protein substitution than other genes, most likely reflecting their function at the forefront of defence against viruses and transposable elements (TEs) (Palmer *et al.*, 2018).

Control of viruses by the siRNA pathway in other insects

Antiviral RNAi has been investigated in other insects, including vector insects. Many mosquito-borne viruses are associated with human and animal diseases, raising interest in mosquito antiviral immunity (Aguilar *et al.*, 2016; Powers and Waterman, 2017). Over the past 50 years, the *Aedes albopictus* cell line C6/36, isolated from larvae, has been commonly used for amplification of arboviruses but also to study virus–vector interactions.

Recently, the genome of C6/36 cells has been sequenced and null mutations in the *dicer-2* gene were identified, which makes them incompetent for production of siRNAs (Morazzani *et al.*, 2012; Miller *et al.*, 2018). As a result, C6/36 cells support viral replication to high titres, confirming the important antiviral function of Dicer-2 in mosquitoes.

Injection of dsRNA in the body cavity efficiently silences gene expression in a sequence-specific manner and has been used in pioneer experiments to knock-down expression of components of the siRNA pathway in *Aedes aegypti*. This led to significantly increased SINV and DENV titres, without compromising insect survival (Campbell *et al.*, 2008; Sánchez-Vargas *et al.*, 2009; Khoo *et al.*, 2010). Interestingly, in genotype–phenotype association studies, Lambrechts *et al.* (2013) found that the *dicer-2* genotype is associated with resistance to DENV in a virus isolate-specific manner. By contrast, no such association is found for flanking loci, suggesting that the *dicer-2* gene and the siRNA pathway are important determinants of virus suppression in *Aedes*. Mutants of the core RNAi components have subsequently been established in *Ae. aegypti* mosquitoes using genome editing approaches such as TALEN (Transcription Activator-Like Effector Nuclease) or CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) mutagenesis (Basu *et al.*, 2015). In the context of viral infections, *dicer-2* null mutant mosquitoes exhibit a decrease in survival rate compared to wild-type mosquito infection, that correlates with high Yellow fever virus titres (Samuel *et al.*, 2016). The midgut epithelium is a critical barrier for viruses in insect vectors, as it becomes infected after acquiring an infectious blood meal from the host. After successfully replicating in the midgut, viruses reach the haemolymph and disseminate systemically. Viruses then reach the salivary glands from which they are transmitted to a naïve host upon blood feeding. An early study revealed that silencing *dicer-2* in the midgut of female mosquitoes increased infection and dissemination of SINV (Khoo *et al.*, 2010). Insects can also be vectors for plant viruses (Chapter 6). For example, the small brown planthopper (SBPH) is an incompetent vector for Southern rice black streaked dwarf virus (SRBSDV), a plant virus. SRBSDV is restricted in the midgut epithelium of SBPH. Knock-down of

either *dicer-2* or *AGO2* in SBPH results in the dissemination of the virus to the salivary glands and facilitates transmission to rice plants, revealing the importance of the siRNA pathway in the control of vector competence (Lan *et al.*, 2016). However, the siRNA pathway was found to be largely dispensable for antiviral immunity in the midgut of *Anopheles* mosquitoes exposed to the O'nyong-nyong virus, and to become operative only during the systemic stage of infection (Carissimo *et al.*, 2015). Interestingly, a recent study indicates that the siRNA pathway also fails to efficiently silence DENV in the midgut of *Aedes aegypti*, even though the canonical components of the pathway are expressed in this tissue, and that the pathway is functional when triggered by endogenous (e.g. control of TE by endo siRNAs) and exogenous (e.g. intrathoracic injection of long dsRNA) dsRNAs (Olmo *et al.*, 2018). This discrepancy between antiviral and conventional siRNA pathways in the midgut results from the lack of expression in this tissue of Loqs2, an *Aedes*-specific paralogue of Loqs and R2D2. Importantly,

ectopic expression of Loqs2 in the midgut results in restriction of DENV replication and dissemination (Olmo *et al.*, 2018). Altogether, these results point to an additional level of complexity in the insect siRNA pathway when it comes to the control of viruses (see below).

siRNA as a footprint of antiviral immunity

As mentioned above, virus-derived 21–22 nt-siRNAs (vsiRNAs) produced in the course of viral infection in insects, which can be revealed by High Throughput Sequencing (HTS), provide an excellent read-out of Dicer-2 activity (Box 3.1). These vsiRNAs are strongly reduced or abolished in *dicer-2* mutant flies (Aliyari *et al.*, 2008; Mueller *et al.*, 2010; Bronkhorst *et al.*, 2012; Kemp *et al.*, 2013; Marques *et al.*, 2013). For many RNA viruses, vsiRNAs cover the whole viral genome and the ratio between the number of siRNAs matching the (+) strand and the (–) strand of the genome is close to one (Aliyari *et al.*, 2008; Myles *et al.*, 2008;

Box 3.1 High-throughput sequencing

The emergence of next-generation sequencing technologies set a milestone in the development of a tremendous number of high-throughput 'omics' approaches. Until 2010, RNA sequencing methods were all relying on sequencing by synthesis or semi-conductor technologies (second generation). The advent of the third generation of sequencing with PacBio and Oxford Nanopore opened the way to direct DNA and RNA sequencing without any amplification step. These ever-evolving technologies allow the detection of events that can only be bioinformatically predicted with second generation sequencing methods such as splicing or defective genomes in the case of viruses. However, to detect the specific signature of Dicer-2 (siRNA duplexes), small RNA sequencing HTS (second generation) remains the method of choice because of the high-throughput number of reads generated.

	Advantages	Limitations	Main technologies
2 nd generation	<ul style="list-style-type: none"> • High-throughput number of reads (4M – 20B) • Low error rate (~0,1 - 1%) • Low cost 	<ul style="list-style-type: none"> • Library preparation bias including RT, PCR and optional size exclusion steps • Short reads (max ~400bp) • Sample bleeding due to multiplexing (2-5%: check numbers) 	<ul style="list-style-type: none"> • iSeq – NovaSeq 6000 (Illumina) • Ion PGM (Life Technologies) • SOLiD DNA sequencer (Thermo Fisher Scientific) • 454 GS FLX (Roche)
3 rd generation	<ul style="list-style-type: none"> • Direct RNA/DNA sequencing • Long reads (up to 2Mb so far) allowing precise detection of splicing events for instance • Portability 	<ul style="list-style-type: none"> • High error rate (5 – 15%) • Low number of reads (50k – 1M) 	<ul style="list-style-type: none"> • PacBio • Oxford Nanopore

Mueller *et al.*, 2010; Marques *et al.*, 2013; Ferreira *et al.*, 2018), as would be expected from the processing of long dsRNA formed during viral replication. By contrast, in the case of DNA viruses, hotspots of vsiRNAs are observed on specific regions of the viral genome. These vsiRNAs match both strands of the genome, suggesting the siRNA pathway targets regions transcribed on both strands and producing dsRNA (Bronkhorst *et al.*, 2012, 2013; Jayachandran *et al.*, 2012; Kemp *et al.*, 2013). Secondary structures in the genome or antigenome may also be targeted by Dicer-2 in some viruses (Sabin *et al.*, 2013).

Based on *in vitro* experiments, AGO2 has been suggested to impair viral replication by cleaving the viral RNA through its slicer activity (van Mierlo *et al.*, 2012a). *In vivo* experiments further suggest that viral mRNAs are the primary target of vsiRNA-loaded AGO2 (Marques *et al.*, 2013). As genetics are still tricky in most non-model organisms, HTS provides a convenient readout to monitor activity of RNAi pathways in insects infected with viruses (e.g. Chejanovsky *et al.*, 2014; Zografidis *et al.*, 2015; Ferreira *et al.*, 2018). For example, HTS of small RNAs isolated from SINV-infected *Ae. aegypti* mosquitoes revealed the presence of 21 nt-long vsiRNAs and brought the first evidence that the siRNA pathway is activated during viral infection in vector mosquitoes (Myles *et al.*, 2008). This technique has also been successfully used for virus identification in both plants and insects (Kreuze *et al.*, 2009; Wu *et al.*, 2010). Indeed, large fragments of viral genomes can be reconstituted upon assembly of contigs from sequenced vsiRNAs. Furthermore, such contigs are enriched for viral sequences compared to long RNA sequencing reads, because they are by-products of the detection of viral replication by the insect immune system. Thus, HTS of small RNAs can be used to determine the virome of laboratory and wild populations of insects and possibly also other multicellular eukaryotes (Aguiar *et al.*, 2015; Waldron *et al.*, 2018). As such, HTS of small RNAs represents a powerful tool for virus surveillance in populations of vector insects.

Viral suppressors of RNAi (VSRs)

The study of viruses themselves can provide interesting insight into antiviral defence in insects. Indeed, in the course of their interaction with host cells, viruses have evolved to counter antiviral

defence. Thus, many insect viruses, including *Drosophila* viruses, encode viral suppressors of RNAi (VSRs) (Bronkhorst and van Rij, 2014). Some VSRs (e.g. DCV-1A, FHV-B2, IIV6-340R, *Drosophila* X virus-VP3 and *Culex* Y virus-VP3) directly bind long dsRNA through canonical dsRNA binding domains, dsRBDs and prevent processing by Dicer-2 (Li *et al.*, 2002; van Rij *et al.*, 2006; Bronkhorst *et al.*, 2014; van Cleef *et al.*, 2014; see also Fig. 3.4A and B). Interestingly, with the exception of DCV-1A, these VSRs bind siRNA duplexes as well, suggesting that they also inhibit the pathway after long dsRNAs have been processed into siRNAs (Morazzani *et al.*, 2012; Valli *et al.*, 2012; Bronkhorst and van Rij, 2014). Other VSRs (e.g. CrPV-1A and Nora-VP1) bind directly to AGO2 and inhibit its endonuclease activity (Nayak *et al.*, 2010; van Mierlo *et al.*, 2012b). In addition, CrPV-1A also targets AGO2 to the proteasome through the K48 polyubiquitination pathway (Nayak *et al.*, 2018). This is reminiscent of the mode of action of P0, a VSR from poleroviruses, which triggers degradation of AGO1 in plant cells (Baumberger *et al.*, 2007; Bortolamiol *et al.*, 2007).

The importance of VSRs for viruses has been particularly well illustrated in the case of FHV-B2. *Nodaviridae* have small bipartite RNA genomes that are easy to manipulate genetically. One segment of the genome, RNA1, encodes the replicase, whereas the second, RNA2, encodes the capsid proteins. A third RNA transcript, RNA3, is also produced from RNA1 once it has replicated and encodes the VSR B2 (Chao *et al.*, 2005). Whereas wild-type FHV is highly pathogenic upon injection into the body cavity of flies, viral mutants unable to express B2 are completely attenuated (Galiana-Arnoux *et al.*, 2006; Wang *et al.*, 2006; Fig. 3.4C). As expected, the virus regains virulence when injected into *dicer-2* or AGO2 mutant flies (Han *et al.*, 2011; Petrillo *et al.*, 2013). Similarly, mutation of residue Phe 114 into Ala in the flexible loop of CrPV-1A, which is involved in the interaction with AGO2 results in an attenuated virus in wild type flies but not in AGO2 mutant flies (Nayak *et al.*, 2018). Of note however, even in RNAi mutants, the B2 deficient virus exhibits reduced virulence compared to wild-type FHV. Indeed, an additional function of B2 is to bind double stranded regions of RNA2, thus preventing its recruitment into poorly characterized cytoplasmic RNA granules where its

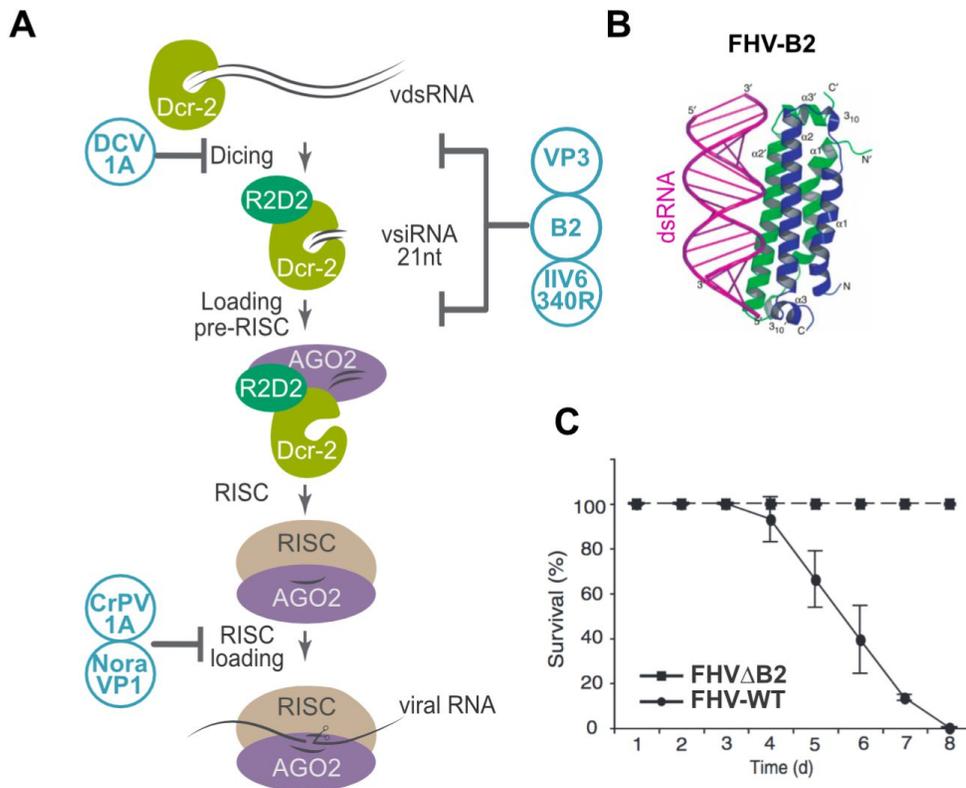


Figure 3.4 Viral suppressors of RNAi neutralize the siRNA pathway in insects. (A) Schematic representation of the action of a set of insects VSRs. (B) Crystal structure of FHV-B2 dimer associated with dsRNA (figure from Chao *et al.*, 2005). (C) Survival curve of flies injected with WT FHV or a mutant version deleted for B2 (adapted from Galiana-Arnoux *et al.*, 2006).

translation would be repressed. As a result, translation of the capsid protein is impaired in the absence of B2, even in RNAi deficient cells (Petrillo *et al.*, 2013). Interestingly, the FHV-B2 VSR can be used to neutralize the siRNA pathway upon ectopic expression. For example, when *Ae. aegypti* mosquitoes were challenged with SINV recombined with FHV-B2 (SINV-B2) either by injection or infected blood meal, the virus titre was increased compared to the control virus. The recombinant SINV-B2 caused high mortality among the mosquitoes at 4–6 days post-infection, highlighting that the RNAi pathway is essential to control arboviruses replication (Myles *et al.*, 2008; Cirimotich *et al.*, 2009). In *Ae. aegypti* mosquitoes, the constitutive and ubiquitous expression of FHV-B2 impaired the siRNA pathway and enhanced replication of both SINV and DENV (Khoo *et al.*, 2013).

One consequence of the direct interaction of some VSRs with protein factors of the siRNA pathway to alter their ability to neutralize viral RNAs or trigger their degradation (Singh *et al.*, 2009; Nayak *et al.*, 2010, 2018; van Mierlo *et al.*, 2012b, 2014) is that both VSRs and components of the siRNA

pathway evolve rapidly (Obbard *et al.*, 2006). As a result, the activity of viral suppressors can be host-specific. For example, the VP1 protein from a divergent Nora virus isolated from *D. immigrans* interacts with and suppresses *D. immigrans* AGO2, but not *D. melanogaster* AGO2 (van Mierlo *et al.*, 2014). This provides an excellent example for the co-evolution of the host RNAi machinery and viral suppression mechanisms. Because of their intimate association with key components of the siRNA pathway, VSRs provide promising tools to decipher the regulation and molecular mechanisms of antiviral RNAi. For example, development of single-molecule approaches can shed light on VSR mode of action and how they discriminate viral from cellular RNA (Fareh *et al.*, 2018). VSRs may also be used to visualize and track dsRNA in live plant and animal cells, as recently shown for an FHV-B2–GFP fusion protein (Monsion *et al.*, 2018).

Systemic RNAi in insects?

In both plants and *C. elegans*, systemic RNAi contributes to the control of viral infections. The

mechanism at play involves spreading of siRNAs generated in infected cells to neighbour healthy cells. There, these siRNAs prime the synthesis of dsRNAs by host-encoded RNA-dependent RNA polymerases (RdRPs) (Ding, 2010). Insect genomes do not encode such RdRPs and clonal analyses in *Drosophila* revealed that the siRNA pathway is a cell autonomous pathway (Roignant *et al.*, 2003). Nevertheless, the spread of antiviral RNAi has been proposed to contribute to the control of viral infections in insects. Indeed, injection of exogenous dsRNA in the body cavity of most insects, or even feeding dsRNA, leads to gene silencing through the siRNA pathway (reviewed in Zotti *et al.*, 2018). Thus, viral infection, which is known to trigger transcriptional responses, may induce a mechanism of systemic RNAi. In *Drosophila*, uptake of exogenous dsRNA is mediated by the endocytic pathway (Saleh *et al.*, 2006). Nanotube like structures have also been reported to transfer dsRNA and components of the RNAi machinery between cells (Karlikow *et al.*, 2016). Thus, dsRNA released from infected dying cells may trigger RNAi in distant cells, upon internalization by the dsRNA uptake pathway (Saleh *et al.*, 2009). Surprisingly, however, infection of flies with a sublethal dose of DCV, which should prime the antiviral siRNA pathway, did not induce protection against a challenge with a lethal dose of virus (Longdon *et al.*, 2013). Of note, viral RNA can be reverse transcribed into DNA (vDNA) in *Drosophila* and in mosquitoes, through the action of the reverse transcriptase from transposable elements (TEs) (Goic *et al.*, 2013, 2016), acting together with Dicer-2 (Poirier *et al.*, 2018) or AGO2 (Tassetto *et al.*, 2017). Transcription of vDNA has been proposed to result in the production of secondary siRNAs bearing a 5'-triphosphate mark, associated with systemic antiviral effect (Tassetto *et al.*, 2017). However, the existence of such secondary siRNAs was not confirmed by an independent study (Mondotte *et al.*, 2018). Overall, the mechanisms involved in systemic antiviral RNAi remain to be characterized and genetic evidence for the importance of the contribution of vDNA and secondary siRNAs in antiviral immunity is still lacking.

In summary, it now appears that the antiviral siRNA pathway in insects is more complex than previously thought. A number of host and viral factors, which affect the stability, the binding affinity

to viral nucleic acids, and even the cytoplasmic localization of ribonucleoprotein complexes mediating RNA silencing can influence antiviral defence. Biochemistry, genetics and live imaging will clarify the mode of action of the components of the siRNA pathway in the context of infected cells, in different types of tissues and in various insects.

The piRNA pathway in antiviral immunity

Production of primary and secondary piRNAs

Another small RNA pathway, the piRNA pathway is able to sense foreign nucleic acids and has been proposed to participate in antiviral immunity (Miesen *et al.*, 2016a). This pathway involves 24–30 nt small RNAs and Argonaute proteins from the PIWI clade. The founding member of this subfamily, PIWI, was initially characterized as a *Drosophila* gene essential for male fertility (Lin and Spradling, 1997). This phenotype results from derepression and mobilization of transposable elements in the germline. It was subsequently found that PIWI works with two other Argonaute proteins from the same clade, namely Aub and AGO3 (Saito *et al.*, 2006; Vagin *et al.*, 2006; Brennecke *et al.*, 2007; Gunawardane *et al.*, 2007). The PIWI-mediated mechanism of genome maintenance in the germline is conserved in all animals. The piRNAs were initially identified in mouse testis and in *Drosophila*, where they were first called repeat-associated small interfering (rasi)RNAs. They interact with PIWI proteins (Aravin *et al.*, 2006; Girard *et al.*, 2006; Grivna *et al.*, 2006; Vagin *et al.*, 2006). *Drosophila* genetics coupled with HTS, have shed light on the molecular mechanism generating piRNAs. This involves a primary processing pathway that primes the production of phased piRNAs during the so-called ping-pong amplification mechanism (Czech and Hannon, 2016; Fig. 3.5). Of note, HTS of cell lines or mosquitoes revealed the existence of virus-derived piRNAs (Wu *et al.*, 2010; Hess *et al.*, 2011; Morazzani *et al.*, 2012; Vodovar *et al.*, 2012; Léger *et al.*, 2013; Aguiar *et al.*, 2015).

The primary processing pathway targets Pol II-dependent transcripts generated from genomic loci rich in transposon remnants known as piRNA cluster (Bucheton, 1995; Brennecke *et al.*, 2007;

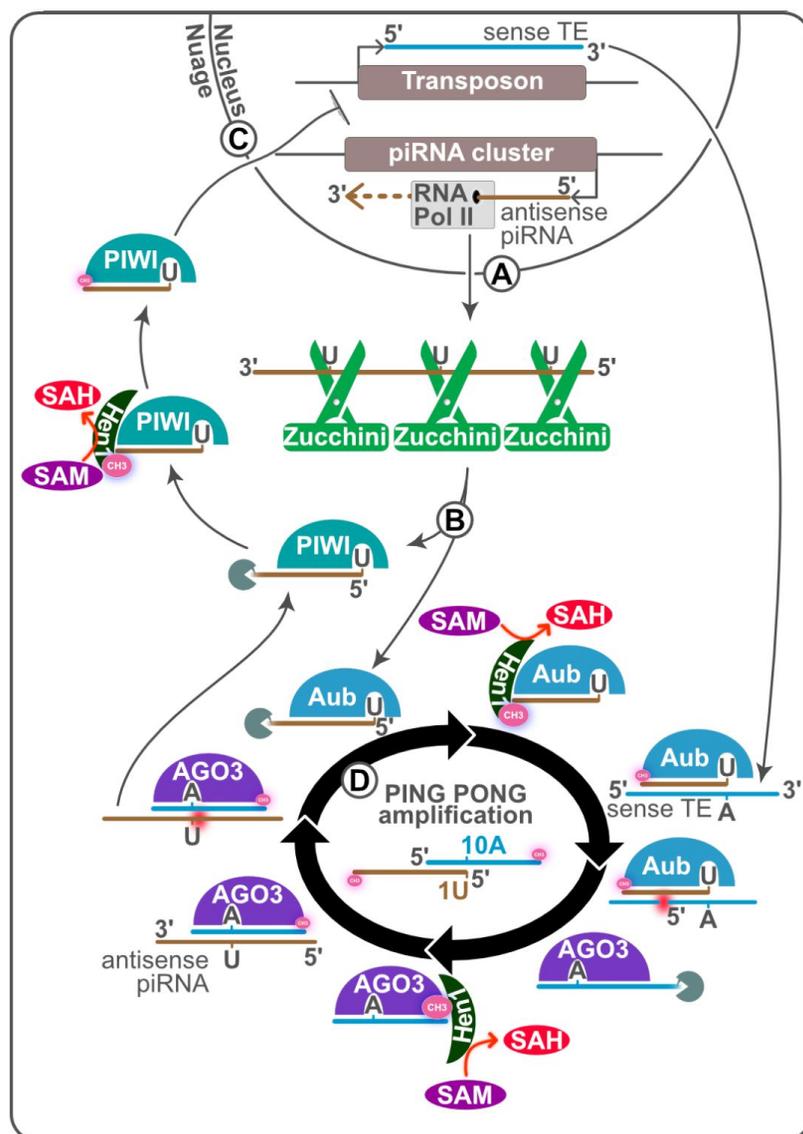


Figure 3.5 The piRNA pathway in *Drosophila* ovaries. piRNA clusters are mainly composed of defective transposon copies integrated in the genome and serve as a genetic memory of transposon exposition in a population. Transcription of these genomic clusters by the host RNA polymerase II generates long RNAs that are exported in the cytoplasm (A). There, they are processed by Zucchini, an endonuclease with a strong preference for cleavage 5' of a uridine. This leads to the production of phased piRNAs, which can either be loaded on PIWI or on Aub (B). PIWI- and Aub-loaded piRNAs undergo trimming and 2'O methylation of their 3' extremity. piRNA-loaded PIWI translocates to the nucleus where it participates in transcriptional silencing of transposons through deposition of repressive histone modifications (C). On the other hand, piRNA-loaded Aub will initiate the Ping-Pong amplification cycle (D). Briefly, loaded Aub will recognize a complementary transposon transcript and induce endonucleolytic slicing of the target between nucleotide 10 and 11 of the piRNA. This slicing generates the 5' end of a new sense piRNA with a 10nt 5' overlap with the initial antisense piRNA and an adenosine residue at position 10. This newly formed piRNA is loaded on AGO3, trimmed and 2'O methylated at its 3' extremity. Finally, piRNA-loaded AGO3, using a similar mechanism, will generate Ago-bound piRNAs from piRNA clusters.

Péligon *et al.*, 2007). These precursors are processed by the endonuclease Zucchini, which preferentially cleaves 5' of a uridine residue. As a result, the piRNA intermediates produced are enriched for 5' uridine residues (1U) (Pane *et al.*, 2007; Han *et al.*, 2015; Mohn *et al.*, 2015; Fig. 3.5A). In fruit flies, these piRNA intermediates are

loaded onto PIWI and Aub in an electron dense perinuclear region of the germline called 'nuage'. Indeed, the binding pocket of the MID domain of these two proteins preferentially accommodates 5' uridine residues (Cora *et al.*, 2014; Matsumoto *et al.*, 2016; Fig. 3.5B). piRNA-loaded PIWI translocates to the nucleus where it participates in transcriptional

silencing of transposons through deposition of repressive histone modifications (Sienski *et al.*, 2012; Dönertas *et al.*, 2013; Le Thomas *et al.*, 2013; Ohtani *et al.*, 2013; Sienski *et al.*, 2015; Yu *et al.*, 2015; Fig. 3.5C). By contrast, piRNA-loaded Aub remains in the nuage and initiates the ping-pong amplification mechanism (Fig. 3.5D). The Aub pi-RISC recognizes and cleaves complementary transposon mRNAs. The resulting cleavage product corresponds to the precursor of a secondary piRNA and associates with AGO3 (Brennecke *et al.*, 2007; Gunawardane *et al.*, 2007; Lim and Kai, 2007). Of note, because cleavage mediated by enzymes of the PIWI clade occurs specifically between nucleotide 10 and 11, AGO3 bound piRNAs are enriched for adenosine residues in position 10 (10A). piRNAs loaded into AGO3 target and cleave antisense piRNA precursors, thus generating the 5' end of new sense piRNAs and resulting in the ping-pong amplification cycle (Fig. 3.5D). In summary, the combination of slicer activity of PIWI proteins together with the activity of endo- and exonucleases explains the Dicer-independent production of 24–30 nt long piRNAs. Of note, the 1U/10A signature characteristic of the ping-pong amplification provides a convenient way to monitor the activity of this pathway by HTS and was instrumental in revealing the existence of virus-derived piRNAs.

Virus-derived piRNA

Transposable elements share with viruses the property of being selfish genetic units encoding proteins that enable their proliferation and spread. Hence, one can wonder whether the piRNA pathway also participates in antiviral immunity. Indeed, transposable elements are targeted by the siRNA pathway in *Drosophila* somatic tissues as revealed by accumulation of siRNAs matching transposable elements in HTS analysis (Chung *et al.*, 2008). Conversely, virus-derived piRNAs could be observed in the OSS cell line derived from *Drosophila* ovarian tissue (Wu *et al.*, 2010). However, in *Drosophila*, activity of the piRNA pathway is restricted to the germline and neither genetics nor HTS support an antiviral function of the piRNA pathway in *Drosophila* (Petit *et al.*, 2016; van den Beek *et al.*, 2018). Nonetheless, there are significant differences between piRNA pathways in *Drosophila* and other insects. Indeed, a recent study investigating 20 species across the

arthropod phylum revealed that TEs are commonly targeted by somatic piRNAs unlike in *Drosophila* (Lewis *et al.*, 2018). Consistent with this observation, the piRNA pathway components differ between insect species. Notably, it is apparent that the PIWI clade has significantly expanded in *Ae. aegypti* mosquitoes, where it contains seven PIWIs (instead of two in *Drosophila*, Piwi and Aub) and one AGO3 (reviewed in Miesen *et al.*, 2016a). This, together with identification of virus derived piRNAs in mosquito Aag2 and C6/36 cell lines, led to the suggestion that the piRNA pathway could form a second layer of antiviral defence in mosquitoes (Wu *et al.*, 2010; Hess *et al.*, 2011; Morazzani *et al.*, 2012; Vodovar *et al.*, 2012; Léger *et al.*, 2013; Aguiar *et al.*, 2015). Depletion in AGO3 and PIWI5 resulting in the decreased production of viral piRNAs revealed that these enzymes mediate recognition and processing of SINV RNAs, although the functional consequence of this processing on viral replication was not reported (Miesen *et al.*, 2015). A subsequent report focusing on DENV indicated that the knockdown of PIWI proteins did not significantly affect viral RNA levels (Miesen *et al.*, 2016b). An independent study proposed a role for PIWI4 in antiviral immunity in Aag2 cells, an *Aedes aegypti* derived cell line. Interestingly, PIWI4 behaves as an atypical member of the PIWI family and is not involved in piRNA production but associates with the siRNA pathway core components AGO2 and Dicer-2 (Varjak *et al.*, 2017a,b, 2018). This protein may reveal a cross-talk between piRNA and siRNA pathways in the context of viral infections. The presence of virus derived somatic piRNAs in other arthropods was tested by Jiggins and colleagues in their landmark paper (Lewis *et al.*, 2018) and virus derived siRNAs could be identified in 9 of the 20 species investigated. Among them, 5 species also produced 24–30 nt 5'U biased small RNAs derived from viruses. However, only in *Aedes aegypti* did these piRNAs bare the 1U/10A signature of ping-pong amplification. In the four other species, piRNAs mapped to one strand only, similar to primary piRNAs (Lewis *et al.*, 2018). Therefore, the predominant role of siRNAs compared to piRNAs in antiviral defence observed in *Drosophila* and *Lutzomyia* is probably relevant across arthropods, with *Aedes* mosquitoes representing a notable exception (Ferreira *et al.*, 2018).

Sensing viral RNA in insects

Viral nucleic acid sensors

In mammals a major molecular pattern associated with viral infection is long dsRNA, generated as a by-product of viral replication. The Toll receptor TLR3 is localized in the endosome compartment and is activated upon binding of long dsRNA (Kawai and Akira, 2011). This receptor probes the content of the endosomes for the presence of endocytosed signs of viral infection in the extracellular milieu. Cross-linking of two TLR3 subunits by dsRNA triggers TRIF-dependent interferon activation. Of note, two other endosomal TLRs, TLR7 and TLR8 detect UG rich short ssRNA fragments (Maeda and Akira, 2016). In the cytosol another family of pattern recognition receptors, the RIG-like receptors, sense viral RNA (Goubau *et al.*, 2013). As mentioned above, a hallmark of these receptors is the presence of a DRA domain phylogenetically related to the Dicer proteins (Paro *et al.*, 2015). This domain is followed by a C-terminal domain (CTD), which participates in RNA binding. In addition, two of the three RLRs, RIG-I and MDAS contain amino-terminal caspase recruitment domains (CARDs) (Kawai and Akira, 2011). This enables them to recruit the signalling adaptor MAVS and to activate expression of interferon genes. Although the three RLRs can bind dsRNA *in vitro*, other molecular features found on viral RNA are necessary to activate RIG-I. Indeed, the CTD of RIG-I, which forms a tighter pocket than the one from MDAS, detects the presence of 5' di- or triphosphate at the extremity of viral RNAs (Kowalinski *et al.*, 2011; Goubau *et al.*, 2014). This provides an efficient means to discriminate between capped cellular mRNAs and uncapped viral RNAs. This biochemical distinction is supported by genetic data that point to non-redundant functions of RIG-I and MDAS. Indeed, RIG-I mutant mice have impaired interferon responses following infection by viruses such as influenza, VSV, or Japanese encephalitis virus (Kato *et al.*, 2005, 2006). Of note, most of these viruses have ssRNA genomes of negative polarity and do not produce detectable amounts of long dsRNA in infected cells (Weber *et al.*, 2006). By contrast, RIG-I mutant mice respond normally to picornaviruses. The 5' end of the genome and antigenome strand of these viruses is covalently linked to a VpG protein, which prevents

recognition of the termini by RIG-I. On the other hand, MDAS mutant mice are highly susceptible to picornaviruses, which generate large quantities of dsRNA (Kato *et al.*, 2006). The third RLR, LGP2, functions together with MDAS (Deddouche *et al.*, 2014).

In insects, the only viral nucleic acid sensor identified so far is Dicer-2, suggesting that long dsRNA is the major molecular pattern used to detect viral infection (Fig. 3.3A). Importantly, even in the case of the negative strand RNA virus VSV, the profile of virus-derived siRNAs reveals a typical long dsRNA signature, with siRNAs covering in equal amounts the whole length of both genome and antigenome strands (Ferreira *et al.*, 2018; Marques *et al.*, 2013; Mueller *et al.*, 2010). However, sensing of viral RNA by Dicer-2 appears not to be limited only to the recognition of dsRNA.

Sensing viral RNA by Dicer-2

In vitro dicing assays using recombinant versions of Dicer-2 with or without its cofactors incubated with diverse dsRNA showed that Dicer-2 preferentially accesses its substrate at the extremities (Sinha *et al.*, 2015). Furthermore, dsRNA termini are crucial determinants of Dicer-2 mode of action *in vitro* (Fig. 3.3B). Indeed, while blunt end dsRNA triggers an efficient, DRA domain and ATP-dependent processive activity of Dicer-2, a dsRNA molecule with 3' overhang termini promotes a slow, ATP-independent distributive activity (Cenik *et al.*, 2011; Welker *et al.*, 2011). Cryo-EM studies revealed that dsRNAs presenting 3' overhangs are repeatedly bound by Dicer-2 platform-PAZ domain to be sequentially cleaved while blunt dsRNAs are threaded through the DRA domain in an ATP-dependent manner and successively diced to generate many phased siRNAs (Sinha *et al.*, 2018a). This distinct mode of dicing is driven by an extensive conformational change of the protein upon binding of a blunt dsRNA extremity, which cannot occur if the extremity harbours a 3' overhang (Fig. 3.3B). Loqs-PD interacts with the DRA domain of Dicer-2 in an RNA independent manner through its C-terminal FDF motif and allows the enzyme to process RNA substrates normally refractory to cleavage, such as dsRNA with blocked, structured or frayed ends (Sinha *et al.*, 2015; Trettin *et al.*, 2017). In summary, *in vitro* experiments with purified recombinant proteins point to a model

where subtle changes in the substrate result in tremendous differences in the dicing mechanism. How this model can be reconciled with the complexity of Dicer-2 natural substrates in the cellular context is an important challenge for future studies. This is particularly relevant in the context of viral infection as the RNA extremities of viruses and viral replication complexes are highly variable. A striking example is the case of picornaviruses and dicistroviruses, which, as mentioned above, display a covalently linked VpG protein at the 5' extremities of their genome and antigenome (Virgen-Slane *et al.*, 2012). This covalent modification at the extremities of the genome from *Dicistroviridae* such as DCV or CrPV is expected to impact sensing by Dicer-2. One asset to solve this important question is the characteristic siRNA signature of Dicer-2, which is amenable to bioinformatic analysis following small RNA HTS. Bioinformatic analysis of the pattern of vsiRNAs produced in wild-type or mutant flies (e.g. inactivated ATP binding site in the DRA domain of Dicer-2) may provide insights on an alternative access point of Dicer-2 on DCV or CrPV RNA (see Box 3.1) (e.g. Aliyari *et al.*, 2008; Mueller *et al.*, 2010; Bronkhorst *et al.*, 2012; Sabin *et al.*, 2013; Aguiar *et al.*, 2015).

Another indication that features other than double strandedness are sensed lies in the importance of the cofactors acting together with Dicer-2. For example, Loqs-PD is not required for silencing viral RNA, in spite of its essential role in RNA interference triggered by endogenous or *in vitro* synthesized dsRNA (Marques *et al.*, 2010, 2013). This points to the existence of differences between viral RNAs and other dsRNAs, produced from endogenous sources or synthesized *in vitro*. Of note, similar observations were made in the nematode *Caenorhabditis elegans*, where a RIG-I-like factor known as DRH1 is essential for antiviral RNAi yet dispensable for the other silencing pathways (Ashe *et al.*, 2013; Guo *et al.*, 2013), but also in *Ae. aegypti* mosquitoes. As mentioned before, the siRNA pathway silences targets of endogenous and exogenous dsRNAs in the midgut of these mosquitoes but fails to suppress viruses (Olmo *et al.*, 2018). These results confirm the existence of differences between silencing triggered by exogenous or endogenous dsRNA and viral RNA, with the latter specifically requiring the dsRBP Loqs2. We note that an isoform of Staufen that evolved in coleopterans is

another example of a species-specific dsRBP that regulates the siRNA pathway (Yoon *et al.*, 2018). The fact that *Drosophila* (Loqs-PD), *Aedes* (Loqs2) and coleopteran insects (StaufenC) all require specific cofactors to define the activity of siRNA pathway is intriguing and highlights that an important facet of this pathway remains ill characterized.

The discrimination between the dsRNA precursors of endo- or exo-siRNAs and the dsRNA generated during viral infection may reflect differences either in the receptor complex sensing dsRNAs or in the viral RNA itself. On the protein side, it will be important to characterize biochemically the role of Loqs and R2D2 proteins in *Drosophila* and *Aedes*. On the RNA side, the pervasiveness of RNA modifications begs the question of how much they contribute to the discrimination of viral RNA by the siRNA pathway (Gokhale and Horner, 2017; Helm and Motorin, 2017; Fig. 3.6). For example, ADAR is an RNA editing enzyme catalysing the conversion of A to I within dsRNA regions of cellular RNAs, which prevents unwanted activation of immunity in mice and *C. elegans* (Liddicoat *et al.*, 2015; Reich *et al.*, 2018). In mammals, the presence of epitranscriptomic marks such as 2'O-methylation on viral RNA affects sensing by TLR7 or MDA5 whereas N6-methyladenosine (m6A) in dsRNA reduces activation of TLR3 (Karikó *et al.*, 2005; Gonzales-van Horn and Sarnow, 2017). Therefore, it will be interesting to investigate the presence of post transcriptional modifications in viral RNAs in insect cells and their impact on antiviral RNAi. This will be particularly interesting in the case of flaviviruses such as dengue and Zika viruses, which have been shown to contain m6A modified nucleosides when grown in mammalian cells (Gokhale *et al.*, 2016; Fig. 3.6).

Future directions

A great deal of progress was made over the past decade on the genetic characterization of antiviral innate immunity in insects. Yet, lots of questions remain, paramount among them the identification of the receptors that sense viral infections. Among insects, the fruit fly *D. melanogaster* offers a fantastic model for unbiased, large scale, mutagenesis screens (Wieschaus and Nüsslein-Volhard, 2016). Indeed, such screens were instrumental in defining the components of the IMD and Toll pathways

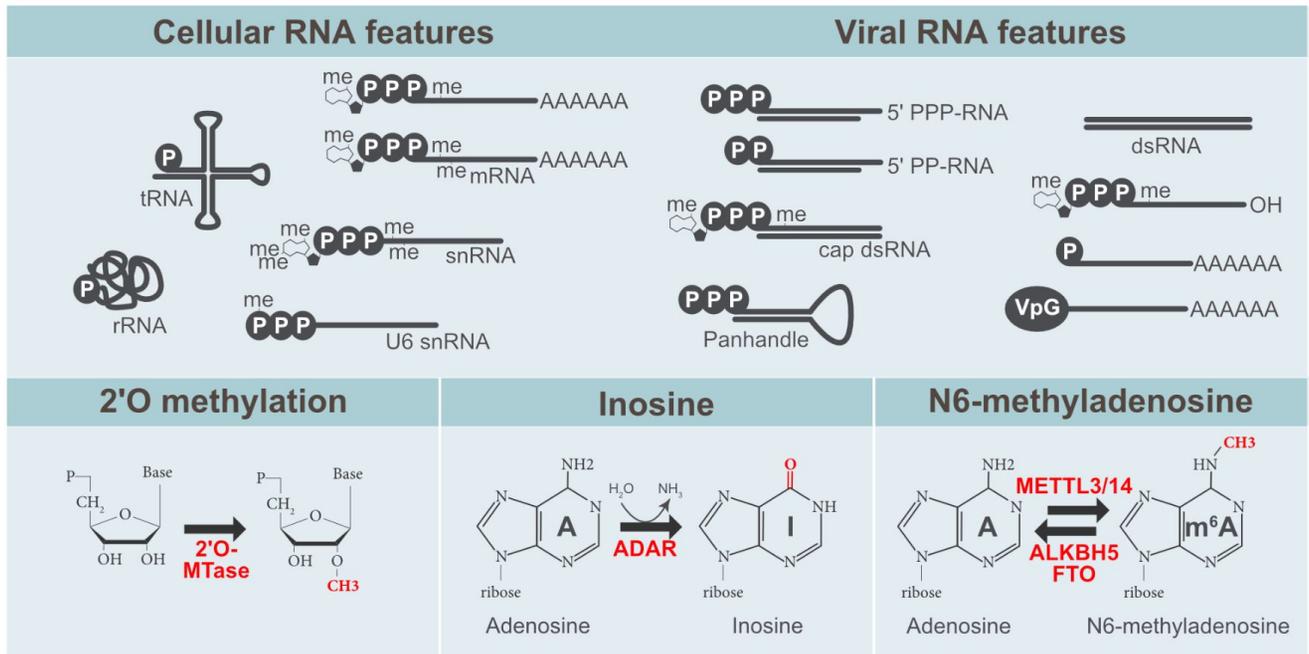


Figure 3.6 Overview of cellular and viral RNAs. Schematic representation of self and non-self RNAs adapted from Gebhardt *et al.* (2017). 2'O methylation, Inosine and N6-methyladenosine are the three main RNA modifications described on viral RNAs affecting their recognition by the innate immune system.

of innate immunity, and in identifying the PRRs activating them in response to bacterial or fungal infections (e.g. Leulier *et al.*, 2000; Rutschmann *et al.*, 2000a,b; Lu *et al.*, 2001; Vidal *et al.*, 2001; Choe *et al.*, 2002; Gottar *et al.*, 2002, 2006). Genetic screens are however time consuming, especially when viruses must be injected into the body cavity of the flies (Merkling and van Rij, 2015). This limitation could, to some extent, be bypassed by using natural infections with fly pathogens, although this would require production of large amounts of virus and the control of viral uptake would be challenging. Alternatively, transgenic viral replicons expressing fluorescent proteins could be used as a proxy for infection, with the caveat that critical steps of the viral cycle (binding, entry, uncoating, assembly and budding of viral particles) would be bypassed (Avadhanula *et al.*, 2009; Wernet *et al.*, 2014). The recent advances in mass spectrometry (MS) technologies provide other opportunities to decipher antiviral innate immunity in insects. These methods can be used to further characterize the antiviral siRNA pathway by defining the interactome of the canonical components Dicer-2, R2D2 and AGO2, but also to identify PRRs sensing nucleic acids.

The genetic characterization of antiviral RNAi in both *Drosophila* and *Aedes* mosquitoes points

to differences between the siRNA-dependent response triggered by dsRNA of endogenous (endo-siRNA pathway) or exogenous (exo-siRNA pathway) origin on one hand, and the antiviral siRNA pathway on the other (Marques *et al.*, 2013; Olmo *et al.*, 2018). Cell imaging of the canonical components Dicer-2, AGO2, R2D2 and Loqs in the context of cells treated with exogenous dsRNA or infected with viruses is likely to reveal differences in the dynamics of these proteins. This information can subsequently be used to define the interactome of the siRNA pathway at critical steps of the infection cycle, to identify novel regulatory co-factors of the pathway participating in the sensing or neutralization of viral RNAs. Of note, this approach requires synchronized infections, and will have to be carried out in tissue culture cells, a caveat considering the possible existence of tissue specific regulators (e.g. Olmo *et al.* 2018).

MS can also be used to identify host proteins that sense nucleic acids of viral origin. Some foreign nucleic acids bear specific marks absent in cellular nucleic acids, such as, in the case of RNA, long double strandedness, the presence of 5' triphosphates or missing methylation marks (reviewed in Habjan and Pichlmair, 2015; Gebhardt *et al.*, 2017). Protein binding to these marked, viral-like nucleic acids can be identified

by nucleic acid affinity purification coupled to MS-based identification of proteins. For example, such an approach in mammals resulted in the identification of the host proteins IFIT1 and AIM2 as viral RNA binding protein with antiviral function and a DNA sensor of the innate immune system, respectively (Bürckstümmer *et al.*, 2009; Pichlmair *et al.*, 2011). Applied to insect cell lines or whole animals, this strategy, coupled to RNAi or CRISPR/Cas9 functional screens may reveal novel nucleic acid binding proteins involved in antiviral immunity in insects.

Finally, a fascinating question for future work pertains to the sensing of DNA in insect cells. Indeed, activation of the IMD pathway in flies mutant for the enzyme DNaseII (significant but moderate compared to a bacterial infection) suggests that cytosolic DNA can activate an NF- κ B-dependent pathway in insect cells (Mukae *et al.*, 2002; Liu *et al.*, 2012). In mammals, a dedicated pathway, the cGAS-STING pathway, activates synthesis of interferons upon detecting cytosolic DNA (Hornung *et al.*, 2014). So far, the data available indicate that viral DNA produces dsRNAs and activation of the siRNA pathway, akin to the sensing of herpes virus derived dsRNA by TLR3 in mammals (Tabeta *et al.*, 2004; Zhang *et al.*, 2007; Bronkhorst *et al.*, 2012; Kemp *et al.*, 2013). Yet, important questions remain. For example, is viral DNA, which is expected to be transcribed in the nucleus possibly by a host RNA polymerase, activating the endo-siRNA pathway (Loqs-PD dependent) or the antiviral siRNA pathway (Loqs-PD independent)? Is there another pathway sensing DNA in the cytosol, as in mammals? The cGAS-like enzymes identified in insect genomes so far do not contain the zinc ribbon motif mediating interaction with DNA in mammalian cGAS (Wu *et al.*, 2014; Margolis *et al.*, 2017). However, the recent discovery that STING carries immune functions in some insects, in particular against DNA viruses in *Bombyx mori*, raises the possibility that dedicated receptors operate in the cytosol of insect cells to sense DNA (Goto *et al.*, 2018; Hua *et al.*, 2018; Liu *et al.*, 2018; Martin *et al.*, 2018). Characterization of the ligands activating the insect STINGs now at hand, which do not appear to bind cyclic dinucleotides (Kranzusch *et al.*, 2015), should help to identify these receptors.

Acknowledgements

We thank Ludmila dos Santos Silva and Joao Marques for critical reading of the manuscript and insightful suggestions. Work in our laboratory is supported by the CNRS, the ANR, the Fondation ARC and the Programme Investissements d'Avenir (Labex NetRNA and Equipex I2MC). LT acknowledges financial support from the Hoffmann Infinitus Program.

References

- Aguiar, E.R., Olmo, R.P., Paro, S., Ferreira, F.V., de Faria, I.J., Todjro, Y.M., Lobo, F.P., Kroon, E.G., Meignin, C., Gatherer, D., *et al.* (2015). Sequence-independent characterization of viruses based on the pattern of viral small RNAs produced by the host. *Nucleic Acids Res.* 43, 6191–6206. <https://doi.org/10.1093/nar/gkv587>
- Aguiar, E.R., Olmo, R.P., and Marques, J.T. (2016). Virus-derived small RNAs: molecular footprints of host-pathogen interactions. *Wiley Interdiscip. Rev. RNA* 7, 824–837. <https://doi.org/10.1002/wrna.1361>
- Aliyari, R., Wu, Q., Li, H.W., Wang, X.H., Li, F., Green, L.D., Han, C.S., Li, W.X., and Ding, S.W. (2008). Mechanism of induction and suppression of antiviral immunity directed by virus-derived small RNAs in *Drosophila*. *Cell Host Microbe* 4, 387–397. <https://doi.org/10.1016/j.chom.2008.09.001>
- 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.
- Asad, S., Parry, R., and Asgari, S. (2018). Upregulation of *Aedes aegypti* Vago1 by *Wolbachia* and its effect on dengue virus replication. *Insect Biochem. Mol. Biol.* 92, 45–52.
- Ashe, A., BÉlicard, T., Le Pen, J., Sarkies, P., Frézal, L., Lehrbach, N.J., Félix, M.A., and Miska, E.A. (2013). A deletion polymorphism in the *Caenorhabditis elegans* RIG-I homolog disables viral RNA dicing and antiviral immunity. *Elife* 2, e00994. <https://doi.org/10.7554/eLife.00994>
- Avadhanula, V., Weasner, B.P., Hardy, G.G., Kumar, J.P., and Hardy, R.W. (2009). A novel system for the launch of alphavirus RNA synthesis reveals a role for the Imd pathway in arthropod antiviral response. *PLOS Pathog.* 5, e1000582. <https://doi.org/10.1371/journal.ppat.1000582>
- Barribeau, S.M., Sadd, B.M., du Plessis, L., Brown, M.J., Buechel, S.D., Cappelle, K., Carolan, J.C., Christiaens, O., Colgan, T.J., Erler, S., *et al.* (2015). A depauperate immune repertoire precedes evolution of sociality in bees. *Genome Biol.* 16, 83. <https://doi.org/10.1186/s13059-015-0628-y>
- Basu, S., Aryan, A., Overcash, J.M., Samuel, G.H., Anderson, M.A., Dahlem, T.J., Myles, K.M., and Adelman, Z.N. (2015). Silencing of end-joining repair for efficient site-specific gene insertion after TALEN/CRISPR mutagenesis in *Aedes aegypti*. *Proc. Natl. Acad. Sci.*

- U.S.A. 112, 4038–4043. <https://doi.org/10.1073/pnas.1502370112>
- Baumberger, N., Tsai, C.H., Lie, M., Havecker, E., and Baulcombe, D.C. (2007). The Plover virus silencing suppressor P0 targets ARGONAUTE proteins for degradation. *Curr. Biol.* 17, 1609–1614.
- 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. <https://doi.org/10.1038/35053110>
- Bingeli, O., Neyen, C., Poidevin, M., and Lemaitre, B. (2014). Prophenoloxidase activation is required for survival to microbial infections in *Drosophila*. *PLOS Pathog.* 10, e1004067. <https://doi.org/10.1371/journal.ppat.1004067>
- Bortolamiol, D., Pazhouhandeh, M., Marrocco, K., Genschik, P., and Ziegler-Graff, V. (2007). The Plover virus F box protein P0 targets ARGONAUTE1 to suppress RNA silencing. *Curr. Biol.* 17, 1615–1621.
- Boyer, L., Magoc, L., De Jardin, S., Cappillino, M., Paquette, N., Hinault, C., Charriere, G.M., Ip, W.K., Fracchia, S., Hennessy, E., *et al.* (2011). Pathogen-derived effectors trigger protective immunity via activation of the Rac2 enzyme and the IMD or Rip kinase signaling pathway. *Immunity* 35, 536–549. <https://doi.org/10.1016/j.immuni.2011.08.015>
- Bradshaw, C.J., Leroy, B., Bellard, C., Roiz, D., Albert, C., Fournier, A., Barbet-Massin, M., Salles, J.M., Simard, F., and Courchamp, F. (2016). Massive yet grossly underestimated global costs of invasive insects. *Nat. Commun.* 7, 12986. <https://doi.org/10.1038/ncomms12986>
- 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.
- Bronkhorst, A.W., and van Rij, R.P. (2014). The long and short of antiviral defense: small RNA-based immunity in insects. *Curr. Opin. Virol.* 7, 19–28. <https://doi.org/10.1016/j.coviro.2014.03.010>
- Bronkhorst, A.W., van Cleef, K.W., Vodovar, N., Ince, I.A., Blanc, H., Vlask, J.M., Saleh, M.C., and van Rij, R.P. (2012). The DNA virus Invertebrate iridescent virus 6 is a target of the *Drosophila* RNAi machinery. *Proc. Natl. Acad. Sci. U.S.A.* 109, E3604–13. <https://doi.org/10.1073/pnas.1207213109>
- Bronkhorst, A.W., Miesen, P., and van Rij, R.P. (2013). Small RNAs tackle large viruses: RNA interference-based antiviral defense against DNA viruses in insects. *Fly* 7, 216–223. <https://doi.org/10.4161/fly.25708>
- Bronkhorst, A.W., van Cleef, K.W., Venselaar, H., and van Rij, R.P. (2014). A dsRNA-binding protein of a complex invertebrate DNA virus suppresses the *Drosophila* RNAi response. *Nucleic Acids Res.* 42, 12237–12248. <https://doi.org/10.1093/nar/gku910>
- Bucheton, A. (1995). The relationship between the flamenco gene and gypsy in *Drosophila*: how to tame a retrovirus. *Trends Genet.* 11, 349–353.
- Bürkstümmer, T., Baumann, C., Blüml, S., Dixit, E., Dürnberger, G., Jahn, H., Panyavsky, M., Bilban, M., Colinge, J., Bennett, K.L., *et al.* (2009). An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome. *Nat. Immunol.* 10, 266–272. <https://doi.org/10.1038/ni.1702>
- Campbell, C.L., Keene, K.M., Brackney, D.E., Olson, K.E., Blair, C.D., Wilusz, J., and Foy, B.D. (2008). *Aedes aegypti* uses RNA interference in defense against Sindbis virus infection. *BMC Microbiol.* 8, 47. <https://doi.org/10.1186/1471-2180-8-47>
- Cao, C., Cogni, R., Barbier, V., and Jiggins, F.M. (2017). Complex Coding and Regulatory Polymorphisms in a Restriction Factor Determine the Susceptibility of *Drosophila* to Viral Infection. *Genetics* 206, 2159–2173. <https://doi.org/10.1534/genetics.117.201970>
- Carissimo, G., Pondeville, E., McFarlane, M., Dietrich, I., Mitri, C., Bischoff, E., Antoniewski, C., Bourgooin, C., Failloux, A.B., Kohl, A., *et al.* (2015). Antiviral immunity of *Anopheles gambiae* is highly compartmentalized, with distinct roles for RNA interference and gut microbiota. *Proc. Natl. Acad. Sci. U.S.A.* 112, E176–185. <https://doi.org/10.1073/pnas.1412984112>
- Carré-Mlouka, C., Gaumer, S., Gay, P., Petitjean, A.M., Couloude, C., Dru, P., Bras, F., Dezélee, S., and Contamine, D. (2007). Control of sigma virus multiplication by the ref(2)P gene of *Drosophila melanogaster*: an in vivo study of the PB1 domain of Ref(2)P. *Genetics* 176, 409–419.
- Carrillo-Tripp, J., Dolezal, A.G., Goblirsch, M.J., Miller, W.A., Toth, A.L., and Bonning, B.C. (2016). In vivo and in vitro infection dynamics of honey bee viruses. *Sci. Rep.* 6, 22265. <https://doi.org/10.1038/srep22265>
- Cenik, E.S., Fukunaga, R., Lu, G., Dutcher, R., Wang, Y., Tanaka Hall, T.M., and Zamore, P.D. (2011). Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. *Mol. Cell* 42, 172–184. <https://doi.org/10.1016/j.molcel.2011.03.002>
- Chao, J.A., Lee, J.H., Chapados, B.R., Debler, E.W., Schneemann, A., and Williamson, J.R. (2005). Dual modes of RNA-silencing suppression by Flock House virus protein B2. *Nat. Struct. Mol. Biol.* 12, 952–957.
- Chejanovsky, N., Ophir, R., Schwager, M.S., Slabezki, Y., Grossman, S., and Cox-Foster, D. (2014). Characterization of viral siRNA populations in honey bee colony collapse disorder. *Virology* 454–455, 176–183. <https://doi.org/10.1016/j.virol.2014.02.012>
- Choe, K.M., Werner, T., Stöven, S., Hultmark, D., and Anderson, K.V. (2002). Requirement for a peptidoglycan recognition protein (PGRP) in Relish activation and antibacterial immune responses in *Drosophila*. *Science* 296, 359–362. <https://doi.org/10.1126/science.1070216>
- Chung, W.J., Okamura, K., Martin, R., and Lai, E.C. (2008). Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr. Biol.* 18, 795–802. <https://doi.org/10.1016/j.cub.2008.05.006>
- Cirimotich, C.M., Scott, J.C., Phillips, A.T., Geiss, B.J., and Olson, K.E. (2009). Suppression of RNA interference increases alphavirus replication and virus-associated mortality in *Aedes aegypti* mosquitoes. *BMC Microbiol.* 9, 49. <https://doi.org/10.1186/1471-2180-9-49>
- 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. <https://doi.org/10.1261/rna.044701.114>

- Costa, A., Jan, E., Sarnow, P., and Schneider, D. (2009). The Imd pathway is involved in antiviral immune responses in *Drosophila*. *PLOS ONE* 4, e7436. <https://doi.org/10.1371/journal.pone.0007436>
- Coste, F., Kemp, C., Bobezeau, V., Hetru, C., Kellenberger, C., Imler, J.L., and Roussel, A. (2012). Crystal structure of Diedel, a marker of the immune response of *Drosophila melanogaster*. *PLOS ONE* 7, e33416. <https://doi.org/10.1371/journal.pone.0033416>
- Czech, B., and Hannon, G.J. (2016). One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. *Trends Biochem. Sci.* 41, 324–337.
- Deddouche, S., Matt, N., Budd, A., Mueller, S., Kemp, C., Galiana-Arnoux, D., Dostert, C., Antoniewski, C., Hoffmann, J.A., and Imler, J.L. (2008). The DExD/H-box helicase Dicer-2 mediates the induction of antiviral activity in *drosophila*. *Nat. Immunol.* 9, 1425–1432. <https://doi.org/10.1038/ni.1664>
- Deddouche, S., Goubau, D., Rehwinkel, J., Chakravarty, P., Begum, S., Maillard, P.V., Borg, A., Matthews, N., Feng, Q., van Kuppeveld, F.J., et al. (2014). Identification of an LGP2-associated MDAs agonist in picornavirus-infected cells. *Elife* 3, e01535. <https://doi.org/10.7554/eLife.01535>
- Ding, S.W. (2010). RNA-based antiviral immunity. *Nat. Rev. Immunol.* 10, 632–644. <https://doi.org/10.1038/nri2824>
- Dönertas, D., Sienski, G., and Brennecke, J. (2013). *Drosophila* Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. *Genes Dev.* 27, 1693–1705. <https://doi.org/10.1101/gad.221150.113>
- Dostert, C., Jouanguy, E., Irving, P., Troxler, L., Galiana-Arnoux, D., Hetru, C., Hoffmann, J.A., and Imler, J.-L. (2005). The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of *Drosophila*. *Nat. Immunol.* 6, 946–953. <https://doi.org/10.1038/ni1237>
- Dudas, G., and Obbard, D.J. (2015). Are arthropods at the heart of virus evolution? *ELife* 4 <https://doi.org/10.7554/eLife.06837>
- El Chamy, L., Leclerc, V., Caldelari, I., and Reichhart, J.-M. (2008). Sensing of 'danger signals' and pathogen-associated molecular patterns defines binary signaling pathways 'upstream' of Toll. *Nat. Immunol.* 9, 1165–1170 <https://doi.org/10.1038/ni.1643>
- Fareh, M., van Lopik, J., Katechis, I., Bronkhorst, A.W., Haagsma, A.C., van Rij, R.P., and Joo, C. (2018). Viral suppressors of RNAi employ a rapid screening mode to discriminate viral RNA from cellular small RNA. *Nucleic Acids Res.* 46, 3187–3197. <https://doi.org/10.1093/nar/gkx1316>
- Ferrandon, D., Imler, J.L., Hetru, C., and Hoffmann, J.A. (2007). The *Drosophila* systemic immune response: sensing and signalling during bacterial and fungal infections. *Nat. Rev. Immunol.* 7, 862–874.
- Ferreira, Á.G., Naylor, H., Esteves, S.S., Pais, I.S., Martins, N.E., and Teixeira, L. (2014). The Toll-dorsal pathway is required for resistance to viral oral infection in *Drosophila*. *PLOS Pathog.* 10, e1004507. <https://doi.org/10.1371/journal.ppat.1004507>
- Ferreira, F.V., Aguiar, E.R.G.R., Olmo, R.P., de Oliveira, K.P.V., Silva, E.G., Sant'Anna, M.R.V., Gontijo, N.F., Kroon, E.G., Imler, J.L., and Marques, J.T. (2018). The small non-coding RNA response to virus infection in the Leishmania vector *Lutzomyia longipalpis*. *PLOS Negl. Trop. Dis.* 12, e0006569. <https://doi.org/10.1371/journal.pntd.0006569>
- 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. <https://doi.org/10.1038/35888>
- Förstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D. (2005). Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLOS Biol.* 3, e236.
- Fragkoudis, R., Chi, Y., Siu, R.W., Barry, G., Attarzadeh-Yazdi, G., Merits, A., Nash, A.A., Fazakerley, J.K., and Kohl, A. (2008). Semliki Forest virus strongly reduces mosquito host defence signaling. *Insect Mol. Biol.* 17, 647–656. <https://doi.org/10.1111/j.1365-2583.2008.00834.x>
- Fukunaga, R., Colpan, C., Han, B.W., and Zamore, P.D. (2014). Inorganic phosphate blocks binding of pre-miRNA to Dicer-2 via its PAZ domain. *EMBO J.* 33, 371–384. <https://doi.org/10.1002/emboj.201387176>
- Galiana-Arnoux, D., Dostert, C., Schneemann, A., Hoffmann, J.A., and Imler, J.L. (2006). Essential function in vivo for Dicer-2 in host defense against RNA viruses in *drosophila*. *Nat. Immunol.* 7, 590–597.
- Gebhardt, A., Laudenbach, B.T., and Pichlmair, A. (2017). Discrimination of self and non-self ribonucleic acids. *J. Interferon Cytokine Res.* 37, 184–197. <https://doi.org/10.1089/jir.2016.0092>
- 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.
- Goic, B., Vodovar, N., Mondotte, J.A., Monot, C., Frangeul, L., Blanc, H., Gausson, V., Vera-Otarola, J., Cristofari, G., and Saleh, M.C. (2013). RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model *Drosophila*. *Nat. Immunol.* 14, 396–403. <https://doi.org/10.1038/ni.2542>
- Goic, B., Stapleford, K.A., Frangeul, L., Doucet, A.J., Gausson, V., Blanc, H., Schemmel-Jofre, N., Cristofari, G., Lambrechts, L., Vignuzzi, M., et al. (2016). Virus-derived DNA drives mosquito vector tolerance to arboviral infection. *Nat. Commun.* 7, 12410. <https://doi.org/10.1038/ncomms12410>
- Gokhale, N.S., and Horner, S.M. (2017). RNA modifications go viral. *PLOS Pathog.* 13, e1006188. <https://doi.org/10.1371/journal.ppat.1006188>
- Gokhale, N.S., McIntyre, A.B.R., McFadden, M.J., Roder, A.E., Kennedy, E.M., Gandara, J.A., Hopcraft, S.E., Quicke, K.M., Vazquez, C., Willer, J., et al. (2016). N6-Methyladenosine in Flaviviridae Viral RNA Genomes Regulates Infection. *Cell Host Microbe* 20, 654–665.
- Gold, K.S., and Brückner, K. (2015). Macrophages and cellular immunity in *Drosophila melanogaster*. *Semin. Immunol.* 27, 357–368. <https://doi.org/10.1016/j.smim.2016.03.010>
- Gonzales-var Horn, S.R., and Sarnow, P. (2017). Making the Mark: The Role of Adenosine Modifications in

- the Life Cycle of RNA Viruses. *Cell Host Microbe* 21, 661–669.
- Gordon, O., Henry, C.M., Srinivasan, N., Ahrens, S., Franz, A., Deddouche, S., Chakravarty, P., Phillips, D., George, R., Kjaer, S., et al. (2018). α -actinin accounts for the bioactivity of actin preparations in inducing STAT target genes in *Drosophila melanogaster*. *ELife* 7 <https://doi.org/10.7554/eLife.38636>.
- Goto, A., Okado, K., Martins, N., Cai, H., Barbier, V., Lamiable, O., Troxler, L., Santiago, E., Kuhn, L., Paik, D., et al. (2018). The Kinase IKK β Regulates a STING- and NF- κ B-Dependent Antiviral Response Pathway in *Drosophila*. *Immunity* 49, 225–234.e4.
- Gottar, M., Gobert, V., Michel, T., Belvin, M., Duyk, G., Hoffmann, J.A., Ferrandon, D., and Royet, J. (2002). The *Drosophila* immune response against Gram-negative bacteria is mediated by a peptidoglycan recognition protein. *Nature* 416, 640–644. <https://doi.org/10.1038/nature734>
- Gottar, M., Gobert, V., Matskevich, A.A., Reichhart, J.M., Wang, C., Butt, T.M., Belvin, M., Hoffmann, J.A., and Ferrandon, D. (2006). Dual detection of fungal infections in *Drosophila* via recognition of glucans and sensing of virulence factors. *Cell* 127, 1425–1437.
- Goubau, D., Deddouche, S., and Reis e Sousa, C. (2013). Cytosolic sensing of viruses. *Immunity* 38, 855–869. <https://doi.org/10.1016/j.immuni.2013.05.007>
- Goubau, D., Schlee, M., Deddouche, S., Pruijssers, A.J., Zillinger, T., Goldeck, M., Schubert, C., Van der Veen, A.G., Fujimura, T., Rehwinkel, J., et al. (2014). Antiviral immunity via RIG-I-mediated recognition of RNA bearing 5'-diphosphates. *Nature* 514, 372–375. <https://doi.org/10.1038/nature13590>
- Grivna, S.T., Pyhtila, B., and Lin, H. (2006). MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13415–13420.
- Gu, S., Jin, L., Huang, Y., Zhang, F., and Kay, M.A. (2012). Slicing-independent RISC activation requires the argonaute PAZ domain. *Curr. Biol.* 22, 1536–1542. <https://doi.org/10.1016/j.cub.2012.06.040>
- Gueguen, G., Kalamarz, M.E., Ramroop, J., Uribe, J., and Govind, S. (2013). Polydnviral ankyrin proteins aid parasitic wasp survival by coordinate and selective inhibition of hematopoietic and immune NF-kappa B signaling in insect hosts. *PLOS Pathog.* 9, e1003580. <https://doi.org/10.1371/journal.ppat.1003580>
- 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.
- Guo, X., Zhang, R., Wang, J., Ding, S.W., and Lu, R. (2013). Homologous RIG-I-like helicase proteins direct RNAi-mediated antiviral immunity in *C. elegans* by distinct mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16085–16090. <https://doi.org/10.1073/pnas.1307453110>
- Habjan, M., and Pichlmair, A. (2015). Cytoplasmic sensing of viral nucleic acids. *Curr. Opin. Virol.* 11, 31–37. <https://doi.org/10.1016/j.coviro.2015.01.012>
- Haley, B., and Zamore, P.D. (2004). Kinetic analysis of the RNAi enzyme complex. *Nat. Struct. Mol. Biol.* 11, 599–606. <https://doi.org/10.1038/nsmb780>
- Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286, 950–952.
- 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. <https://doi.org/10.1038/35005107>
- 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. <https://doi.org/10.1126/science.aaa1264>
- Han, Y.H., Luo, Y.J., Wu, Q., Jovel, J., Wang, X.H., Aliyari, R., Han, C., Li, W.X., and Ding, S.W. (2011). RNA-based immunity terminates viral infection in adult *Drosophila* in the absence of viral suppression of RNA interference: characterization of viral small interfering RNA populations in wild-type and mutant flies. *J. Virol.* 85, 13153–13163. <https://doi.org/10.1128/JVI.05518-11>
- Helm, M., and Motorin, Y. (2017). Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.* 18, 275–291. <https://doi.org/10.1038/nrg.2016.169>
- Hess, A.M., Prasad, A.N., Ptitsyn, A., Ebel, G.D., Olson, K.E., Barbacioru, C., Monighetti, C., and Campbell, C.L. (2011). Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense. *BMC Microbiol.* 11, 45. <https://doi.org/10.1186/1471-2180-11-45>
- Hoffmann, J.A. (2003). The immune response of *Drosophila*. *Nature* 426, 33–38. <https://doi.org/10.1038/nature02021>
- Hoffmann, J.A., Kafatos, F.C., Janeway, C.A., and Ezekowitz, R.A. (1999). Phylogenetic perspectives in innate immunity. *Science* 284, 1313–1318.
- Hornung, V., Hartmann, R., Ablasser, A., and Hopfner, K.P. (2014). OAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids. *Nat. Rev. Immunol.* 14, 521–528. <https://doi.org/10.1038/nri3719>
- Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* 17, 1265–1272.
- Hua, X., Li, B., Song, L., Hu, C., Li, X., Wang, D., Xiong, Y., Zhao, P., He, H., Xia, Q., et al. (2018). Stimulator of interferon genes (STING) provides insect antiviral immunity by promoting Dredd caspase-mediated NF- κ B activation. *J. Biol. Chem.* 293, 11878–11890. <https://doi.org/10.1074/jbc.RA117.000194>
- Huang, X., Fejes Tóth, K., and Aravin, A.A. (2017). piRNA Biogenesis in *Drosophila melanogaster*. *Trends Genet.* 33, 882–894 <https://doi.org/10.1016/j.tig.2017.09.002>
- Huang, Z., Kingsolver, M.B., Avadhanula, V., and Hardy, R.W. (2013). An antiviral role for antimicrobial peptides during the arthropod response to alphavirus replication. *J. Virol.* 87, 4272–4280. <https://doi.org/10.1128/JVI.03360-12>

- Hussain, M., and Asgari, S. (2014). MicroRNAs as mediators of insect host-pathogen interactions and immunity. *J. Insect Physiol.* 70, 151–158. <https://doi.org/10.1016/j.jinsphys.2014.08.003>
- Hutvagner, G., and Zamore, P.D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. *Science* 297, 2056–2060. <https://doi.org/10.1126/science.1073827>
- Ikushima, H., Negishi, H., and Taniguchi, T. (2013). The IRF family transcription factors at the interface of innate and adaptive immune responses. *Cold Spring Harb. Symp. Quant. Biol.* 78, 105–116. <https://doi.org/10.1101/sqb.2013.78.020321>
- Imler, J.L., and Bulet, P. (2005). Antimicrobial peptides in *Drosophila*: structures, activities and gene regulation. *Chem. Immunol. Allergy* 86, 1–21 <https://doi.org/10.1159/000086648>.
- Issa, N., Guillaumot, N., Lauret, E., Matt, N., Schaeffer-Reiss, C., Van Dorsselaer, A., Reichhart, J.M., and Veillard, F. (2018). The Circulating Protease Persephone Is an Immune Sensor for Microbial Proteolytic Activities Upstream of the *Drosophila* Toll Pathway. *Mol. Cell* 69, 539–550.e6.
- Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T., and Tomari, Y. (2010). Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol. Cell* 39, 292–299. <https://doi.org/10.1016/j.molcel.2010.05.015>
- Iwasaki, S., Sasaki, H.M., Sakaguchi, Y., Suzuki, T., Tadakuma, H., and Tomari, Y. (2015). Defining fundamental steps in the assembly of the *Drosophila* RNAi enzyme complex. *Nature* 521, 533–536. <https://doi.org/10.1038/nature14254>
- Jayachandran, B., Hussain, M., and Asgari, S. (2012). RNA interference as a cellular defense mechanism against the DNA virus baculovirus. *J. Virol.* 86, 13729–13734. <https://doi.org/10.1128/JVI.02041-12>
- Jiang, F., Ye, X., Liu, X., Fincher, L., McKearin, D., and Liu, Q. (2005). Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev.* 19, 1674–1679 <https://doi.org/10.1101/gad.1334005>.
- Jiang, H., Patel, P.H., Kohlmaier, A., Grenley, M.O., McEwen, D.G., and Edgar, B.A. (2009). Cytokine/Jak/Stat signaling mediates regeneration and homeostasis in the *Drosophila* midgut. *Cell* 137, 1343–1355. <https://doi.org/10.1016/j.cell.2009.05.014>
- Jupatanakul, N., Sim, S., Angleró-Rodríguez, Y.I., Souza-Neto, J., Das, S., Poti, K.E., Rossi, S.L., Bergren, N., Vasilakis, N., and Dimopoulos, G. (2017). Engineered *Aedes aegypti* JAK/STAT Pathway-Mediated Immunity to Dengue Virus. *PLOS Negl. Trop. Dis.* 11, e0005187. <https://doi.org/10.1371/journal.pntd.0005187>
- Karikó, K., Buckstein, M., Ni, H., and Weissman, D. (2005). Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* 23, 165–175 <https://doi.org/10.1016/j.immuni.2005.06.008>.
- Karlikow, M., Goic, B., Mongelli, V., Salles, A., Schmitt, C., Bonne, I., Zurzolo, C., and Saleh, M.C. (2016). *Drosophila* cells use nanotube-like structures to transfer dsRNA and RNAi machinery between cells. *Sci. Rep.* 6, 27085. <https://doi.org/10.1038/srep27085>
- Kato, H., Sato, S., Yoneyama, M., Yamamoto, M., Uematsu, S., Matsui, K., Tsujimura, T., Takeda, K., Fujita, T., Takeuchi, O., *et al.* (2005). Cell type-specific involvement of RIG-I in antiviral response. *Immunity* 23, 19–28 <https://doi.org/10.1016/j.immuni.2005.04.010>.
- Kato, H., Takeuchi, O., Sato, S., Yoneyama, M., Yamamoto, M., Matsui, K., Uematsu, S., Jung, A., Kawai, T., Ishii, K.J., *et al.* (2006). Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* 441, 101–105.
- Kawai, T., and Akira, S. (2011). Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* 34, 637–650. <https://doi.org/10.1016/j.immuni.2011.05.006>
- Kemp, C., Mueller, S., Goto, A., Barbier, V., Paro, S., Bonnay, F., Dostert, C., Troxler, L., Hetru, C., Meignin, C., *et al.* (2013). Broad RNA interference-mediated antiviral immunity and virus-specific inducible responses in *Drosophila*. *J. Immunol.* 190, 650–658. <https://doi.org/10.4049/jimmunol.1102486>
- Khong, A., Kerr, C.H., Yeung, C.H., Keatings, K., Nayak, A., Allan, D.W., and Jan, E. (2017). Disruption of stress granule formation by the multifunctional cricket paralysis virus 1A protein. *J. Virol.* 91, e01779-16. <https://doi.org/10.1128/JVI.01779-16>.
- Khoo, C.C., Piper, J., Sanchez-Vargas, I., Olson, K.E., and Franz, A.W. (2010). The RNA interference pathway affects midgut infection- and escape barriers for Sindbis virus in *Aedes aegypti*. *BMC Microbiol.* 10, 130. <https://doi.org/10.1186/1471-2180-10-130>
- Khoo, C.C., Doty, J.B., Heersink, M.S., Olson, K.E., and Franz, A.W. (2013). Transgene-mediated suppression of the RNA interference pathway in *Aedes aegypti* interferes with gene silencing and enhances Sindbis virus and dengue virus type 2 replication. *Insect Mol. Biol.* 22, 104–114. <https://doi.org/10.1111/imb.12008>
- Kim, K., Lee, Y.S., and Carthew, R.W. (2007). Conversion of pre-RISC to holo-RISC by Ago2 during assembly of RNAi complexes. *RNA* 13, 22–29. <https://doi.org/10.1261/rna.283207>
- Kowalinski, E., Lunardi, T., McCarthy, A.A., Loubser, J., Brunel, J., Grigorov, B., Gerlier, D., and Cusack, S. (2011). Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* 147, 423–435. <https://doi.org/10.1016/j.cell.2011.09.039>
- Kranzusch, P.J., Wilson, S.C., Lee, A.S., Berger, J.M., Doudna, J.A., and Vance, R.E. (2015). Ancient Origin of cGAS-STING Reveals Mechanism of Universal 2',3' cGAMP Signaling. *Mol. Cell* 59, 891–903. <https://doi.org/10.1016/j.molcel.2015.07.022>
- Kreuze, J.F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., and Simon, R. (2009). Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388, 1–7. <https://doi.org/10.1016/j.virol.2009.03.024>
- Kroemer, J.A., and Webb, B.A. (2005). Ikappabeta-related vankyrin genes in the *Campoletis sonorensis* ichnovirus: temporal and tissue-specific patterns of expression in parasitized *Heliothis virescens* lepidopteran hosts. *J. Virol.* 79, 7617–7628.
- Kroemer, J.A., and Webb, B.A. (2006). Divergences in protein activity and cellular localization within the *Campoletis sonorensis* Ichnovirus Vankyrin family. *J.*

- Viol. 80, 12219–12228. <https://doi.org/10.1128/JVI.01187-06>
- Ktistakis, N.T., and Tooze, S.A. (2016). Digesting the expanding mechanisms of autophagy. *Trends Cell Biol.* 26, 624–635.
- Kwak, P.B., and Tomari, Y. (2012). The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nat. Struct. Mol. Biol.* 19, 145–151. <https://doi.org/10.1038/nsmb.2232>
- Lambrechts, L., Quillery, E., Noël, V., Richardson, J.H., Jarman, R.G., Scott, T.W., and Chevillon, C. (2013). Specificity of resistance to dengue virus isolates is associated with genotypes of the mosquito antiviral gene Dicer-2. *Proc. Biol. Sci.* 280, 20122437. <https://doi.org/10.1098/rspb.2012.2437>
- Lamiable, O., Kellenberger, C., Kemp, C., Troxler, L., Pelte, N., Boutros, M., Marques, J.T., Daefler, L., Hoffmann, J.A., Roussel, A., *et al.* (2016a). Cytokine Dieldel and a viral homologue suppress the IMD pathway in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 113, 698–703. <https://doi.org/10.1073/pnas.1516122113>
- Lamiable, O., Arnold, J., de Faria, I.J.D.S., Olmo, R.P., Bergami, F., Meignin, C., Hoffmann, J.A., Marques, J.T., and Imler, J.L. (2016b). Analysis of the Contribution of Hemocytes and Autophagy to *Drosophila* Antiviral Immunity. *J. Virol.* 90, 5415–5426. <https://doi.org/10.1128/JVI.00238-16>
- Lan, H., Chen, H., Liu, Y., Jiang, C., Mao, Q., Jia, D., Chen, Q., and Wei, T. (2016). Small interfering RNA pathway modulates initial viral infection in midgut epithelium of insect after ingestion of virus. *J. Virol.* 90, 917–929. <https://doi.org/10.1128/JVI.01835-15>
- Le Thomas, A., Rogers, A.K., Webster, A., Marinov, G.K., Liao, S.E., Perkins, E.M., Hur, J.K., Aravin, A.A., and Tóth, K.F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* 27, 390–399. <https://doi.org/10.1101/gad.209841.112>
- Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* 117, 69–81.
- Léger, P., Lara, E., Jagla, B., Sismeiro, O., Mansuroglu, Z., Coppée, J.Y., Bonnefoy, E., and Bouloy, M. (2013). Dicer-2- and Piwi-mediated RNA interference in Rift Valley fever virus-infected mosquito cells. *J. Virol.* 87, 1631–1648. <https://doi.org/10.1128/JVI.02795-12>
- Lemaitre, B., Meister, M., Govind, S., Georgel, P., Steward, R., Reichhart, J.M., and Hoffmann, J.A. (1995). Functional analysis and regulation of nuclear import of dorsal during the immune response in *Drosophila*. *EMBO J.* 14, 536–545.
- Letourneau, M., Lapraz, F., Sharma, A., Vanzo, N., Waltzer, L., and Crozatier, M. (2016). *Drosophila* hematopoiesis under normal conditions and in response to immune stress. *FEBS Lett.* 590, 4034–4051. <https://doi.org/10.1002/1873-3468.12327>
- Leulier, F., Rodriguez, A., Khush, R.S., Abrams, J.M., and Lemaitre, B. (2000). The *Drosophila* caspase Dredd is required to resist gram-negative bacterial infection. *EMBO Rep.* 1, 353–358. <https://doi.org/10.1093/embo-reports/kvd073>
- Lewis, S.H., Quarles, K.A., Yang, Y., Tanguy, M., Frézal, L., Smith, S.A., Sharma, P.P., Cordaux, R., Gilbert, C., Giraud, I., *et al.* (2018). Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nat. Ecol. Evol.* 2, 174–181. <https://doi.org/10.1038/s41559-017-0403-4>
- Li, C.X., Shi, M., Tian, J.H., Lin, X.D., Kang, Y.J., Chen, L.J., Qin, X.C., Xu, J., Holmes, E.C., and Zhang, Y.Z. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife* 4. <https://doi.org/10.7554/eLife.05378>
- Li, H., Li, W.X., and Ding, S.W. (2002). Induction and suppression of RNA silencing by an animal virus. *Science* 296, 1319–1321. <https://doi.org/10.1126/science.1070948>
- Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr. Biol.* 15, 1501–1507.
- Liang, C., Wang, Y., Murota, Y., Liu, X., Smith, D., Siomi, M.C., and Liu, Q. (2015). TAF11 Assembles the RISC Loading Complex to Enhance RNAi Efficiency. *Mol. Cell* 59, 807–818. <https://doi.org/10.1016/j.molcel.2015.07.006>
- Liddicoat, B.J., Piskol, R., Chalk, A.M., Ramaswami, G., Higuchi, M., Hartner, J.C., Li, J.B., Seeburg, P.H., and Walkley, C.R. (2015). RNA editing by ADAR1 prevents MDAS sensing of endogenous dsRNA as nonself. *Science* 349, 1115–1120. <https://doi.org/10.1126/science.aac7049>
- 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. U.S.A.* 104, 6714–6719.
- 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.
- 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. <https://doi.org/10.1126/science.1088710>
- Liu, X., Sano, T., Guan, Y., Nagata, S., Hoffmann, J.A., and Fukuyama, H. (2012). *Drosophila* EYA regulates the immune response against DNA through an evolutionarily conserved threonine phosphatase motif. *PLOS ONE* 7, e42725. <https://doi.org/10.1371/journal.pone.0042725>
- Liu, Y., Ye, X., Jiang, F., Liang, C., Chen, D., Peng, J., Kinch, L.N., Grishin, N.V., and Liu, Q. (2009). C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. *Science* 325, 750–753. <https://doi.org/10.1126/science.1176325>
- Liu, Y., Gordesky-Gold, B., Leney-Greene, M., Weinbren, N.L., Tudor, M., and Cherry, S. (2018). Inflammation-induced, STING-dependent autophagy restricts Zika virus infection in the *Drosophila* brain. *Cell Host Microbe* 24, 57–68.e3 <https://doi.org/10.1016/j.chom.2018.05.022>
- Longdon, B., Cao, C., Martinez, J., and Jiggins, F.M. (2013). Previous exposure to an RNA virus does not protect against subsequent infection in *Drosophila melanogaster*.

- PLoS One 8, e73833 <https://doi.org/10.1371/journal.pone.0073833>.
- Lu, Y., Wu, L.P., and Anderson, K.V. (2001). The antibacterial arm of the *Drosophila* innate immune response requires an IκappaB kinase. *Genes Dev.* 15, 104–110.
- Ma, J.B., Ye, K., and Patel, D.J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* 429, 318–322. <https://doi.org/10.1038/nature02519>
- Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T., and Patel, D.J. (2005). Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* 434, 666–670.
- Maeda, K., and Akira, S. (2016). TLR7 Structure: Cut in Z-Loop. *Immunity* 45, 705–707 <https://doi.org/10.1016/j.immuni.2016.10.003>
- Magwire, M.M., Bayer, F., Webster, C.L., Cao, C., and Jiggins, F.M. (2011). Successive increases in the resistance of *Drosophila* to viral infection through a transposon insertion followed by a Duplication. *PLOS Genet.* 7, e1002337. <https://doi.org/10.1371/journal.pgen.1002337>
- Magwire, M.M., Fabian, D.K., Schweyen, H., Cao, C., Longdon, B., Bayer, F., and Jiggins, F.M. (2012). Genome-wide association studies reveal a simple genetic basis of resistance to naturally coevolving viruses in *Drosophila melanogaster*. *PLOS Genet.* 8, e1003057. <https://doi.org/10.1371/journal.pgen.1003057>
- Margolis, S.R., Wilson, S.C., and Vance, R.E. (2017). Evolutionary Origins of cGAS-STING Signaling. *Trends Immunol.* 38, 733–743 <https://doi.org/10.1016/j.it.2017.03.004>
- Marques, J.T., and Imler, J.-L. (2016). The diversity of insect antiviral immunity: insights from viruses. *Curr. Opin. Microbiol.* 32, 71–76 <https://doi.org/10.1016/j.mib.2016.05.002>
- Marques, J.T., Kim, K., Wu, P.H., Alleyne, T.M., Jafari, N., and Carthew, R.W. (2010). Loqs and R2D2 act sequentially in the siRNA pathway in *Drosophila*. *Nat. Struct. Mol. Biol.* 17, 24–30. <https://doi.org/10.1038/nsmb.1735>
- Marques, J.T., Wang, J.P., Wang, X., de Oliveira, K.P., Gao, C., Aguiar, E.R., Jafari, N., and Carthew, R.W. (2013). Functional specialization of the small interfering RNA pathway in response to virus infection. *PLOS Pathog.* 9, e1003579. <https://doi.org/10.1371/journal.ppat.1003579>
- Martin, M., Hiroyasu, A., Guzman, R.M., Roberts, S.A., and Goodman, A.G. (2018). Analysis of *Drosophila* STING Reveals an Evolutionarily Conserved Antimicrobial Function. *Cell Rep.* 23, 3537–3550. <https://doi.org/10.1016/j.celrep.2018.05.029>
- Martins, N., Imler, J.L., and Meignin, C. (2016). Discovery of novel targets for antivirals: learning from flies. *Curr. Opin. Virol.* 20, 64–70 <https://doi.org/10.1016/j.coviro.2016.09.005>
- Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* 123, 607–620 <https://doi.org/10.1016/j.cell.2005.08.044>
- Matsumoto, N., Nishimasu, H., Sakakibara, K., Nishida, K.M., Hirano, T., Ishitani, R., Siomi, H., Siomi, M.C., and Nureki, O. (2016). Crystal Structure of Silkworm PIWI-Clade Argonaute Siwi Bound to piRNA. *Cell* 167, 484–497. <https://doi.org/10.1016/j.cell.2016.09.002>
- McFarlane, M., Arias-Goeta, C., Martin, E., O'Hara, Z., Lulla, A., Mousson, L., Rainey, S.M., Misbah, S., Schnetzler, E., Donald, C.L., et al. (2014). Characterization of *Aedes aegypti* innate-immune pathways that limit Chikungunya virus replication. *PLOS Negl. Trop. Dis.* 8, e2994. <https://doi.org/10.1371/journal.pntd.0002994>
- Merkling, S.H., and van Rij, R.P. (2015). Analysis of resistance and tolerance to virus infection in *Drosophila*. *Nat. Protoc.* 10, 1084–1097. <https://doi.org/10.1038/nprot.2015.071>
- Merkling, S.H., Bronkhorst, A.W., Kramer, J.M., Overheul, G.J., Schenck, A., and Van Rij, R.P. (2015a). The epigenetic regulator G9a mediates tolerance to RNA virus infection in *Drosophila*. *PLOS Pathog.* 11, e1004692. <https://doi.org/10.1371/journal.ppat.1004692>
- Merkling, S.H., Overheul, G.J., van Mierlo, J.T., Arends, D., Gilissen, C., and van Rij, R.P. (2015b). The heat shock response restricts virus infection in *Drosophila*. *Sci. Rep.* 5, 12758. <https://doi.org/10.1038/srep12758>
- Miesen, P., Girardi, E., and van Rij, R.P. (2015). Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in *Aedes aegypti* mosquito cells. *Nucleic Acids Res.* 43, 6545–6556. <https://doi.org/10.1093/nar/gkv590>
- Miesen, P., Joosten, J., and van Rij, R.P. (2016a). PIWIs Go Viral: Arbovirus-Derived piRNAs in Vector Mosquitoes. *PLOS Pathog.* 12, e1006017. <https://doi.org/10.1371/journal.ppat.1006017>
- Miesen, P., Ivens, A., Buck, A.H., and van Rij, R.P. (2016b). Small RNA Profiling in Dengue Virus 2-Infected *Aedes* Mosquito Cells Reveals Viral piRNAs and Novel Host miRNAs. *PLOS Negl. Trop. Dis.* 10, e0004452. <https://doi.org/10.1371/journal.pntd.0004452>
- Miller, J.R., Koren, S., Dille, K.A., Puri, V., Brown, D.M., Harkins, D.M., Thibaud-Nissen, F., Rosen, B., Chen, X.G., Tu, Z., et al. (2018). Analysis of the *Aedes albopictus* C6/36 genome provides insight into cell line utility for viral propagation. *Gigascience* 7, 1–13. <https://doi.org/10.1093/gigascience/gix135>
- Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., et al. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346, 763–767. <https://doi.org/10.1126/science.1257570>
- Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H., and Siomi, M.C. (2005). Slicer function of *Drosophila* Argonautes and its involvement in RISC formation. *Genes Dev.* 19, 2837–2848. <https://doi.org/10.1101/gad.1370605>
- Miyoshi, K., Miyoshi, T., Hartig, J.V., Siomi, H., and Siomi, M.C. (2010a). Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in *Drosophila*. *RNA* 16, 506–515. <https://doi.org/10.1261/rna.1952110>
- Miyoshi, T., Takeuchi, A., Siomi, H., and Siomi, M.C. (2010b). A direct role for Hsp90 in pre-RISC formation in *Drosophila*. *Nat. Struct. Mol. Biol.* 17, 1024–1026. <https://doi.org/10.1038/nsmb.1875>

- Mlih, M., Khericha, M., Birdwell, C., West, A.P., and Karpac, J. (2018). A virus-acquired host cytokine controls systemic aging by antagonizing apoptosis. *PLoS Biol.* 16, e2005796. <https://doi.org/10.1371/journal.pbio.2005796>
- Mo, X., Yang, X., and Yuan, Y.A. (2018). Structural insights into *Drosophila*-C3PO complex assembly and 'Dynamic Side Port' model in substrate entry and release. *Nucleic Acids Res.* 46, 8590–8604. <https://doi.org/10.1093/nar/gky465>
- 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. <https://doi.org/10.1126/science.aal1039>
- Molina-Cruz, A., Zilversmit, M.M., Neafsey, D.E., Hartl, D.L., and Barillas-Mury, C. (2016). Mosquito Vectors and the Globalization of *Plasmodium falciparum* Malaria. *Annu. Rev. Genet.* 50, 447–465. <https://doi.org/10.1146/annurev-genet-120215-035211>
- Mondotte, J.A., Gausson, V., Frangeul, L., Blanc, H., Lambrechts, L., and Saleh, M.C. (2018). Immune priming and clearance of orally acquired RNA viruses in *Drosophila*. *Nat. Microbiol.* 3, 1394–1403. <https://doi.org/10.1038/s41564-018-0265-9>
- Monsion, B., Incarbone, M., Hleibieh, K., Poignavent, V., Ghannam, A., Dunoyer, P., Daeffler, L., Tilsner, J., and Ritzenthaler, C. (2018). Efficient Detection of Long dsRNA *In Vitro* and *In Vivo* Using the dsRNA Binding Domain from FHV B2 Protein. *Front. Plant Sci.* 9, 70. <https://doi.org/10.3389/fpls.2018.00070>
- Morazzani, E.M., Wiley, M.R., Murreddu, M.G., Adelman, Z.N., and Myles, K.M. (2012). Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. *PLoS Pathog.* 8, e1002470. <https://doi.org/10.1371/journal.ppat.1002470>
- Mueller, S., Gausson, V., Vodovar, N., Deddouche, S., Troxler, L., Perot, J., Pfeffer, S., Hoffmann, J.A., Saleh, M.C., and Imler, J.L. (2010). RNAi-mediated immunity provides strong protection against the negative-strand RNA vesicular stomatitis virus in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19390–19395. <https://doi.org/10.1073/pnas.1014378107>
- Mukae, N., Yokoyama, H., Yokokura, T., Sakoyama, Y., and Nagata, S. (2002). Activation of the innate immunity in *Drosophila* by endogenous chromosomal DNA that escaped apoptotic degradation. *Genes Dev.* 16, 2662–2671. <https://doi.org/10.1101/gad.1022802>
- Müller, S., and Imler, J.L. (2007). Dicing with viruses: microRNAs as antiviral factors. *Immunity* 27, 1–3.
- Mussabekova, A., Daeffler, L., and Imler, J.L. (2017). Innate and intrinsic antiviral immunity in *Drosophila*. *Cell. Mol. Life Sci.* 74, 2039–2054. <https://doi.org/10.1007/s00018-017-2453-9>
- Myles, K.M., Wiley, M.R., Morazzani, E.M., and Adelman, Z.N. (2008). Alphavirus-derived small RNAs modulate pathogenesis in disease vector mosquitoes. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19938–19943. <https://doi.org/10.1073/pnas.0803408105>
- Nainu, F., Tanaka, Y., Shiratsuchi, A., and Nakanishi, Y. (2015). Protection of insects against viral infection by apoptosis-dependent phagocytosis. *J. Immunol.* 195, 5696–5706. <https://doi.org/10.4049/jimmunol.1500613>
- Nakanishi, K. (2016). Anatomy of RISC: how do small RNAs and chaperones activate Argonaute proteins? *Wiley Interdiscip. Rev. RNA* 7, 637–660 <https://doi.org/10.1002/wrna.1356>.
- Nayak, A., Berry, B., Tassetto, M., Kunitomi, M., Acevedo, A., Deng, C., Krutchinsky, A., Gross, J., Antoniewski, C., and Andino, R. (2010). Cricket paralysis virus antagonizes Argonaute 2 to modulate antiviral defense in *Drosophila*. *Nat. Struct. Mol. Biol.* 17, 547–554. <https://doi.org/10.1038/nsmb.1810>
- Nayak, A., Kim, D.Y., Trnka, M.J., Kerr, C.H., Lidsky, P.V., Stanley, D.J., Rivera, B.M., Li, K.H., Burlingame, A.L., Jan, E., et al. (2018). A Viral Protein Restricts *Drosophila* RNAi Immunity by Regulating Argonaute Activity and Stability. *Cell Host Microbe* 24, 542–557.e9 <https://doi.org/10.1016/j.chom.2018.09.006>
- Nishida, K.M., Miyoshi, K., Ogino, A., Miyoshi, T., Siomi, H., and Siomi, M.C. (2013). Roles of R2D2, a cytoplasmic D2 body component, in the endogenous siRNA pathway in *Drosophila*. *Mol. Cell* 49, 680–691. <https://doi.org/10.1016/j.molcel.2012.12.024>
- Obbard, D.J., Jiggins, F.M., Halligan, D.L., and Little, T.J. (2006). Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr. Biol.* 16, 580–585. <https://doi.org/10.1016/j.cub.2006.01.065>
- Ohtani, H., Iwasaki, Y.W., Shibuya, A., Siomi, H., Siomi, M.C., and Saito, K. (2013). DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the *Drosophila* ovary. *Genes Dev.* 27, 1656–1661. <https://doi.org/10.1101/gad.221515.113>
- Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M.C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* 18, 1655–1666. <https://doi.org/10.1101/gad.1210204>
- Olmo, R.P., Ferreira, A.G.A., Izidoro-Toledo, T.C., Aguiar, E.R.G.R., de Faria, I.J.S., de Souza, K.P.R., Osório, K.P., Kuhn, L., Hammann, P., de Andrade, E.G., et al. (2018). Control of dengue virus in the midgut of *Aedes aegypti* by ectopic expression of the dsRNA-binding protein Loqs2. *Nat. Microbiol.* 3, 1385–1393. <https://doi.org/10.1038/s41564-018-0268-6>
- Palmer, W.H., Hadfield, J.D., and Obbard, D.J. (2018). RNA-Interference Pathways Display High Rates of Adaptive Protein Evolution in Multiple Invertebrates. *Genetics* 208, 1585–1599. <https://doi.org/10.1534/genetics.117.300567>
- 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. <https://doi.org/10.1016/j.devcel.2007.03.022>
- Paradkar, P.N., Trinidad, L., Voysey, R., Duchemin, J.B., and Walker, P.J. (2012). Secreted Vago restricts West Nile virus infection in *Culex* mosquito cells by activating the Jak-STAT pathway. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18915–18920. <https://doi.org/10.1073/pnas.1205231109>
- Paradkar, P.N., Duchemin, J.B., Voysey, R., and Walker, P.J. (2014). Dicer-2-dependent activation of *Culex* Vago occurs via the TRAF-Rel2 signaling pathway. *PLoS*

- Negl. Trop. Dis. 8, e2823. <https://doi.org/10.1371/journal.pntd.0002823>
- Park, J.H., and Shin, C. (2015). Slicer-independent mechanism drives small-RNA strand separation during human RISC assembly. *Nucleic Acids Res.* 43, 9418–9433. <https://doi.org/10.1093/nar/gkv937>
- Paro, S., Imler, J.L., and Meignin, C. (2015). Sensing viral RNAs by Dicer/RIG-I like ATPases across species. *Curr. Opin. Immunol.* 32, 106–113. <https://doi.org/10.1016/j.coi.2015.01.009>
- Péligon, A., Sarot, E., Payen-Groschène, G., and Bucheton, A. (2007). A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. *J. Virol.* 81, 1951–1960. <https://doi.org/10.1128/JVI.01980-06>
- Petit, M., Mongelli, V., Frangeul, L., Blanc, H., Jiggins, F., and Saleh, M.-C. (2016). piRNA pathway is not required for antiviral defense in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 113, E4218–E4227. <https://doi.org/10.1073/pnas.1607952113>
- Petrillo, J.E., Venter, P.A., Short, J.R., Gopal, R., Deddouche, S., Lamielle, O., Imler, J.L., and Schneemann, A. (2013). Cytoplasmic granule formation and translational inhibition of nodaviral RNAs in the absence of the double-stranded RNA binding protein B2. *J. Virol.* 87, 13409–13421. <https://doi.org/10.1128/JVI.02362-13>
- Pichlmair, A., Lassnig, C., Eberle, C.A., Górna, M.W., Baumann, C.L., Burkard, T.R., Bürckstümmer, T., Stefanovic, A., Krieger, S., Bennett, K.L., et al. (2011). IFIT1 is an antiviral protein that recognizes 5'-triphosphate RNA. *Nat. Immunol.* 12, 624–630. <https://doi.org/10.1038/ni.2048>
- Poirier, E.Z., Goic, B., Tomé-Poderti, L., Frangeul, L., Boussier, J., Gausson, V., Blanc, H., Vallet, T., Loyd, H., Levi, L.I., et al. (2018). Dicer-2-Dependent Generation of Viral DNA from Defective Genomes of RNA Viruses Modulates Antiviral Immunity in Insects. *Cell Host Microbe* 23, 353–365.e8. <https://doi.org/10.1016/j.chom.2018.02.001>
- Popham, H.J., Ellersieck, M.R., Li, H., and Bonning, B.C. (2016). Evaluation of the Insecticidal Efficacy of Wild Type and Recombinant Baculoviruses. *Methods Mol. Biol.* 1350, 407–444. https://doi.org/10.1007/978-1-4939-3043-2_21
- Posadas, D.M., and Carthew, R.W. (2014). MicroRNAs and their roles in developmental canalization. *Curr. Opin. Genet. Dev.* 27, 1–6. <https://doi.org/10.1016/j.gde.2014.03.005>
- Powers, A.M., and Waterman, S.H. (2017). A decade of arboviral activity—Lessons learned from the trenches. *PLOS Negl. Trop. Dis.* 11, e0005421. <https://doi.org/10.1371/journal.pntd.0005421>
- Rand, T.A., Petersen, S., Du, F., and Wang, X. (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* 123, 621–629. <https://doi.org/10.1016/j.cell.2005.10.020>
- Rao, X.J., Zhan, M.Y., Pan, Y.M., Liu, S., Yang, P.J., Yang, L.L., and Yu, X.Q. (2018). Immune functions of insect β GRPs and their potential application. *Dev. Comp. Immunol.* 83, 80–88. <https://doi.org/10.1016/j.dci.2017.12.007>
- Reich, D.P., Tyc, K.M., and Bass, B.L. (2018). C. elegans ADARs antagonize silencing of cellular dsRNAs by the antiviral RNAi pathway. *Genes Dev.* 32, 271–282. <https://doi.org/10.1101/gad.310672.117>
- Roers, A., Hiller, B., and Hornung, V. (2016). Recognition of endogenous nucleic acids by the innate immune System. *Immunity* 44, 739–754. <https://doi.org/10.1016/j.immuni.2016.04.002>
- Roignant, J.Y., Carré, C., Mugat, B., Szymczak, D., Lepesant, J.A., and Antoniewski, C. (2003). Absence of transitive and systemic pathways allows cell-specific and isoform-specific RNAi in *Drosophila*. *RNA* 9, 299–308.
- Royet, J., Gupta, D., and Dziarski, R. (2011). Peptidoglycan recognition proteins: modulators of the microbiome and inflammation. *Nat. Rev. Immunol.* 11, 837–851. <https://doi.org/10.1038/nri3089>
- Rutschmann, S., Jung, A.C., Hetru, C., Reichhart, J.M., Hoffmann, J.A., and Ferrandon, D. (2000a). The Rel protein DIF mediates the antifungal but not the antibacterial host defense in *Drosophila*. *Immunity* 12, 569–580.
- Rutschmann, S., Jung, A.C., Zhou, R., Silverman, N., Hoffmann, J.A., and Ferrandon, D. (2000b). Role of *Drosophila* IKK gamma in a toll-independent antibacterial immune response. *Nat. Immunol.* 1, 342–347. <https://doi.org/10.1038/79801>
- Sabin, L.R., Zheng, Q., Thekkat, P., Yang, J., Hannon, G.J., Gregory, B.D., Tudor, M., and Cherry, S. (2013). Dicer-2 processes diverse viral RNA species. *PLOS ONE* 8, e55458. <https://doi.org/10.1371/journal.pone.0055458>
- Saito, K., Ishizuka, A., Siomi, H., and Siomi, M.C. (2005). Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells. *PLOS Biol.* 3, e235. <https://doi.org/10.1371/journal.pbio.0030235>
- 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 Dev.* 20, 2214–2222. <https://doi.org/10.1101/gad.1454806>
- Saleh, M.C., van Rij, R.P., Hekele, A., Gillis, A., Foley, E., O'Farrell, P.H., and Andino, R. (2006). The endocytic pathway mediates cell entry of dsRNA to induce RNAi silencing. *Nat. Cell Biol.* 8, 793–802. <https://doi.org/10.1038/ncb1439>
- Saleh, M.C., Tassetto, M., van Rij, R.P., Goic, B., Gausson, V., Berry, B., Jacquier, C., Antoniewski, C., and Andino, R. (2009). Antiviral immunity in *Drosophila* requires systemic RNA interference spread. *Nature* 458, 346–350. <https://doi.org/10.1038/nature07712>
- Samuel, G.H., Wiley, M.R., Badawi, A., Adelman, Z.N., and Myles, K.M. (2016). Yellow fever virus capsid protein is a potent suppressor of RNA silencing that binds double-stranded RNA. *Proc. Natl. Acad. Sci. U.S.A.* 113, 13863–13868. <https://doi.org/10.1073/pnas.1600544113>
- Sánchez-Vargas, I., Scott, J.C., Poole-Smith, B.K., Franz, A.W., Barbosa-Solomieu, V., Wilusz, J., Olson, K.E., and Blair, C.D. (2009). Dengue virus type 2 infections of *Aedes aegypti* are modulated by the mosquito's RNA interference pathway. *PLOS Pathog.* 5, e1000299. <https://doi.org/10.1371/journal.ppat.1000299>

- Schirle, N.T., and MacRae, I.J. (2012). The crystal structure of human Argonaute2. *Science* 336, 1037–1040. <https://doi.org/10.1126/science.1221551>
- Schirle, N.T., Sheu-Gruttadauria, J., and MacRae, I.J. (2014). Structural basis for microRNA targeting. *Science* 346, 608–613. <https://doi.org/10.1126/science.1258040>
- Schwarz, D.S., Tomari, Y., and Zamore, P.D. (2004). The RNA-induced silencing complex is a Mg²⁺-dependent endonuclease. *Curr. Biol.* 14, 787–791. <https://doi.org/10.1016/j.cub.2004.03.008>
- Shelly, S., Lukinova, N., Bambina, S., Berman, A., and Cherry, S. (2009). Autophagy is an essential component of *Drosophila* immunity against vesicular stomatitis virus. *Immunity* 30, 588–598. <https://doi.org/10.1016/j.immuni.2009.02.009>
- Shi, M., Lin, X.D., Tian, J.H., Chen, L.J., Chen, X., Li, C.X., Qin, X.C., Li, J., Cao, J.P., Eden, J.S., et al. (2016). Redefining the invertebrate RNA virosphere. *Nature* [Epub ahead of print]. <https://doi.org/10.1038/nature20167>
- Sienski, G., Dönertas, D., and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* 151, 964–980. <https://doi.org/10.1016/j.cell.2012.10.040>
- Sienski, G., Batki, J., Senti, K.A., Dönertas, D., Tirian, L., Meixner, K., and Brennecke, J. (2015). Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. *Genes Dev.* 29, 2258–2271. <https://doi.org/10.1101/gad.271908.115>
- Singh, G., Popli, S., Hari, Y., Malhotra, P., Mukherjee, S., and Bhatnagar, R.K. (2009). Suppression of RNA silencing by Flock house virus B2 protein is mediated through its interaction with the PAZ domain of Dicer. *FASEB J.* 23, 1845–1857. <https://doi.org/10.1096/fj.08-125120>
- Sinha, N.K., Trettin, K.D., Aruscavage, P.J., and Bass, B.L. (2015). *Drosophila* dicer-2 cleavage is mediated by helicase- and dsRNA termini-dependent states that are modulated by Loquacious-PD. *Mol. Cell* 58, 406–417. <https://doi.org/10.1016/j.molcel.2015.03.012>
- Sinha, N.K., Iwasa, J., Shen, P.S., and Bass, B.L. (2018). Dicer uses distinct modules for recognizing dsRNA termini. *Science* 359, 329–334. <https://doi.org/10.1126/science.aag0921>
- 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. <https://doi.org/10.1126/science.1102514>
- Souza-Neto, J.A., Sim, S., and Dimopoulos, G. (2009). An evolutionary conserved function of the JAK-STAT pathway in anti-dengue defense. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17841–17846. <https://doi.org/10.1073/pnas.0905006106>
- Stein, D., Goltz, J.S., Jurcsak, J., and Stevens, L. (1998). The Dorsal-related immunity factor (Dif) can define the dorsal-ventral axis of polarity in the *Drosophila* embryo. *Development* 125, 2159–2169.
- Steiner, H. (2004). Peptidoglycan recognition proteins: on and off switches for innate immunity. *Immunol. Rev.* 198, 83–96.
- Steiner, H., Hultmark, D., Engström, A., Bennich, H., and Boman, H.G. (1981). Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature* 292, 246–248.
- Tabeta, K., Georgel, P., Janssen, E., Du, X., Hoebe, K., Crozat, K., Mudd, S., Shamel, L., Sovath, S., Goode, J., et al. (2004). Toll-like receptors 9 and 3 as essential components of innate immune defense against mouse cytomegalovirus infection. *Proc. Natl. Acad. Sci. U.S.A.* 101, 3516–3521. <https://doi.org/10.1073/pnas.0400525101>
- Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D. (2003). A biochemical framework for RNA silencing in plants. *Genes Dev.* 17, 49–63. <https://doi.org/10.1101/gad.1048103>
- Tanji, T., Yun, E.Y., and Ip, Y.T. (2010). Heterodimers of NF-kappaB transcription factors DIF and Relish regulate antimicrobial peptide genes in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14715–14720. <https://doi.org/10.1073/pnas.1009473107>
- Tants, J.N., Fesser, S., Kern, T., Stehle, R., Geerlof, A., Wunderlich, C., Juen, M., Hartlmüller, C., Böttcher, R., Kunzelmann, S., et al. (2017). Molecular basis for asymmetry sensing of siRNAs by the *Drosophila* Loqs-PD/Dcr-2 complex in RNA interference. *Nucleic Acids Res.* 45, 12536–12550. <https://doi.org/10.1093/nar/gkx886>
- Tassetto, M., Kunitomi, M., and Andino, R. (2017). Circulating Immune Cells Mediate a Systemic RNAi-Based Adaptive Antiviral Response in *Drosophila*. *Cell* 169, 314–325.e13. <https://doi.org/10.1016/j.cell.2017.03.033>
- Theopold, U., Krautz, R., and Dushay, M.S. (2014). The *Drosophila* clotting system and its messages for mammals. *Dev. Comp. Immunol.* 42, 42–46. <https://doi.org/10.1016/j.dci.2013.03.014>
- Tomari, Y., Du, T., Haley, B., Schwarz, D.S., Bennett, R., Cook, H.A., Koppetsch, B.S., Theurkauf, W.E., and Zamore, P.D. (2004a). RISC assembly defects in the *Drosophila* RNAi mutant armitage. *Cell* 116, 831–841.
- Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P.D. (2004b). A protein sensor for siRNA asymmetry. *Science* 306, 1377–1380. <https://doi.org/10.1126/science.1102755>
- Treiber, T., Treiber, N., and Meister, G. (2019). Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. *Nat. Rev. Mol. Cell Biol.* 20, 5–20. <https://doi.org/10.1038/s41580-018-0059-1>
- Trettin, K.D., Sinha, N.K., Eckert, D.M., Apple, S.E., and Bass, B.L. (2017). Loquacious-PD facilitates *Drosophila* Dicer-2 cleavage through interactions with the helicase domain and dsRNA. *Proc. Natl. Acad. Sci. U.S.A.* 114, E7939–E7948. <https://doi.org/10.1073/pnas.1707063114>
- Tsuyoyama, K., Tadakuma, H., and Tomari, Y. (2018). Conformational activation of Argonaute by distinct yet coordinated actions of the Hsp70 and Hsp90 chaperone systems. *Mol. Cell* 70, 722–729.e4. <https://doi.org/10.1016/j.molcel.2018.04.010>
- 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. <https://doi.org/10.1126/science.1129333>

- Valli, A., Busnadiego, I., Maliogka, V., Ferrero, D., Castón, J.R., Rodríguez, J.F., and García, J.A. (2012). The VP3 factor from viruses of Birnaviridae family suppresses RNA silencing by binding both long and small RNA duplexes. *PLOS ONE* 7, e45957. <https://doi.org/10.1371/journal.pone.0045957>
- van Cleef, K.W., van Mierlo, J.T., Miesen, P., Overheul, G.J., Fros, J.J., Schuster, S., Marklewitz, M., Pijlman, G.P., Junglen, S., and van Rij, R.P. (2014). Mosquito and *Drosophila* entomobirnaviruses suppress dsRNA- and siRNA-induced RNAi. *Nucleic Acids Res.* 42, 8732–8744. <https://doi.org/10.1093/nar/gku528>
- van den Beek, M., da Silva, B., Pouch, J., Ali Chaouche, M.E.A., Carré, C., and Antoniewski, C. (2018). Dual-layer transposon repression in heads of *Drosophila melanogaster*. *RNA* 24, 1749–1760. <https://doi.org/10.1261/rna.067173.118>
- van Mierlo, J.T., Bronkhorst, A.W., Overheul, G.J., Sadanandan, S.A., Ekström, J.O., Heestermans, M., Hultmark, D., Antoniewski, C., and van Rij, R.P. (2012a). Convergent evolution of argonaute-2 slicer antagonism in two distinct insect RNA viruses. *PLOS Pathog.* 8, e1002872. <https://doi.org/10.1371/journal.ppat.1002872>
- van Mierlo, J.T., Bronkhorst, A.W., Overheul, G.J., Sadanandan, S.A., Ekström, J.O., Heestermans, M., Hultmark, D., Antoniewski, C., and van Rij, R.P. (2012b). Convergent evolution of argonaute-2 slicer antagonism in two distinct insect RNA viruses. *PLOS Pathog.* 8, e1002872. <https://doi.org/10.1371/journal.ppat.1002872>
- van Mierlo, J.T., Overheul, G.J., Obadia, B., van Cleef, K.W., Webster, C.L., Saleh, M.C., Obbard, D.J., and van Rij, R.P. (2014). Novel *Drosophila* viruses encode host-specific suppressors of RNAi. *PLOS Pathog.* 10, e1004256. <https://doi.org/10.1371/journal.ppat.1004256>
- van Rij, R.P., Saleh, M.C., Berry, B., Foo, C., Houk, A., Antoniewski, C., and Andino, R. (2006). The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in *Drosophila melanogaster*. *Genes Dev.* 20, 2985–2995. <https://doi.org/10.1101/gad.1482006>
- Varjak, M., Maringer, K., Watson, M., Sreenu, V.B., Frederick, A.C., Pondeville, E., Donald, C.L., Sterk, J., Kean, J., Vazeille, M., et al. (2017a). *Aedes aegypti* Piwi4 is a noncanonical PIWI protein involved in antiviral responses. *mSphere* 2. <https://doi.org/10.1128/mSphere.00144-17>
- Varjak, M., Donald, C.L., Mottram, T.J., Sreenu, V.B., Merits, A., Maringer, K., Schnettler, E., and Kohl, A. (2017b). Characterization of the Zika virus induced small RNA response in *Aedes aegypti* cells. *PLOS Negl. Trop. Dis.* 11, e0006010. <https://doi.org/10.1371/journal.pntd.0006010>
- Varjak, M., Dietrich, I., Sreenu, V.B., Till, B.E., Merits, A., Kohl, A., and Schnettler, E. (2018). Spindle-E Acts Antivirally Against Alphaviruses in Mosquito Cells. *Viruses* 10, E88. <https://doi.org/10.3390/v10020088>
- Vidal, S., Khush, R.S., Leulier, F., Tzou, P., Nakamura, M., and Lemaître, B. (2001). Mutations in the *Drosophila* dTAK1 gene reveal a conserved function for MAPKKs in the control of rel/NF-kappaB-dependent innate immune responses. *Genes Dev.* 15, 1900–1912. <https://doi.org/10.1101/gad.203301>
- Virgen-Slane, R., Rozovics, J.M., Fitzgerald, K.D., Ngo, T., Chou, W., van der Heden van Noort, G.J., Filippov, D.V., Gershon, P.D., and Semler, B.L. (2012). An RNA virus hijacks an incognito function of a DNA repair enzyme. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14634–14639. <https://doi.org/10.1073/pnas.1208096109>
- Vodovar, N., Bronkhorst, A.W., van Cleef, K.W., Miesen, P., Blanc, H., van Rij, R.P., and Saleh, M.C. (2012). Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. *PLOS ONE* 7, e30861. <https://doi.org/10.1371/journal.pone.0030861>
- Waldron, F.M., Stone, G.N., and Obbard, D.J. (2018). Metagenomic sequencing suggests a diversity of RNA interference-like responses to viruses across multicellular eukaryotes. *PLOS Genet.* 14, e1007533. <https://doi.org/10.1371/journal.pgen.1007533>
- Wang, H., Smagge, G., and Meeus, I. (2017). The role of a single gene encoding the Single von Willebrand factor C-domain protein (SVC) in bumblebee immunity extends beyond antiviral defense. *Insect Biochem. Mol. Biol.* 91, 10–20. <https://doi.org/10.1016/j.ibmb.2017.10.002>
- Wang, X.H., Aliyari, R., Li, W.X., Li, H.W., Kim, K., Carthew, R., Atkinson, P., and Ding, S.W. (2006). RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science* 312, 452–454. <https://doi.org/10.1126/science.1125694>
- Weavers, H., Evans, I.R., Martin, P., and Wood, W. (2016). Corpse engulfment generates a molecular memory that primes the macrophage Inflammatory Response. *Cell* 165, 1658–1671. <https://doi.org/10.1016/j.cell.2016.04.049>
- Weber, F., Wagner, V., Rasmussen, S.B., Hartmann, R., and Paludan, S.R. (2006). Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. *J. Virol.* 80, 5059–5064. <https://doi.org/10.1128/JVI.80.10.5059-5064.2006>
- Welker, N.C., Maity, T.S., Ye, X., Aruscavage, P.J., Krauchuk, A.A., Liu, Q., and Bass, B.L. (2011). Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. *Mol. Cell* 41, 589–599. <https://doi.org/10.1016/j.molcel.2011.02.005>
- Wernet, M.F., Klovstad, M., and Clandinin, T.R. (2014). Generation of infectious virus particles from inducible transgenic genomes. *Curr. Biol.* 24, R107–8. <https://doi.org/10.1016/j.cub.2013.12.009>
- West, C., and Silverman, N. (2018). p38b and JAK-STAT signaling protect against Invertebrate iridescent virus 6 infection in *Drosophila*. *PLOS Pathog.* 14, e1007020. <https://doi.org/10.1371/journal.ppat.1007020>
- Wieschaus, E., and Nüsslein-Volhard, C. (2016). The Heidelberg SCREEN FOR PATTERN MUTANTS OF *Drosophila*: A personal account. *Annu. Rev. Cell Dev. Biol.* 32, 1–46. <https://doi.org/10.1146/annurev-cellbio-113015-023138>
- Wong, M.T., and Chen, S.S. (2016). Human Choline Kinase- α Promotes Hepatitis C Virus RNA Replication through Modulation of Membranous Viral Replication Complex Formation. *J. Virol.* 90, 9075–9095. <https://doi.org/10.1128/JVI.00960-16>

- Wong, M.T., and Chen, S.S. (2017). Hepatitis C Virus Subverts Human Choline Kinase- α To Bridge Phosphatidylinositol-4-Kinase III α (PI4KIII α) and NSSA and Upregulates PI4KIII α Activation, Thereby Promoting the Translocation of the Ternary Complex to the Endoplasmic Reticulum for Viral Replication. *J. Virol.* *91*, e00355-17. <https://doi.org/10.1128/JVI.00355-17>
- Wu, Q., Luo, Y., Lu, R., Lau, N., Lai, E.C., Li, W.X., and Ding, S.W. (2010). Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc. Natl. Acad. Sci. U.S.A.* *107*, 1606–1611. <https://doi.org/10.1073/pnas.0911353107>
- Wu, X., Wu, F.H., Wang, X., Wang, L., Siedow, J.N., Zhang, W., and Pei, Z.M. (2014). Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. *Nucleic Acids Res.* *42*, 8243–8257. <https://doi.org/10.1093/nar/gku569>
- Ye, X., Huang, N., Liu, Y., Paroo, Z., Huerta, C., Li, P., Chen, S., Liu, Q., and Zhang, H. (2011). Structure of C3PO and mechanism of human RISC activation. *Nat. Struct. Mol. Biol.* *18*, 650–657. <https://doi.org/10.1038/nsmb.2032>
- Yoon, J.S., Mogilicherla, K., Gurusamy, D., Chen, X., Chereddy, S.C.R.R., and Palli, S.R. (2018). Double-stranded RNA binding protein, Staufen, is required for the initiation of RNAi in coleopteran insects. *Proc. Natl. Acad. Sci. U.S.A.* *115*, 8334–8339. <https://doi.org/10.1073/pnas.1809381115>
- Yu, Y., Gu, J., Jin, Y., Luo, Y., Preall, J.B., Ma, J., Czech, B., and Hannon, G.J. (2015). Panoramix enforces piRNA-dependent cotranscriptional silencing. *Science* *350*, 339–342. <https://doi.org/10.1126/science.aab0700>
- Yuan, Y.R., Pei, Y., Ma, J.B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.Y., Dauter, Z., Tuschl, T., and Patel, D.J. (2005). Crystal structure of *A. aeolicus* argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol. Cell* *19*, 405–419. <https://doi.org/10.1016/j.molcel.2005.07.011>
- Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* *101*, 25–33. [https://doi.org/10.1016/S0092-8674\(00\)80620-0](https://doi.org/10.1016/S0092-8674(00)80620-0)
- Zhang, S.Y., Jouanguy, E., Ugolini, S., Smahi, A., Elain, G., Romero, P., Segal, D., Sancho-Shimizu, V., Lorenzo, L., Puel, A., *et al.* (2007). TLR3 deficiency in patients with herpes simplex encephalitis. *Science* *317*, 1522–1527. <https://doi.org/10.1126/science.1139522>
- Zhang, X., Guo, R., Kumar, D., Ma, H., Liu, J., Hu, X., Cao, G., Xue, R., and Gong, C. (2016). Identification, gene expression and immune function of the novel Bm-STAT gene in virus-infected *Bombyx mori*. *Gene* *577*, 82–88. <https://doi.org/10.1016/j.gene.2015.11.027>
- Zhou, R., Czech, B., Brennecke, J., Sachidanandam, R., Wohlschlegel, J.A., Perrimon, N., and Hannon, G.J. (2009). Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform. *RNA* *15*, 1886–1895. <https://doi.org/10.1261/rna.1611309>
- Zografidis, A., Van Nieuwerburgh, F., Koliopoulou, A., Apostolou-Karampelis, K., Head, S.R., Deforce, D., Smagghe, G., and Swevers, L. (2015). Viral Small-RNA Analysis of *Bombyx mori* larval midgut during persistent and pathogenic cytoplasmic polyhedrosis virus infection. *J. Virol.* *89*, 11473–11486. <https://doi.org/10.1128/JVI.01695-15>
- Zotti, M., Dos Santos, E.A., Cagliari, D., Christiaens, O., Taning, C.N.T., and Smagghe, G. (2018). RNA interference technology in crop protection against arthropod pests, pathogens and nematodes. *Pest Manag. Sci.* *74*, 1239–1250. <https://doi.org/10.1002/ps.4813>

Outro (*term commonly heard in music designing the opposite of an introduction*)

While extensive work is being done at trying to unravel new sensors of viruses in drosophila, my Ph.D work focused on the characterization of the only known sensor of viral nucleic acids that fuels the antiviral RNAi pathway: Dicer-2. *In vitro* studies previously proposed models for Dicer-2 mechanisms of action on dsRNA substrates depending on their nature, length and termini but also on interaction with several cofactors. All these parameters that can be tightly controlled in *in vitro* conditions are as many unknowns when it comes to *in vivo* studies. Indeed, although endogenous, exogenous and viral Dicer-2 targets have been identified, little is known about the exact characteristics of these dsRNAs. Thus, two main questions guided the writing of this manuscript:

How Dicer-2 accesses its substrates and what are their characteristics?

Because of its dual role in the endo-siRNA and antiviral RNAi pathways, Dicer-2 must be able to sense and discriminate a potentially wide diversity of dsRNA molecules. The detection of viral RNAs is made even more difficult by their high adaptability potential led by the mutation rate of their replicating enzyme. Thus, a lot of counter mechanisms have evolved and complicate their detection. Still, Dicer-2 is able to access many viruses and process the dsRNA molecules formed during their infection cycle. How and where does Dicer-2 manages to bypass viral defenses (**Chapter I**)? What are the characteristics of these potential weak points (**Chapter II**)? I tried to answer these questions by using the dicistroviruses Drosophila C Virus and Cricket Paralysis Virus as infection models.

What is the role of Dicer-2 helicase and its associated cofactors?

It is striking that proteins involved in the antiviral immunity of diverse organisms all possess the same specific helicase domain organization. In the case of mammalian RIG-like receptors for instance, this domain was shown to be required for recognition of their target and regulation of their activity. Thus, involvement of drosophila Dicer-2 helicase domain in the sensing of its viral and endogenous RNA targets was investigated. In addition, two distinct mutations of this domain were studied in an attempt to uncouple its possible sensing role from its ATPase activity (**Chapter III**).

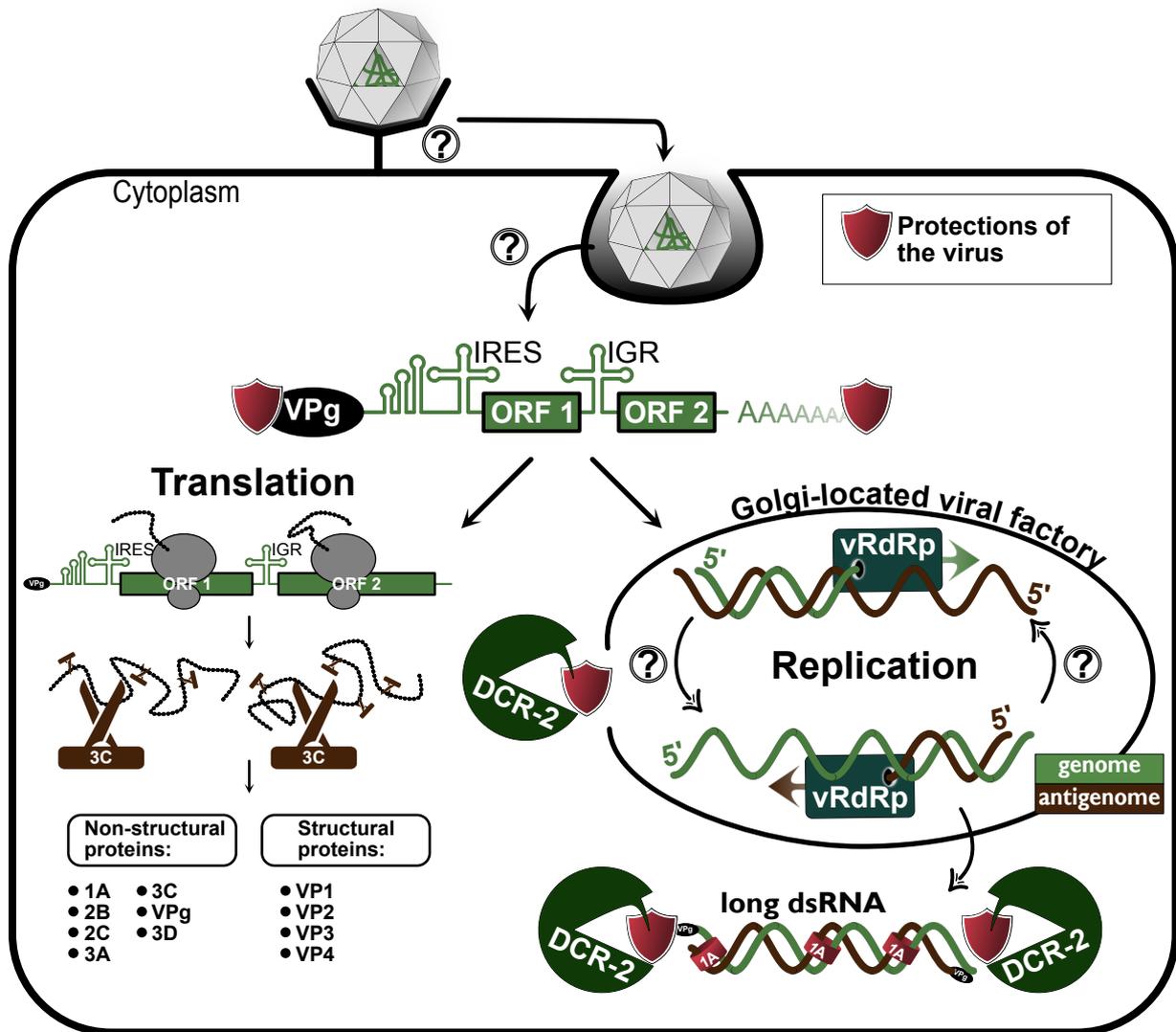


Figure 1: Drosophila C Virus life cycle as an example for dicistroviruses.

When DCV enters the host cell, it releases its protected viral genome in the cytoplasm. A viral protein covalently linked to the 5' extremity (VPg) and a poly-A tail at the 3' extremity ensure protection against 5'-3' and 3'-5' RNA degradation mechanisms. Cap-independent translation is mediated by the host ribosome on the two ORFs independently. The two resulting polyproteins are cleaved by the 3C-like protease generating structural and non-structural mature proteins.

Brief description of the role of each non-structural protein: **1A** = inhibition of RNAi; **2B** = predicted membrane remodeling activity; **2C** = predicted role as RNA helicase and/or membrane permeabilization; **3A** = predicted role in viral replication complex formation; **3C** = viral poly protein processing; **3D** = RNA dependent RNA polymerase; **VPg** = covalently linked to the 5' end of genomic RNA to protect it from degradation and initiate protein-primed viral replication.

During the viral replication, the RdRP catalyzes the syntheses of genomic and antigenomic strands that lead to the formation of a dsRNA intermediate of replication (further explained in Chapter II). Dicistroviruses encode for viral suppressors of RNAi with various mechanisms of action. In the case of DCV, its suppressor of RNAi (DCV-1A) coats the long dsRNA intermediate of replication and inhibits its processing by Dicer-2.

Introduction

Dicistroviruses are non-enveloped viruses with a relatively small (~8-10kb) monopartite positive ssRNA genome and a small (~30nm) icosahedral capsid. They have a broad tropism in the Arthropoda phylum, are tremendously diverse and represent major agricultural and economic threats worldwide. Due to their characteristics and to the similarity in term of diseases symptoms (e.g. paralysis), they were initially classified in the *Picornaviridae* family. However, their genomic organization do not compare which justifies the classification of these viruses in distinct families. Indeed, picornaviruses have a monocistronic genome encoding structural proteins first and then non-structural proteins while dicistroviruses have a bicistronic genome encoding non-structural proteins in ORF1 and structural proteins in ORF2. Despite the extensive study of dicistroviruses interactions with their host, many gaps remain in our understanding of the mechanisms behind the different infection steps.

Most of the mechanistic details of dicistroviruses infection cycle come from studies conducted using *Drosophila* and two viruses from the *Cripavirus* genus: Drosophila C Virus (DCV) and Cricket Paralysis Virus (CrPV). Extended description of dicistroviruses entry, replication, translation and packaging was reviewed in Warsaba et al., 2019. Briefly, viruses most likely first attach to a cell surface receptor not identified yet and enter the cell using the clathrin-dependent endocytosis pathway (Cherry and Perrimon, 2004; Yang et al., 2018). Then, the viral genome is released in the cytoplasm and targeted for translation and replication. Cap-independent translation of the two ORFs is mediated by the hijacked host ribosome and produces two polyproteins (Moore et al., 1980, 1981). Further processing of these polyproteins by the virally encoded 3C-like protease give rise to non-structural (ORF1) and structural (ORF2) mature proteins (Nakashima and Ishibashi, 2010; Nakashima and Nakamura, 2008). In parallel, replication is conducted by the viral RNA-dependent RNA polymerase interacting with host proteins and the 5' viral protein genome-linked VPg. The mechanism by which sense and antisense strands are produced is still obscure and was mostly inferred from studies in closely-related picornaviruses (further explained in Chapter II and reviewed in Paul and Wimmer, 2015). Finally, dicistrovirus family encompasses lytic and non-lytic viruses which complicates our understanding of their transmission strategies (*Figure 1*).

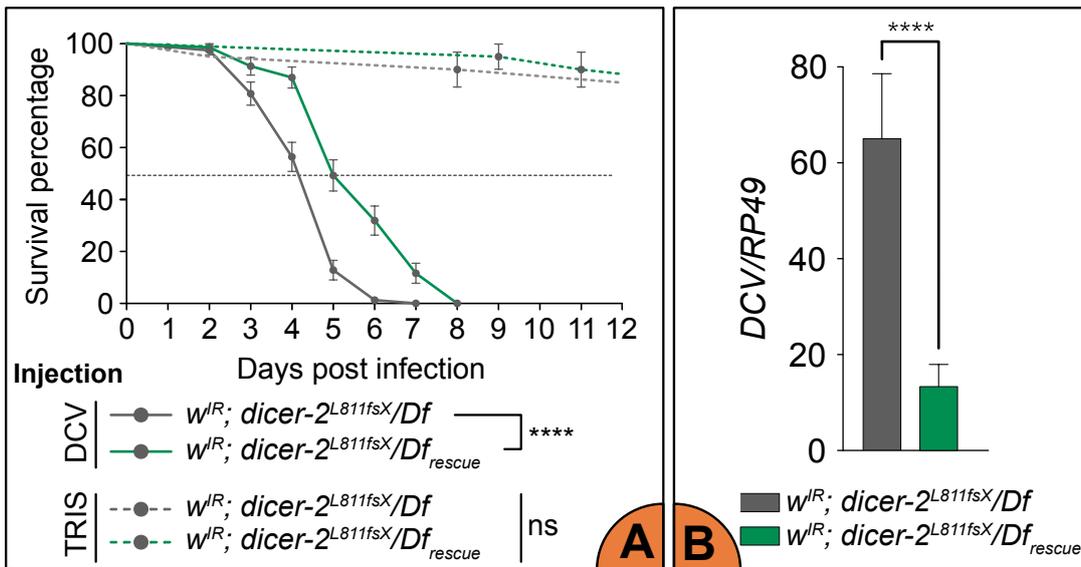


Figure 2: Dicer-2 plays an important role in the defense against DCV.

A) Survival of *dicer-2* wild-type ($w^{IR}; dicer-2^{L811fsX}/Df_{rescue}$) and *dicer-2* null mutant ($w^{IR}; dicer-2^{L811fsX}/Df$) flies after DCV (500PFU) or TRIS injection (n=60; statistical test: logrank). In both A and B, error bars represent standard deviation. **B)** Relative DCV RNA accumulation measured by RT-qPCR in control and *dicer-2* null mutant flies 3d post infection with DCV 500PFU. (n=2, biological triplicates, statistical test: one way ANOVA).

As presented in the general introduction of this manuscript, insects and most specifically *Drosophila* fruit flies are able to mount antiviral responses against a wide diversity of viruses. Several evolutionarily conserved innate immune mechanisms could be identified as antiviral, namely the JAK-STAT, Toll, IMD and Heat-Shock proteins pathways. In addition to these, constitutively expressed restriction factors can also participate in the control of viruses in insect. While all these defense mechanisms are most of the time virus specific, the small interfering RNA pathway remains the major broad antiviral response in insects. As previously explained, Dicer-2 is the only known sensor of viral nucleic acids and activator of the antiviral siRNA pathway in *Drosophila*. This enzyme is described as a dsRNA endonuclease entering from its substrate extremity and generating virus-derived siRNAs. Impaired expression of this catalytic enzyme results in a higher viral load as well as a reduced survival rate of flies injected with the dicistrovirus DCV (*Figure 2* and Galiana-Arnoux et al., 2006; van Rij et al., 2006). Thus, Dicer-2 must be able to sense and dice (i.e. cleave) the dsRNA intermediate of replication produced by DCV (*Figure 1*). However, dicistroviruses are highly protected and hidden from the immune system and we still do not understand how Dicer-2 is able to detect them.

First, dicistrovirus genomic RNAs are protected against the Pacman-mediated 5'-3' cellular RNA decay machinery by a viral protein genome-linked (VPg) at their 5' extremity (King and Moore, 1988; Nakashima and Shibuya, 2006). The lack of a 7-methylguanosine mRNA cap at the 5' extremity of their genome also protects them from Dcp1 or Dcp2 decapping enzymes. In addition to its protective role, the VPg is proposed as a protein primer used by the viral RdRp to initiate the synthesis of genomic and antigenomic strands (Paul and Wimmer, 2015). Similarly, it is thought that the 3' poly A tail of dicistroviruses genomes is bound by cellular poly A binding proteins to promote circularization of the viral RNA and subsequent antigenome-strand synthesis by the RdRp (Herold and Andino, 2001). This poly A tail is identical to the one present at the 3' extremity of cellular mRNAs and most likely protects the viral genome against exosome-mediated 3'-5' degradation. In addition to these genomic stabilizing features, the replication of dicistroviruses often occurs on host membranes structures. In the case of DCV, the remodeling of the Golgi apparatus results in the formation of small viral replication factories (~115nm diameter) that are hard to access for the immune system (Cherry et al., 2006). Finally, viruses have evolved their own defense mechanisms against RNAi, namely viral suppressors of RNAi (VSRs). Numerous strategies have evolved to

counter the antiviral RNAi at multiple levels and even closely related dicistroviruses like DCV and CrPV encode for VSRs that have different modes of action. While DCV 1A binds viral dsRNA intermediate of replication to block Dicer-2 processing and RISC loading, CrPV 1A directly interacts with the effector protein Ago2 to suppress its cleavage activation and to target it for proteosomal degradation (Nayak et al., 2010, 2018; van Rij et al., 2006; Watanabe et al., 2017). As a consequence, dicistroviruses appear to be extremely well protected against antiviral immune pathways (*Figure 1*).

Despite all these layers of protection, strong evidences show that Dicer-2 plays an important role in the defense against dicistroviruses. Thus, one major question drove my Ph.D during four years: **how is Dicer-2 able to detect protected RNAs from dicistroviruses?**

One of the hallmarks of antiviral RNAi is that the virus-derived siRNAs produced by Dicer-2 provide a footprint of the action of the immune system. Thus, being able to identify a siRNA signature could provide us with mechanistic insights on the detection and entry of Dicer-2 on viruses. The current method of choice to investigate small RNAs is undoubtedly small RNA high-throughput sequencing (HTS). During my Ph.D, I used two complementary approaches based on this technology in order to decipher Dicer-2 entry point on two dicistroviruses, DCV and CrPV. First, a kinetic study of the apparition of siRNAs was performed using the simplified S2 cells model. These macrophage-like cells can be infected in a synchronized manner, making possible the identification of siRNAs at early time points of infection. Then, small RNA HTS of DCV infected flies expressing different variants of Dicer-2 was performed to check for *in vivo* relevance of the data obtained in cells and to gain insights in the molecular mechanisms at play.

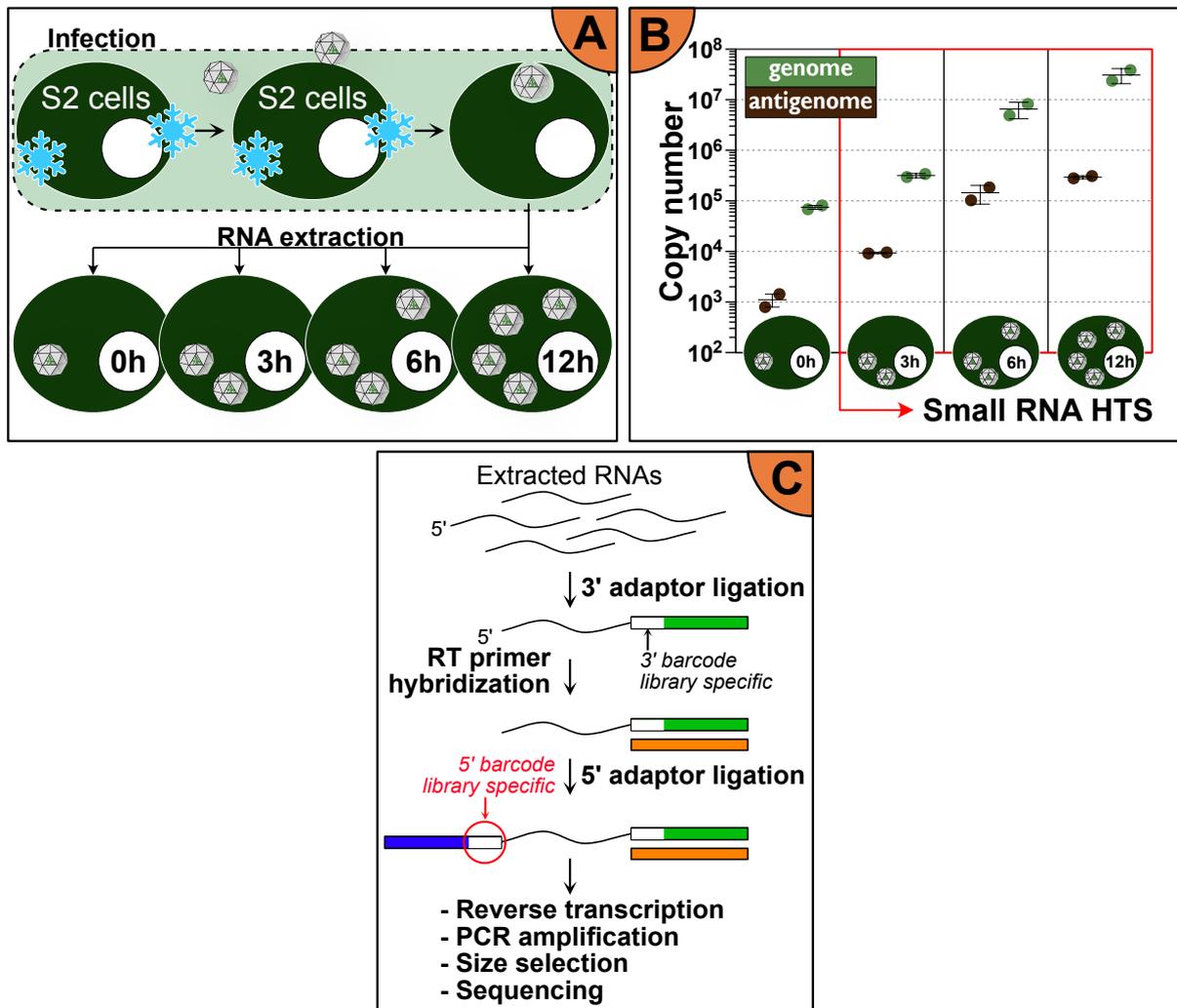


Figure 3: Preparation of the RNAs for small RNA HTS.

A) Schematic representation of the experimental setup for S2 cells synchronized infection. Cells are kept on ice while the infective solution is being added (DCV MOI 10 or CrPV MOI 0.05). Viruses bind to the cell membranes but are not able to internalize. Wash of the unbound viruses is performed before removing cells from ice. Finally, cells are put in the incubator ($t=0h$ of the synchronized infection) and the RNAs are extracted after 0, 3, 6 or 12h of infection. **B)** Strand-specific RT-qPCR designed to detect separately genome and antigenome DCV strands. The y axis is the absolute strand copy number. The error bars represent the standard deviation between the biological duplicates ($n=1$). Time points 3 -12h were used to perform small RNA HTS. **C)** Schematic representation of the **adapted** Illumina protocol to produce double-indexed small RNA libraries.

Results

I. Kinetics of siRNAs apparition in S2 cells during DCV and CrPV infection

a. Preliminary work

Synchronized DCV infection of S2 cells

A detail of uttermost importance for this study was that all cells had to be infected at the same time in order for the subsequent RNA sequencing to depict the situation at precise time points. Thus, the preparation of S2 cells infection by DCV was performed on ice to prevent cell-bound viruses to internalize and release their genome in a non-synchronized manner (*Figure 3A*). Moreover, the infection was carried on with a MOI of 10 to make sure that all the cells would be infected with at least one viral particle. All the unbound viruses were washed before starting the incubation at t=0h.

Thereafter, the presence of DCV RNAs was checked at four different time points (0, 3, 6 and 12h pi) using strand specific RT-qPCR to detect separately genomic and antigenomic strands (*Figure 3B*). To prevent false priming effect of this technic that would result in an inaccurate measurement of the antigenomic strand, we used tag-specific primers (Plaskon et al., 2009; Tuiskunen et al., 2010). DCV being a positive ssRNA virus, it is not surprising to observe a majority of genomic strands at 0h pi. However, the presence of antigenomic strands at this time point might reflect a slight contamination of the purified DCV viral stock with antigenomic strands and will be discussed later. The following time points (3-12h pi) show an occurring productive infection with the virus replicating at an exponential rate and keeping a ~100-fold difference between genome and antigenome strand abundance. The experiment was conducted in biological duplicates and two independent libraries for each 3, 6 and 12h pi time points were constructed to be sent for Illumina HTS (*Table 1*).

Sample bleeding: definition, importance and solution

Illumina-based next generation sequencing drastically increased the output number of reads and thus, the sequence coverage one could expect from a sequencing experiment. Many different procedures have been developed to perform DNA, short RNA or long RNA sequencing. The possibility to individually label different samples made the

Sequencing lane	Virus	Time point	Seq ID	Total number of reads	Trimming surviving	Second demultiplexing surviving	21-22-23 miRNAs (1M)	Normalization factor
#1	DCV	3h	SNBN482	35461457	9363086	7935011	200274	0,76
			SNBN483	32802208	15394917	14497894	221606	0,84
		6h	SNBN484	28076461	9602839	8610308	133291	0,50
			SNBN485	35186428	15297869	13887545	278362	1,05
		12h	SNBN486	31820875	11346320	10016871	150050	0,57
			SNBN487	23906611	7446817	6716570	123381	0,47
	CrPV	3h	SNBN488	36781012	19634757	18213046	440355	1,66
			SNBN489	31118678	13911153	12835867	229097	0,86
		6h	SNBN490	77777124	44180319	41874789	587163	2,21
			SNBN491	47964471	12067871	11039456	302240	1,14
		12h	SNBN492	46643861	19461501	17634468	374307	1,41
			SNBN493	23611276	9064387	8429532	142916	0,54

Mean =
265253,5

Table 1: Summary table of the small RNA HTS performed on libraries constructed from DCV or CrPV infected S2 cells extracted RNAs.

All the samples were multiplexed and sequenced in the same sequencing lane. Biological duplicates were done for each experimental condition, hence the preparation of 12 libraries for 6 experimental conditions. **Total number of reads:** the number of all small RNA reads in the fastq file delivered by the sequencing platform. This number is obtained after demultiplexing all the reads from the sequencing lane according to the 3' adaptor barcode. **Trimming surviving:** all the reads that were not identified as primers dimers or 2S ribosomal RNA. **Second demultiplexing surviving:** all the reads sequenced with the custom 5' adaptor barcode expected from this specific library reads. **21-22-23 miRNAs (1M):** alignment of the surviving reads on all miRNA hairpin sequences was done with 1 mismatch allowed. The miRNA identified reads of 21-23nt long were used for our normalization strategy (see Materials & Methods). **Normalization factor:** factor applied to data of each individual library of the sequencing lane in order to make comparisons between them possible. Of note, all the study was first performed without normalization of the data and showed similar results.

multiplexing of samples a common practice in laboratories. However, it is now of public knowledge that this method has a 0.1 to 10% misattribution rate of reads to the wrong libraries (Griffiths et al., 2018; Sinha et al., 2017; van der Valk et al., 2018). This phenomenon called “index hopping” or more dramatically “sample bleeding” was mostly identified in long RNA sequencing or DNA sequencing and results in a cross contamination of all the libraries ran in the same sequencing lane. In most cases, sample bleeding is not an issue as the high number of reads we are looking at strongly reduces the impact of this technical flaw. However, in our case, libraries made from cells infected for 3h were going to be sequenced in the same lane as libraries made from cells infected for 12h. Thus, the danger was that the 0.1-10% spillover of the 12h libraries would mask the small number of virus derived small RNAs generated at 3h.

One of the experimental solutions found so far to minimize the impact of sample bleeding is double indexing (Kircher et al., 2012). This method relies on the use of different 5' adaptors than the ones originally used in Illumina protocols and kits. These adaptors contain a 6nt long internal barcode that will be sequenced together with the small RNA read and later used to perform a second demultiplexing (*Figure 3C*). Double indexing theoretically increases the confidence in the attribution of reads to specific libraries up to 99.99%. Thus, even though sample bleeding was not proven yet in small RNA HTS, a double indexing strategy was used to prepare the libraries from virus infected S2 cells.

Trimming and second demultiplexing of the libraries

The libraries were constructed using the double indexing strategy described above which allowed the safe multiplexing of all the samples in a single sequencing lane (*Table 1*). It has to be noted that this lane contained libraries constructed from DCV infected cells but also from CrPV infected cells extracted RNAs that will be used later on. All libraries were considered as a whole for the trimming, second demultiplexing and miRNA-based normalization (see Materials and Methods). That is why the graphics corresponding to these steps will display the 12 libraries.

Between 43 to 74% of the total reads obtained were reads mapping on the drosophila 2S ribosomal RNA sequence or adapters dimers (*Figure 4A e³ Table 1*). These reads were dropped during the trimming of the libraries. Among each library, we searched for the surviving reads presenting a wrong internal label (*Figure 4B e³ Table 1*). Most of the reads that did not have a perfect match with the corresponding internal label were reads

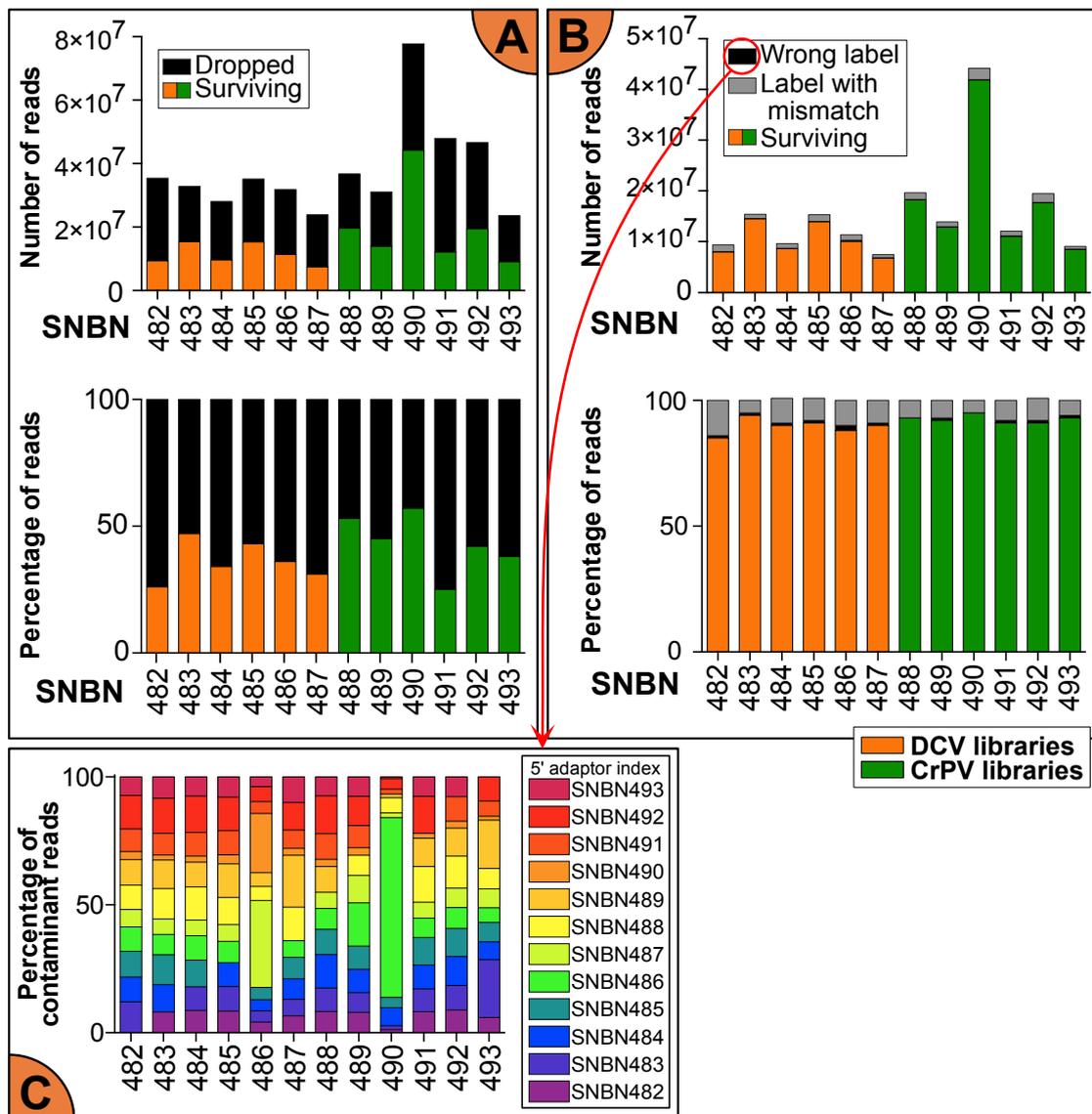


Figure 4: Sample bleeding demonstration and libraries trimming.

A) Trimming: reads corresponding to adapters dimers or drosophila 2S ribosomal RNA were removed from the data. **B) Second demultiplexing:** reads with either mismatches in the 5' internal label or internal label from another library were removed from the data. For **A** and **B**, the two graphics are different representations of the same data. **C)** Graphical representation of the "sample bleeding" observed in the small RNA HTS. On this graphic, only the reads with a 5' internal label from another library are represented.

SNBNXXX: library identification number

SNBN482-487: libraries constructed from **DCV** infected S2 cells extracted RNAs.

SNBN488-493: libraries constructed from **CrPV** infected S2 cells extracted RNAs.

presenting at least one mismatch in it, most likely coming from sequencing errors. Even though these reads could have been quite safely attributed to their library, they were discarded from the datasets. However, between 1 to 2% of the reads presented an internal label corresponding to another library. This result justifies the use of a second demultiplexing method and shows the existence of sample bleeding in Illumina-based small RNA HTS. These contaminating reads are coming equally from all the other libraries and were removed from the datasets (*Figure 4C e³ Table 1*).

The library SNBN490 (CrPV – 6h pi) represents an interesting exception to this rule that requires further explanation. As depicted by *Figure 4B*, sample bleeding of SNBN490 into the other libraries is the weakest. This does not correlate with the fact that the original total number of reads of SNBN490 is much higher and should result in a more important bleeding in the other libraries (*Figure 4A e³ Table 1*). Moreover, the distribution of contaminating reads in SNBN490 does not look like all the other libraries present in the same sequencing lane. Investigations led us to discover that SNBN490 library was sequenced two times and in two different flow cells at the sequencing platform. Thus, the few contaminant reads observed only come from the first sequencing flow cell that contained all our libraries. The second sequencing of SNBN490 was performed in a flow cell that contained other libraries. That is why the reads obtained from this second flow cell could not spillover in our libraries. This interesting exception once again proves the existence and importance of sample bleeding when performing multiplexed small RNA sequencing and end up being a good internal control of the experiment.

The reads that survived both trimming and second demultiplexing (between 22 to 54% of the reads depending on the library) were confidently attributed to each independent library and further used in the bioinformatic analysis.

b. Dicer-2 has a precise entry point on the viral dsRNA of DCV 5' region

Dicer-2 produces a strong siRNA signature in the 5' region of DCV

When processing dsRNA, Dicer-2 produces siRNA duplexes that can be viewed as its footprint or signature. These molecules are composed of two 21nt long RNA strands that are paired with a 2nt 3' overhang. A method of choice to detect Dicer-2 signature and try to decipher its mode of action is small RNA HTS.

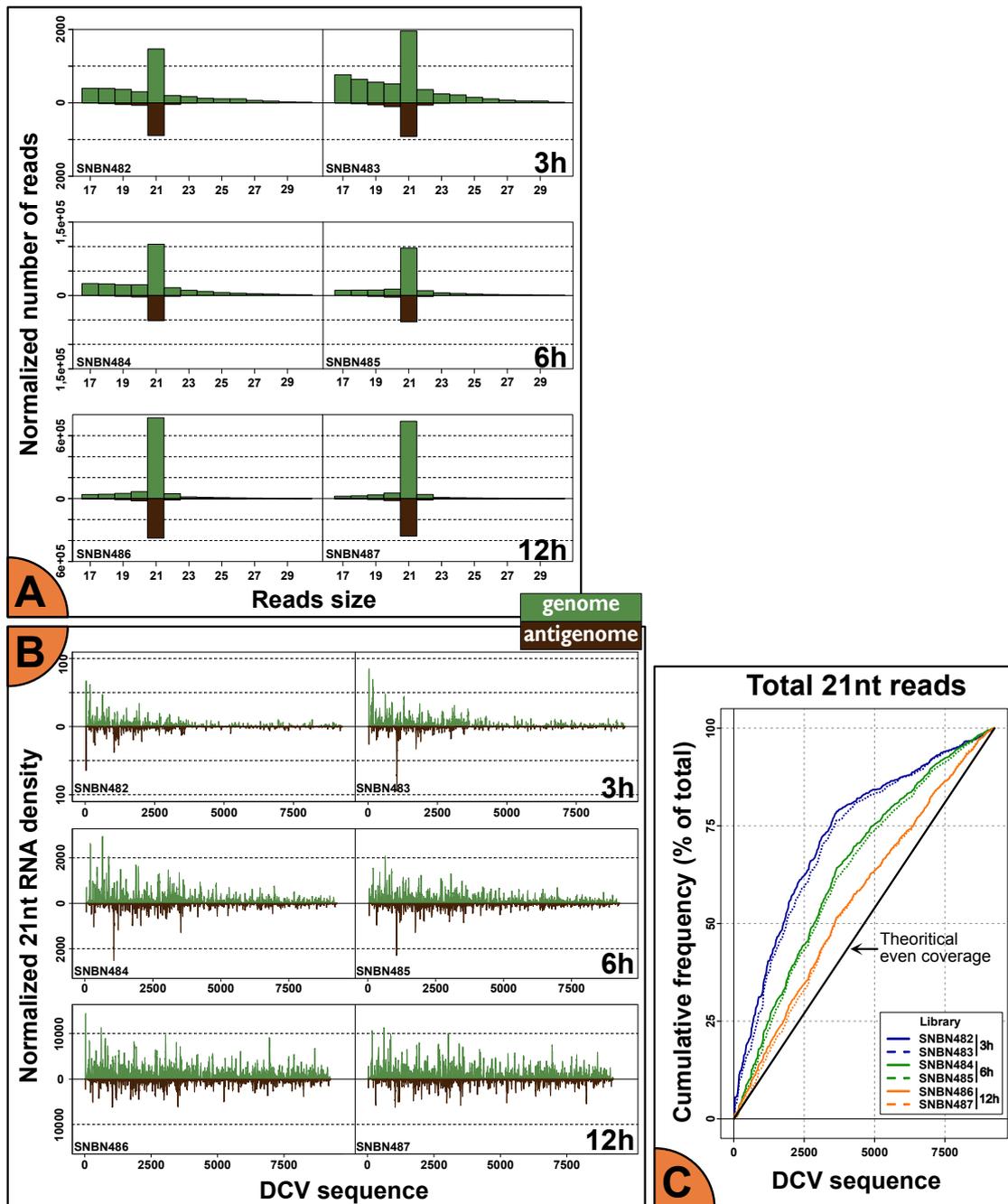


Figure 5: Kinetics of the apparition of DCV mapping 21nt long reads.

A) Normalized DCV mapping reads size distribution. One mismatch was allowed for the alignment. **B)** Normalized DCV coverage by 21nt long reads. **C)** Cumulative frequency of DCV coverage by 21nt long reads. The black line represents what would be a perfectly even coverage of the sequence.

We analyzed libraries SNBN482-487 that were constructed with RNAs extracted at different time points from DCV infected S2 cells (*Table 1*). Alignment of the surviving reads on DCV sequence (1 mismatch allowed) was performed and discriminated between virus derived small RNAs from cell derived RNAs. Because not all libraries had an initial equivalent number of reads, a miRNA-based normalization method was used to make the comparison between libraries possible (*Table 1* & Materials and Methods). Of note, all the study was first performed without any normalization of the data and depicted similar results. The size distributions of DCV mapping reads show a consistent peak at 21nt (*Figure 5A*). The number of these 21nt long reads increases over time and most likely depicts the action of Dicer-2 on its substrate.

An enrichment of DCV mapping 21nt long reads can be identified in the 5' region of the viral sequence after 3h of infection (*Figure 5B*). As the infection continues (6 - 12h pi), this enrichment becomes less and less pronounced to finally end up in a steady state-like situation with an even coverage of DCV genome and antigenome. Another way to look at these data is to plot the cumulative frequency of DCV coverage by 21nt long reads (*Figure 5C*). The advantage of this representation is to make the comparison between the different time points easier because the normalization is based on the percentage of total reads for each library. However, in this case, genomic and antigenomic mapping reads are gathered in the same curve. With this representation, an even coverage of the viral sequence by 21nt long reads would be depicted by a diagonal. This graphic shows (1) high similarity between biological duplicates, (2) enrichment for 21nt long reads in the 5' region of DCV at early time points and (3) flattening of the 6 and 12h pi curves reflecting a tendency to reach an even coverage of the sequence. These two different representations both suggest an early entry point of Dicer-2 in the 5' region of DCV.

To make sure that the 21nt long reads observed are *bona fide* Dicer-2 products, we needed to check whether we could identify two important features of the siRNA signature: (1) the 2nt 3' overhang between genomic and antigenomic strands of the duplex and (2) phasing of the reads. Indeed, *in vitro*, Dicer-2 has been shown to produce siRNAs that are directly following each other and are in phase. Investigation for the first feature (2nt 3' overhang) was done through offset analysis of the 21nt long reads. It has to be noted that the offset and following phasing analyses were done on windows of 500nt long. Analysis of a bigger subsets of the libraries is complicated because the computational time required

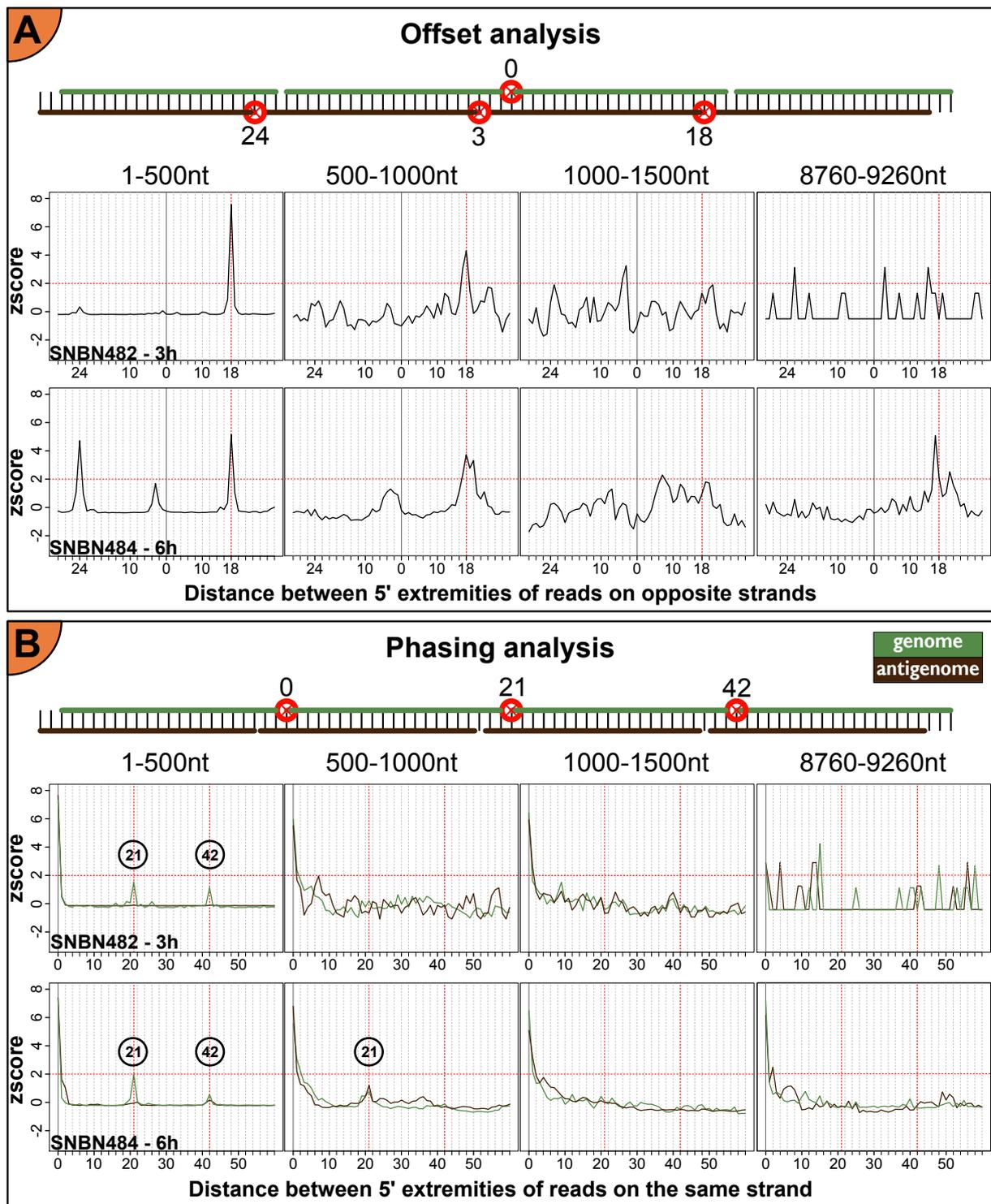


Figure 6: Detection of phased siRNA duplexes.

A) Offset analysis of 2 libraries representative of their corresponding time point. Offset analyses calculates the statistical distance between the 5' extremities of reads that are on opposite strands. **B)** Phasing analysis of 2 libraries representative of their corresponding time point. Phasing analyses calculates the statistical distance between the 5' extremities of reads that are on the same strand. For **A** and **B**, analysis were performed on 21 nt long reads mapping in a 500nt long windows because of computational power issues. The 12h time point presented too many reads making the analysis impossible.

to make the analysis increases exponentially with the number of reads increasing linearly (hence the lack of graphics for the 12h time point in *Figure 6*).

Offset analysis calculates the statistical distance between the 5' extremities of reads that are on opposite strands. As depicted by the cartoon in *Figure 6A*, a distance of 18nt between the 5' extremities of opposite strands is representative of the 2nt 3' overhang feature of siRNA duplexes. In addition, a distance of 3 or 24nt would represent siRNA duplexes that are in phase. A peak at 18nt can be detected when looking at the most 5' 500 nucleotides of DCV sequence 3 and 6h pi (*Figure 6A*). In the same window (first 500nt), peaks at 24 and 3nt can be observed 6h pi. This signal progressively dims with the analysis window being shifted toward the 3' extremity. Thus, this offset analysis by itself already confirms a strong phased Dicer-2 signature in the very 5' region of DCV.

We then wanted to independently check for the phasing signature of a processive Dicer-2, hence the phasing analysis done on the 21nt long reads. Phasing analysis calculates the statistical distance between the 5' extremities of reads that are on the same strand. As depicted by the cartoon in *Figure 6B*, distances of 21 and 42nt are representative of reads in phase. Contrary to the offset analysis, phasing analysis is done independently on genomic and antigenomic strand mapping reads. Analysis of the first 500 nucleotides shows us two peaks at 21 and 42nt for both time points analyzed (*Figure 6B*). This phasing signature is mostly observed on the genome and is quickly lost when looking at the following 500 nucleotides in the 3h pi time point. A peak at 21nt is still visible in this window of analysis for the 6h pi time point but is lost in the following window. Thus, this analysis highlights a strong phasing of the 21 nt long reads mapping on the 5' genomic strand of DCV.

Together, these analyses suggest that Dicer-2 is able to access and processively dice the 5' region of DCV dsRNA as early as 3h pi. If a dsRNA extremity was the only requirement for Dicer-2 to enter, one could also expect a processive siRNA signature in the 3' region of DCV. However, offset and phasing analyses performed on the last 500 nucleotides of DCV 3' extremity did not show an siRNA duplex signature (*Figure 6A e' B*). Thus, Dicer-2 accessibility to DCV dsRNA is somehow initially restricted to its 5' region.

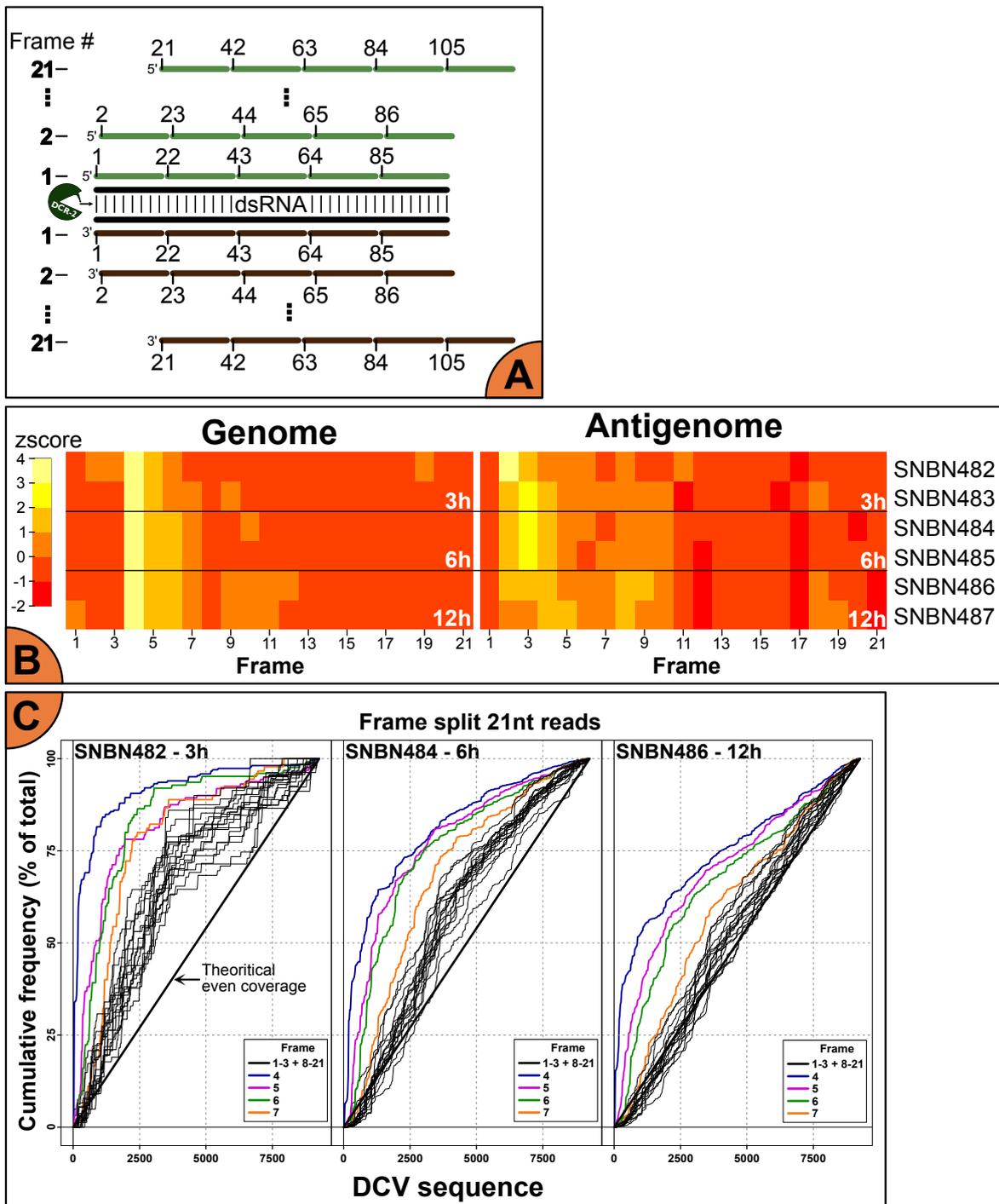


Figure 7: Frame-based analyses highlights a precise entry point of Dicer-2 in the 5' region of DCV.

A) Schematic explanation of the "frame" notion. All the 21nt long reads directly following each other are considered in phase. The frame number attributed to this phase depends on the distance to the 5' extremity of the viral genome. For genomic (resp. antigenomic) reads it is their 5' (resp. 3') extremity that is considered to sort them by frame. **B)** Frame enrichment calculated individually for DCV genome and antigenome mapping 21nt long reads. **C)** Cumulative frequency of DCV coverage by independent frames. The three showed libraries are representative of their respective time point. The black line represents what would be a perfectly even coverage of the sequence.

Dicer-2 has a precise entry point on viral dsRNA

The previous analysis showed that Dicer-2 is dicing the viral dsRNA in a processive manner, in its 5' region and at early time points. We can further characterize the phased reads by their relative distance to the 5' extremity of the viral genome. Thus, if reads are generated in phase starting from the 5' extremity of the viral dsRNA, we will consider these reads as in frame 1 (*Figure 7A*). Because Dicer-2 products are 21nt long, one can distinguish 21 possible dicing frames. Thereafter, the following hypothesis was conjectured: if (1) Dicer-2 is working in a processive manner and if (2) it has a precise entry point on the viral dsRNA, we should observe a dominant frame in the DCV mapping 21nt long reads.

Frame enrichment was calculated for DCV genome and antigenome mapping 21nt long reads (*Figure 7B*). Consistently between biological duplicates and time points, the dominant frame in the reads mapping on DCV genome is frame 4. One can also notice the abundance of reads in the following frames 5 and 6. Doing this analysis on the antigenome mapping reads shows us an enrichment of frames 2 – 4. As a reminder, Dicer-2 produces small RNA duplexes with 2nt 3' overhang. Thus, this frame enrichment analysis revealed a dominant frame for DCV-derived siRNA duplexes, namely frame “4/2”. This information is in accordance with the proposed hypothesis and suggests a precise entry point of Dicer-2 on viral dsRNA. Moreover, if this entry point was located at the very 5' extremity of viral dsRNA, we could expect the frame 1/20 or 3/1 (depending on 5' or 3' measuring rule) to be dominant. However, the dominant frame identified is frame 4/2 which is a first hint for Dicer-2 entry point not being at the 5' extremity of DCV.

Dicer-2 precise entry point is located in DCV 5' region

We then wanted to precisely identify where is this entry point of Dicer-2 on DCV dsRNA located. Because we are looking at Dicer-2 signature of action, we first gathered 21nt long reads mapping on genome and antigenome in groups according to the 2nt 3' overhang rule. For instance, reads mapping on the genome in frame 4 were gathered together with reads mapping on the antigenome in frame 2. Reads were split in 21 groups according to this rule. For the sake of simplicity, on the graphics the group containing reads in frame 4/2 for instance was named “frame 4”. This sorting of the reads allowed us to look at the cumulative frequency of DCV coverage for all the frames independently.

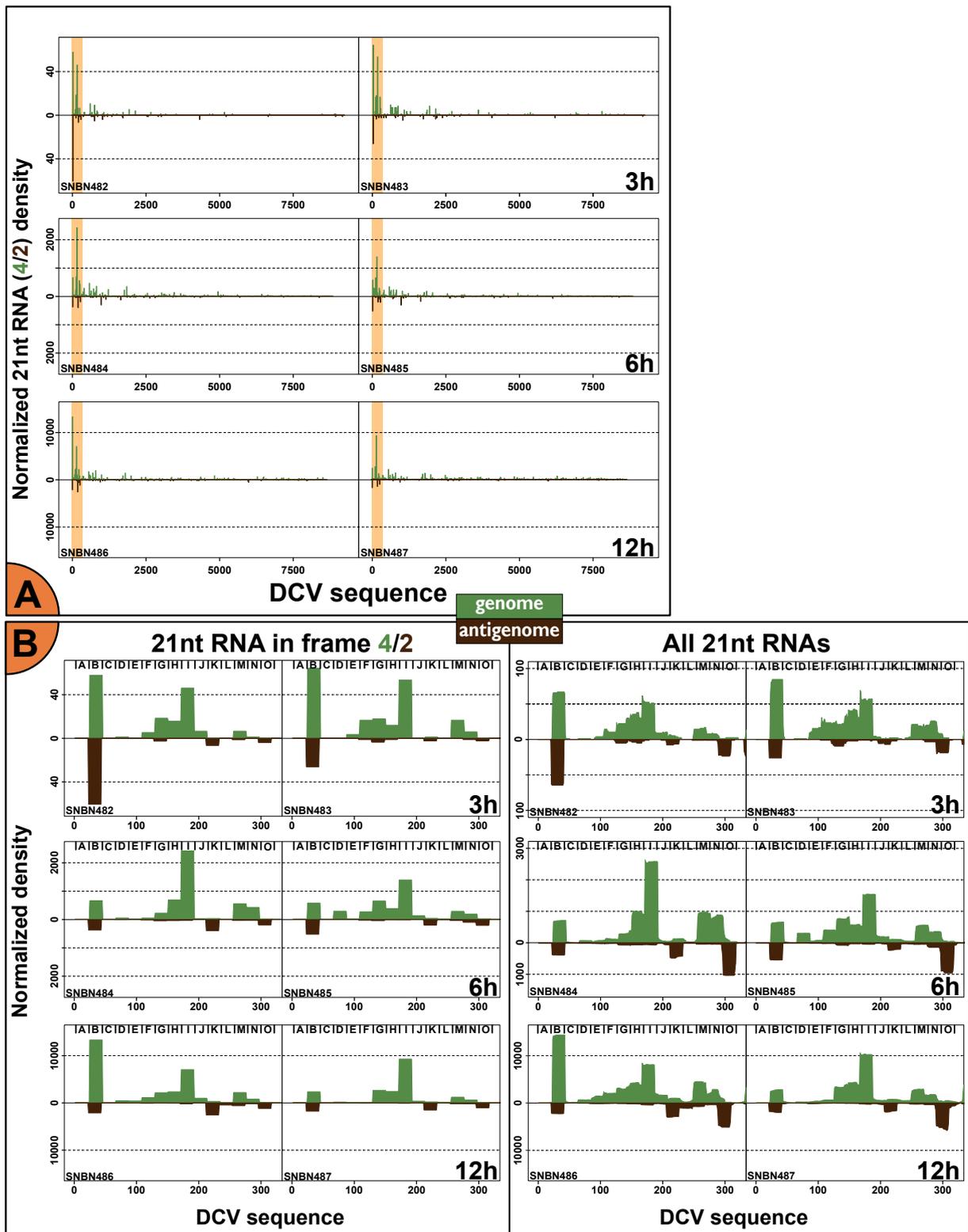


Figure 8: Characterization of the siRNAs mapping in the 5' region of DCV.

A) Normalized DCV coverage by 21nt long reads in frame 4/2. **B)** Zoom on the 1-318nt region of the **Figures 8A** (left) and **5B** (right). Portions A to O on these graphics are arbitrary determined to make the interpretation easier. They correspond to the theoretical reads in frame 4/2.

The frames 1-3 and 8-21 are basically following the trends observed when looking at all 21nt long reads together (*Figure 5C & 7C*). This suggests an even distribution of the corresponding reads on DCV sequence and no region-specific enrichment of these frames. However, the frames 4-6 do not behave the same way and stick out of the theoretical diagonal at the 3 time points of DCV infection. From these graphics, it is clear that the enrichment for frame 4 is strongly biased toward the 5' region of the virus. Interestingly, frame 4 is the first one to be enriched but is quickly followed by frames 5 and 6. This point is in accordance with the previous frame enrichment analysis and will be further discussed. Finally, frame 7 is still a bit above the diagonal but clearly shows a trend of going back to even coverage. Thus, all the results obtained so far highlight a precise entry point of Dicer-2 on DCV dsRNA located in its very 5' region and in frame 4/2.

We then looked at the distribution of reads in frame 4/2 on DCV sequence. Without a surprise an enrichment of reads in the very 5' region of DCV could be observed (*Figure 8A*). We zoomed in on this highly covered region (~1-300nt) in an attempt to find the precise entry point of Dicer-2 (*Figure 8B*). To make the interpretation easier, the graphics were split in chunks of 21nt long (annotated from A to O) corresponding to the theoretical frame 4/2 phased reads (*Figure 8B – left*). In order to grasp the importance of this frame in the DCV 5' region, the same zoom was done on the *Figure 5B* representing the total 21nt long reads distribution (*Figure 8B – right*). Thus, a first general remark concerning these graphics lies in the fact that the vast majority of 21nt long reads mapping to this part of DCV sequence are in frame 4/2.

Regardless of the library we are looking at, no 21nt long reads are mapping at the 5' extremity of the virus (position A). However, at this point, we need to step back and think about Dicer-2 mechanism. In particular, we know that Dicer-2 is producing the peculiar siRNA duplex signature with 2nt 3' overhang but we still do not know (1) what are the 5' extremities that can be produced by the virus during the replication cycle and (2) what are the lengths of the first products made from a blunt dsRNA template. As depicted in *Figure 9A*, we could easily imagine Dicer-2 producing either 19 or 23nt long reads when processing blunt dsRNA from its extremity. However, in our sequencing data, no enrichment for a specific small RNA read mapping at the 5' extremity of DCV sequence could be observed regardless of the size of the reads we looked at.

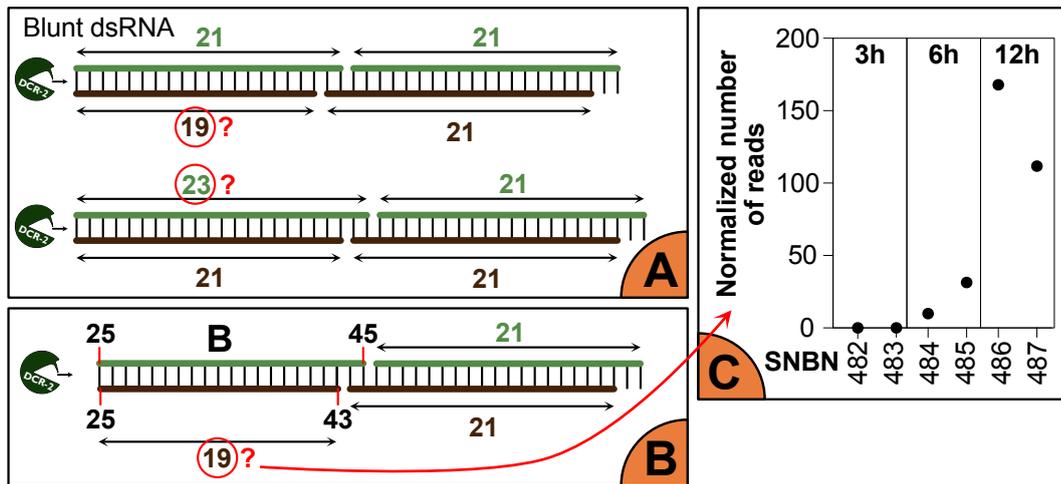
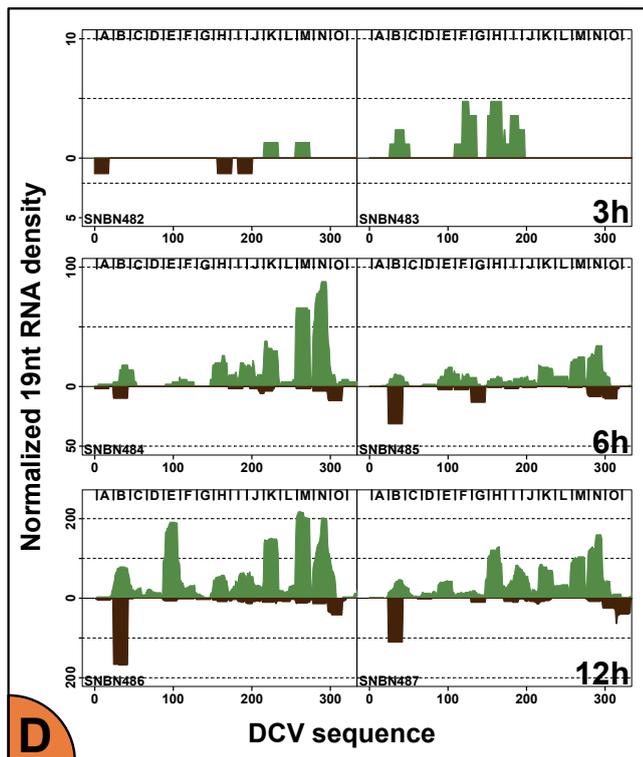
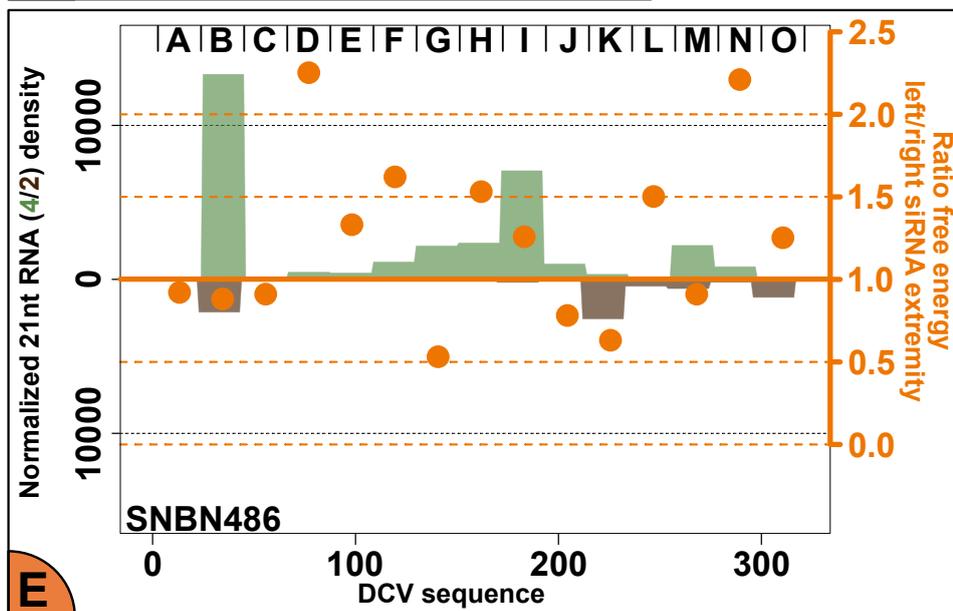


Figure 9: Further characterization of RNA reads mapping in the 5' region of DCV.



A and B) Schematic explanation of the hypothetical Dicer-2 products from a blunt dsRNA. **C)** Normalized number of the specific 19nt long read mapping at position 25-43 on the antigenome of DCV. **D)** Normalized DCV coverage by 19nt long reads. A zoom in region 1-318 is presented. **E)** Example of library SNBN486. Comparison of the predicted siRNA in frame 4/2 strand loading (right y scale) and empirical data (left y scale). A thermodynamic ratio above (resp. below) 1 favors the loading on Ago2 of sense (resp. antisense) strand of the siRNA duplex.



The first siRNA signature that can be observed in frame 4/2 lies at position **B** with approximately equal number of reads mapping on the genome and antigenome, suggesting a Dicer-2 product from a dsRNA. We asked ourselves whether this siRNA could be the entry point of Dicer-2 on the viral dsRNA. To investigate this question, we followed the same reasoning as before: if Dicer-2 enters on a blunt dsRNA at position B (nucleotide 25), we might be able to see a 19nt long read mapping on the antigenome (*Figure 9B*). Interestingly, this very specific read was observed in the 6h and 12h pi libraries in increasing numbers (*Figure 9C*). Moreover, this 19nt long read is the only accumulating on the antigenome in close proximity to position B suggesting that it is a real product and not just some kind of degradation product (*Figure 9D*). The low abundance of this read in the 6h and 12h pi condition as well as its absence in the 3h pi condition can be explained by the fact that 19nt long reads are not loaded on Ago2 and thus, not protected from degradation. The coexistence of the 19 and 21nt long reads mapping on the antigenome at position **B** is puzzling and will be further discussed but suggests that the siRNA at position **B** might be the entry point of Dicer-2 on a blunt dsRNA of DCV.

Of note, no genomic read of 23nt long mapping from 25 to 47 was identified. This information suggests that if Dicer-2 indeed enters on a blunt dsRNA, it is the 5' measuring rule that is used by Dicer-2 to generate precise 21nt long siRNAs. This information is in accordance with the previously published study highlighting the role of Dicer-2 PAZ-platform domain for the recognition of the 5'P moiety to generate high-fidelity siRNAs (Kandasamy and Fukunaga, 2016).

The siRNA signature at position B is directly followed by a surprising gap in the sequence coverage regardless of the libraries and frames (*Figure 8B*). Of note, this gap cannot be explained by mismatches within the reference DCV sequence that was checked notably by the 5' RACE experiment performed in Chapter II. Then, an increasing number of reads mapping on the genome can then be observed from position **D** to **I**. Oddly, all these reads on the genome in frame 4 nearly never have their antigenomic counterpart on frame 2 (*Figure 8B – left*). As a general remark, only a few antigenomic mapping reads were identified in the 5' region of DCV.

In an attempt to explain this peculiar asymmetric distribution of the reads, thermodynamic analysis of the free energy of frame 4/2 siRNA extremities was calculated.

		Frame 4/2 siRNAs										Ratio											
		5' 3'										Left/Right											
A	Genome	A	A	U	A	U	C	G	U	G	U	G	U	A	C	A	U	A	U	A	A	A	U
	Antigenome	A	U	U	A	U	A	G	C	A	C	A	C	A	U	G	U	A	U	A	U	U	
		-5,98										-5,49					0,92						
B	Genome	U	A	U	G	U	A	C	A	C	A	C	G	G	C	U	U	U	U	A	G	G	U
	Antigenome	U	A	U	A	C	A	U	G	U	G	U	G	C	C	G	A	A	A	U	C		
		-7,48										-6,57					0,88						
C	Genome	C	A	G	A	A	U	A	U	U	G	U	U	U	U	C	A	A	U	G	U	U	G
	Antigenome	C	A	U	C	U	U	A	U	A	A	C	A	A	A	A	G	U	U	A	C	A	
		-7,16										-6,48					0,91						
D	Genome	A	C	U	A	A	A	A	U	U	C	C	A	U	U	G	G	A	A	C	C		
	Antigenome	A	C	U	A	A	A	A	U	U	U	C	C	A	U	U	G	G	A	A	C	C	
		-3,99										-8,98					2,25						
E	Genome	A	A	U	A	A	U	A	C	G	A	A	A	U	G	C	C	A	A	A	G		
	Antigenome	A	A	U	A	A	U	A	C	G	A	A	A	U	G	C	C	A	A	A	G		
		-5,16										-6,84					1,33						
F	Genome	U	A	A	C	A	A	C	U	A	C	C	A	U	A	A	A	C	G	C	C	U	
	Antigenome	U	A	A	C	A	A	C	U	A	C	C	A	U	A	A	A	C	G	C	C	U	
		-8,19										-13,26					1,62						
G	Genome	G	A	C	G	C	C	A	G	G	G	A	U	U	A	A	C	A	C	U	U		
	Antigenome	G	A	C	G	C	C	A	G	G	G	A	U	U	A	A	C	A	C	U	U		
		-12,98										-6,92					0,53						
H	Genome	A	A	U	A	A	A	U	A	A	G	A	C	U	A	U	G	C	A	A	C	A	
	Antigenome	A	A	U	A	A	A	U	A	A	G	A	C	U	A	U	G	C	A	A	C	A	
		-4,99										-7,62					1,53						
I	Genome	A	A	A	G	U	A	A	C	U	A	C	C	A	U	U	C	C	A	U			
	Antigenome	A	A	A	G	U	A	A	C	U	A	C	C	A	U	U	C	C	A	U			
		-7,29										-9,22					1,26						
J	Genome	A	A	G	A	A	A	A	A	C	C	U	C	A	C	C	A	A	A	A			
	Antigenome	A	A	G	A	A	A	A	A	C	C	U	C	A	C	C	A	A	A	A			
		-6,54										-5,13					0,78						
K	Genome	A	G	U	C	U	U	C	U	A	U	G	A	G	A	U	U	U	A	U	A	G	A
	Antigenome	A	G	U	C	U	U	C	U	A	U	G	A	G	A	U	U	U	A	U	A		
		-8,14										-5,16					0,63						
L	Genome	C	U	U	A	A	U	A	C	G	G	A	A	U	A	A	C	A	A	A	A	G	
	Antigenome	C	U	U	A	A	U	A	C	G	G	A	A	U	A	A	C	A	A	A	A	G	
		-4,56										-6,84					1,50						
M	Genome	U	U	A	C	A	A	C	U	A	C	C	G	G	A	A	G	C	A	A	A	C	
	Antigenome	U	U	A	C	A	A	C	U	A	C	C	G	G	A	A	G	C	A	A	A	C	
		-7,99										-7,31					0,91						
N	Genome	U	U	A	U	A	A	A	A	A	C	A	A	U	U	A	C	A	C	G	G	U	A
	Antigenome	U	U	A	U	A	A	A	A	A	C	A	A	U	U	A	C	A	C	G	G	U	A
		-4,89										-10,81					2,21						
O	Genome	A	U	U	A	G	U	U	U	C	U	A	A	U	G	U	A	G	A	G	U	U	
	Antigenome	A	U	U	A	G	U	U	U	C	U	A	A	U	G	U	A	G	A	G	U	U	
		-6,59										-8,27					1,25						

Table 2: Summary table of the calculation of free energy at extremities of siRNA duplexes in frame 4/2.

Calculation of the free energy at both extremities of each siRNA duplex in frame 4/2 was done following the nearest neighbor parameter database recommendations (Turner and Mathews, 2010). The orange nucleotides were taken into account to determine this value expressed in kcal/mol. The lower a value is, the more stable this extremity of the duplex is. Thus, a ratio of left/right value above 1 (resp. below) will favor the sense (resp. antisense) strand to be loaded on Ago2.

It was shown that the selection of the guide RNA to be loaded on Ago2 is influenced by the thermodynamic stability of the siRNA duplex extremities (Tomari et al., 2004). Thus, the siRNA strand with the 5' extremity showing the lower stability (higher free energy) will be preferentially loaded on Ago2 and be protected from degradation (guide strand). The passenger strand will be quickly degraded and might be hard to detect by small RNA HTS. Thus, we wanted to investigate whether this asymmetrical distribution of the reads in the 5' region of DCV could be explained by the thermal stability of the siRNA duplexes generated. For each siRNA duplex in frame 4/2, the free energy of both extremities was calculated following the nearest neighbor parameter database recommendations (Turner and Mathews, 2010 – *Table 2*). The ratio between the free energy of genomic and antigenomic strands gave us a value predicting which strand should be preferentially loaded on Ago2. Thus, a ratio above 1 (resp. below) will favor the sense (resp. antisense) strand to be loaded. However, no clear-cut correlation between the theoretical and the empirical distribution of the reads could be observed (*Figure 9E*). The thermodynamic stability of siRNA extremities cannot explain the asymmetric distribution of frame 4/2 reads.

The small RNA HTS performed in DCV infected S2 cells allowed us to establish a kinetic of the apparition of virus derived siRNAs. Interestingly, an siRNA signature can be detected as early as 3h pi depicting a very fast and efficient sensing and processing of the viral dsRNA by Dicer-2. All the bioinformatic analyses done point toward a precise entry point of Dicer-2 on the viral dsRNA in the 5' region of DCV. More precisely, reads in frame 4/2 are highly enriched in the 1-318nt region suggesting an entry of Dicer-2 on a viral dsRNA which extremity does not correspond to the virus 5' extremity. The first siRNA signature identified is interesting because in addition to the $21_{\text{genome}}/21_{\text{antigenome}}$ canonical duplex, an enrichment of $21_{\text{genome}}/19_{\text{antigenome}}$ signature could depict an entry of Dicer-2 on a blunt dsRNA. In conclusion, this kinetic analysis of the apparition of siRNAs in DCV infected S2 cells allowed us to propose a precise entry point of Dicer-2 on a blunt dsRNA extremity located in the 5' region of the virus but different from its 5' extremity.

Kinetics of siRNAs apparition in S2 cells during CrPV infection

At this point, we wanted to test whether the data obtained with DCV could be similar to the ones obtained with a closely related dicistrovirus: CrPV. Thus, kinetics of siRNAs apparition in S2 cells during CrPV infection was investigated using the same small RNA HTS approach.

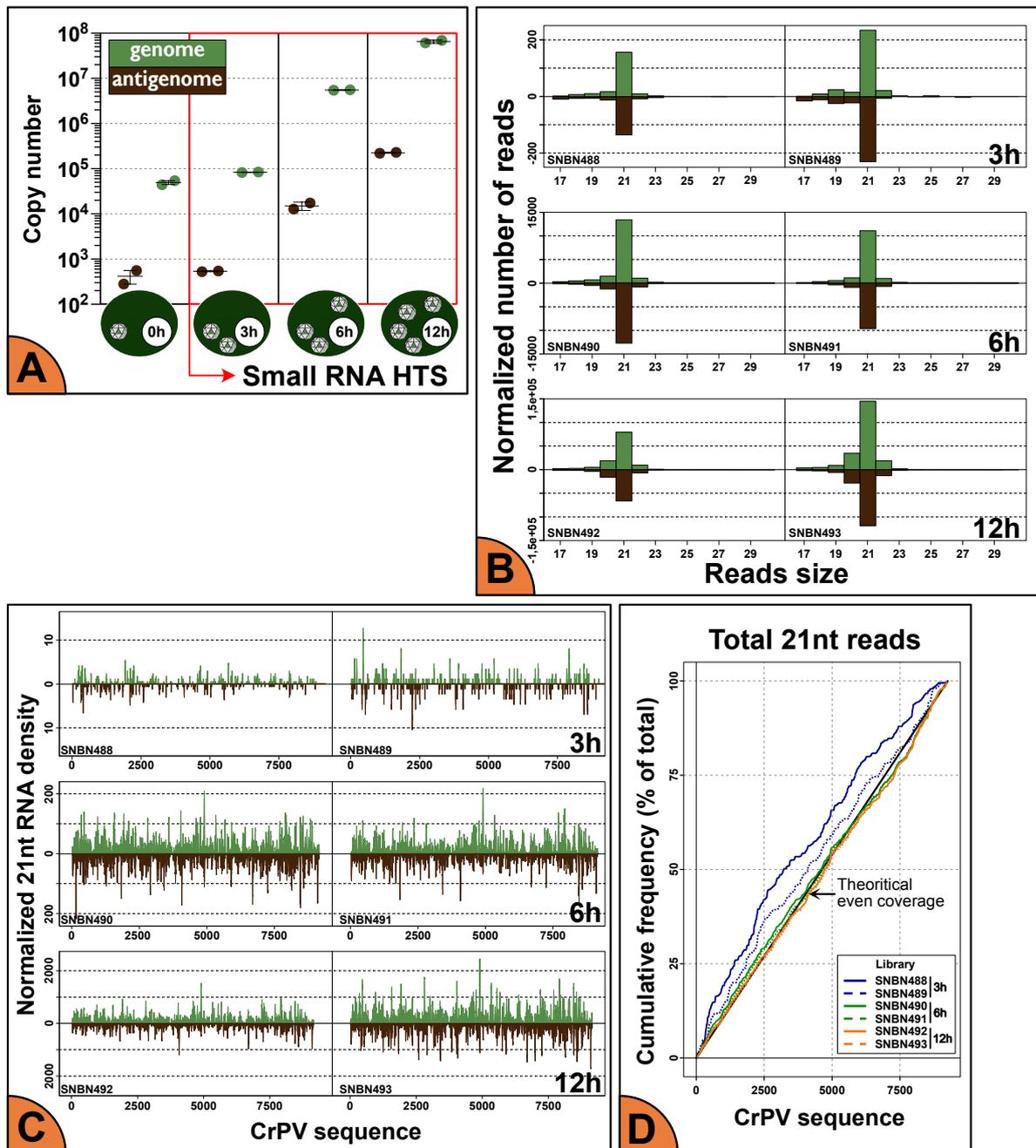


Figure 10: Kinetics of the apparition of CrPV mapping 21nt long reads.

A) Strand-specific RT-qPCR designed to detect separately genome and antigenome CrPV strands. The y axis is the absolute strand copy number. The error bars represent the standard deviation between the biological replicates ($n=1$). Time points 3 -12h were used to perform small RNA HTS. **B)** Normalized CrPV mapping reads size distribution. One mismatch was allowed for the alignment. **C)** Normalized CrPV coverage by 21nt long reads. **D)** Cumulative frequency of DCV coverage by 21nt long reads. The black line represents what would be a perfectly even coverage of the sequence.

Figure 11: Detection of phased siRNA duplexes.

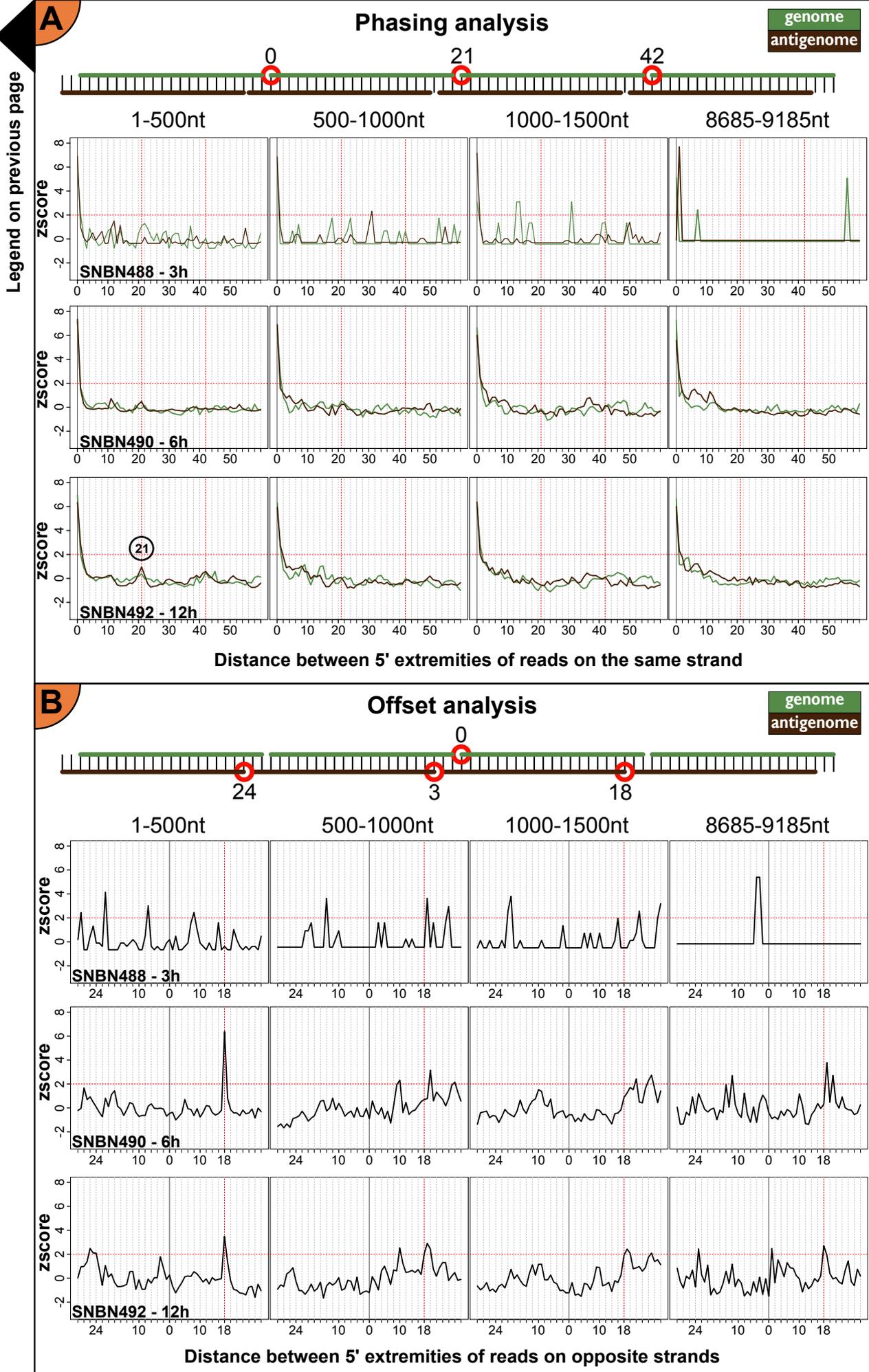
A) Phasing analysis of 3 libraries representative of their corresponding time point. **B)** Offset analysis of 3 libraries representative of their corresponding time point.

This experiment was performed nearly identically and at the same time than the DCV one with the only noticeable difference being the MOI of infection used (0.05), justified by the lytic effect of the virus on cells (*Figure 3A*). Infection status was checked at every time point by strand-specific RT-qPCR and showed an occurring productive infection (*Figure 10A*). The libraries from the 3, 6 and 12h pi time points were constructed following the same double index modification of Illumina protocol. Trimming and the second demultiplexing of the data was done by the same bioinformatic pipeline (*Figure 4 – Table 1*). This time, we analyzed more specifically libraries SNBN488-493 that were constructed with RNAs extracted from CrPV infected S2 cells. All the following figures were assembled as described above for DCV. Thus, we will directly emphasize the result of each figure without going in the detail of the principle behind.

To begin with, the analysis of CrPV mapping reads (1 mismatch allowed) size distributions shows a symmetrical peak at 21 (*Figure 10B*). The number of these reads is raising with time but are significantly lower than the ones observed in the DCV analysis. This reflects the lower MOI of infection we used. No obvious enrichment of 21nt long reads could be observed on the CrPV sequence, regardless of the time point considered (*Figure 10C e³ D*). No clear phasing of the 21nt long reads could be observed (*Figure 11A*). However, a siRNA duplex signature could be detected when searching for offset in the first 500nt of the 6 and 12h pi time points (peak at 18 – *Figure 11B*). In the same window and at 12h pi, peaks at 3 and 24 also suggest phased siRNA duplexes. This first part of the analysis highlights the fact that it is more complicated to detect Dicer-2 signature in these CrPV libraries. However, we can still detect it in the very 5' region of CrPV suggesting that Dicer-2 is able to access and dice the corresponding dsRNA.

Frame 3 enrichment was detected when looking at the CrPV antigenomic mapping reads (*Figure 12A*). However, no consistent frame enrichment was observed in the genomic strand mapping reads. We still decided to plot the cumulative frequency coverage of CrPV for all the frames separately according to the 2nt 3'overhang rule. Thus, antigenomic reads in frame 3 are here gathered in the frame 5/3. Interestingly, the cumulative coverage graphics show an enrichment for frame 5/3 reads in the 5' region at 6 and 12h pi (*Figure 12B*). This accumulation is directly followed by an enrichment of reads in frame 6/4. Again, this result suggests a precise entry point of Dicer-2 in the 5' region of CrPV dsRNA different from the 5' extremity and in frame 5/3. The distribution

Figure II



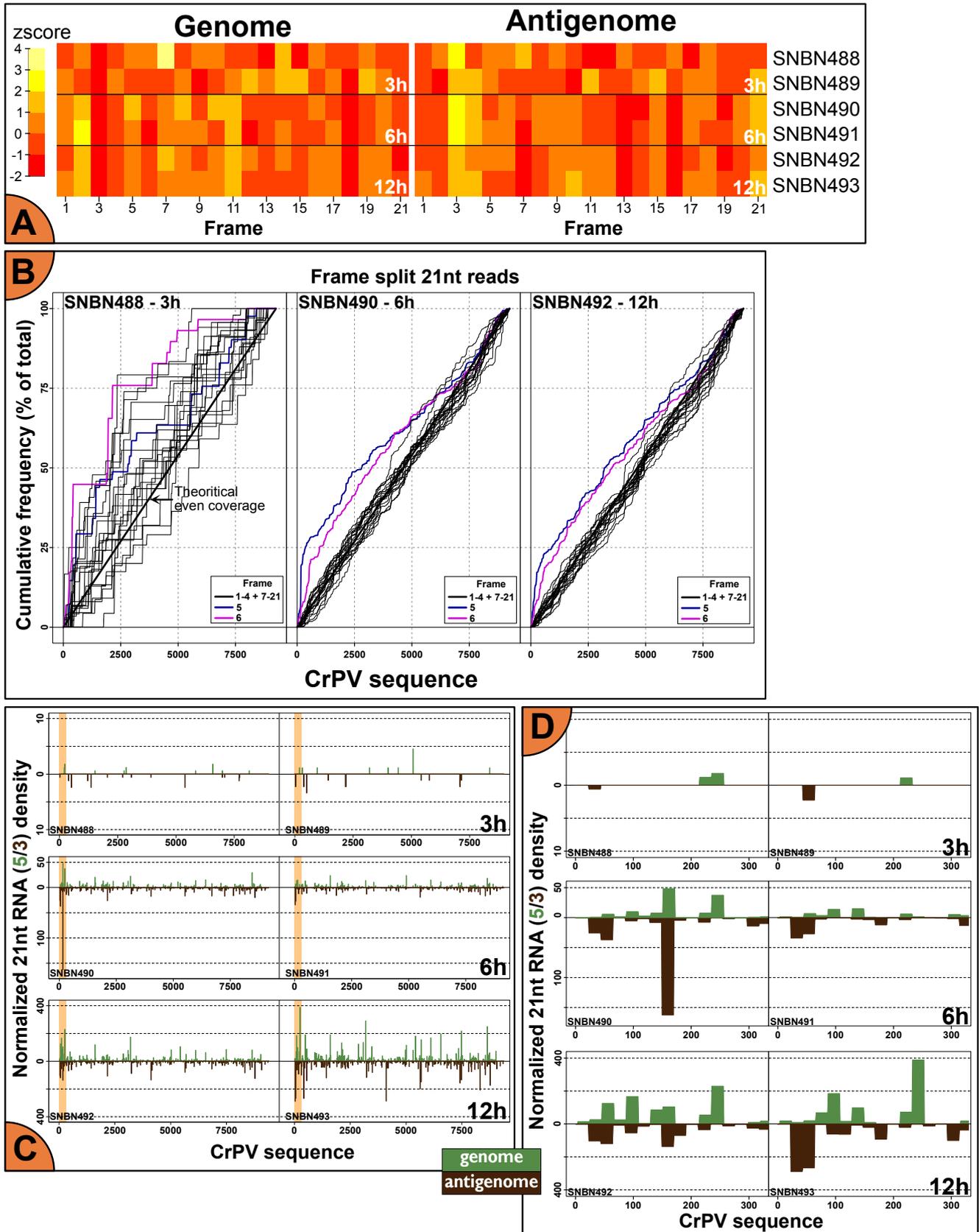


Figure 12: Frame-based analyses highlights a precise entry point of Dicer-2 in the 5' region of CrPV.

A) Frame enrichment calculated individually for CrPV genome and antigenome mapping 21nt long reads. **B)** Cumulative frequency of CrPV coverage by independent frames. **C)** Normalized CrPV coverage by 21nt long reads in frame 5/3. The three showed libraries are representative of their respective time point. The black line represents what would be a perfectly even coverage of the sequence. **D)** Zoom on the 1-318nt region of the **Figures 24C**.

of the reads in frame 5/3 on CrPV sequence shows an enrichment in the 5' region (*Figure 12C*). However, when zooming in on this region (1-318), it was hard to detect a consistent pattern of mapping reads between time points and biological duplicates (*Figure 12D*). Thus, it was decided not to further investigate the mechanism by which these reads could be created.

In conclusion, the small RNA HTS performed in CrPV infected S2 cells allowed us once again to establish a kinetic of the apparition of virus derived siRNAs. Probably due to the lower number of reads, the overall analysis delivered results that were less clear cut than the DCV analysis. In addition, a major difference between these viruses lies in their VSRs that have different modes of action and probably shape the antiviral response signature. Thus, the dsRNA coating mechanism of DCV 1A might be able to block Dicer-2 processing of the entire molecule, which would result in an accumulation of reads at its entry point. On the contrary, CrPV 1A inhibits Ago2 action but leaves the intermediate of replication devoid of protections, hence a complete processing action of Dicer-2.

However, similarly to DCV, the results obtained with CrPV suggest a precise entry point of Dicer-2 on viral dsRNA in the 5' region but not at the 5' extremity. These two complementary experiments highlight what could be a weak point in the dicistroviruses defense. Before going in a detailed characterization of this region, we wanted to make sure that the results obtained are not due to a bias in the model we used. Indeed, S2 cells cannot compare with the *in vivo* complexity of a pluricellular organism. That is why it was decided to perform a final small RNA HTS but this time in DCV infected adult flies.

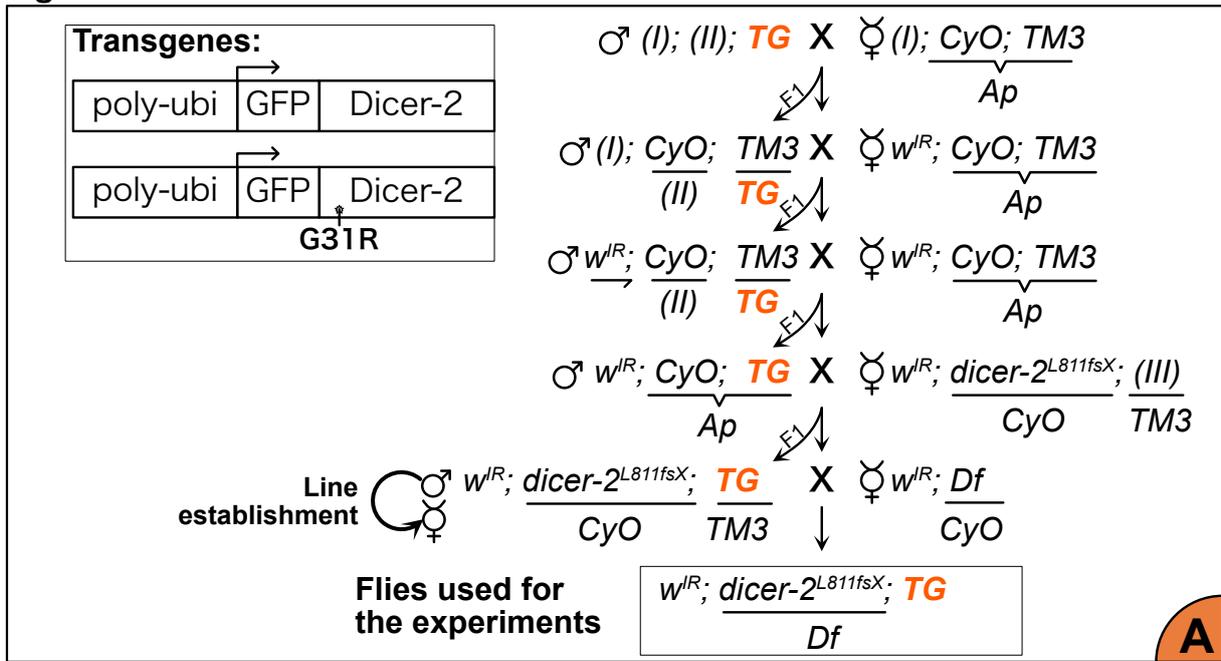
II. Small RNA high throughput sequencing (HTS) of DCV infected adult flies

a. Preliminary work

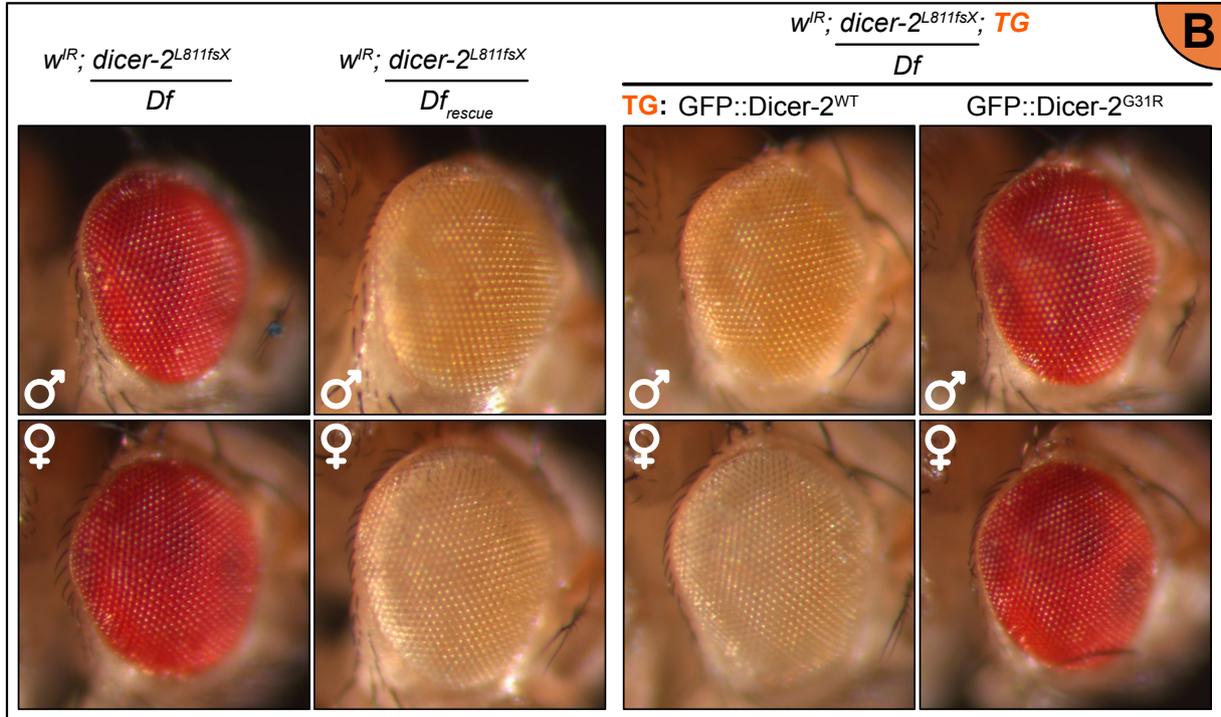
DCV infection of adult flies using different genetic background

One of the advantages of using *Drosophila melanogaster* as a model for this study is the high versatility of genetic tools that have been developed for it. This allowed us to work with 3 different genotypes of flies regarding *dicer-2* gene:

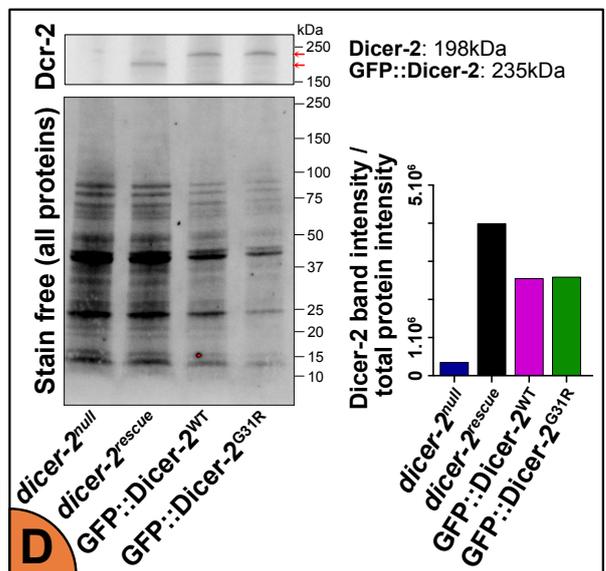
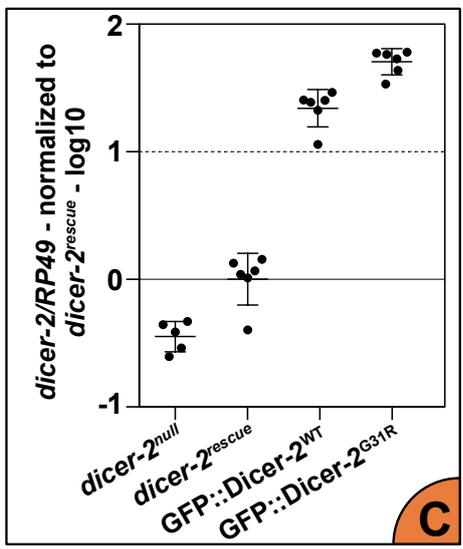
Figure 13



A



Legend on next page



Genotype	Simplified name
$w^{IR}; \text{dicer-2}^{L811fsX}/Df$	dicer-2^{null}
$w^{IR}; \text{dicer-2}^{L811fsX}/Df; \text{GFP::Dicer-2}^{WT}$	GFP::Dicer-2^{WT}
$w^{IR}; \text{dicer-2}^{L811fsX}/Df; \text{GFP::Dicer-2}^{G51R}$	$\text{GFP::Dicer-2}^{G51R}$

Briefly, w^{IR} refers to the transgene used to monitor RNAi efficiency, a “white inverted repeat” previously described by Lee et al., 2004. This transgene, when transcribed and spliced, will fold in a perfect hairpin dsRNA structure targeted by Dicer-2. The resulting siRNAs are complementary to *white* mRNA and will, with the help of the siRNA machinery, downregulate the expression of this gene. White gene encodes for an ABC transporter essential for the entry of red pigment precursors in cell (Ewart and Howells, 1998). The w^{IR} transgene is under the control of a GMR promoter restricting its expression to the eyes’ cells. As a result, wild type flies will have white eyes, whereas mutants for the siRNA pathway will have red eyes. $\text{dicer-2}^{L811fsX}$ mutation is a frameshift at Leucine 811 resulting in a premature STOP codon (Lee et al., 2004). Moreover, the facing chromosome is a deficiency chromosome (**Df(2R)BSC45**) bearing a *dicer-2* deletion (Kemp et al., 2013). Thus, all the flies used in this study are null mutant for endogenous *dicer-2* gene. These flies have been complemented or not with transgenes inserted at the same genomic position on the third chromosome to avoid any position effect. These constructs express different versions of Dicer-2 under the control of a poly-ubiquitin promoter and N-terminally fused to a GFP tag. Random insertion of GFP::Dicer-2 in drosophila genome was previously shown to rescue a dicer-2^{null} mutant (Girardi et al., 2015). Finally, **Dicer-2^{G51R}** contains a point mutation in the helicase domain which inhibits ATP binding and processive activity of Dicer-2 (Cenik et al., 2011; Lee et al., 2004). These flies were used to test whether Dicer-2 entry on its viral dsRNA substrate is dependent of its ATP activity. These flies were obtained by the crossing strategy presented in **Figure 13A**. As a control, flies rescued with a wild-type version of endogenous *dicer-2* recombined with the deficiency (**Df(2R)BSC45-dcr-2**) were used (dicer-2^{rescue} , Kemp et al., 2013).

Only flies expressing a wild-type version of Dicer-2 had white eyes while the two other genotypes presented red eyes (**Figure 13B**). Eye color quantification was done in Chapter III with an additional fly genotype (**Figure 35C**). This result is a first indicator that GFP::Dicer-2^{G51R} is not able to efficiently trigger the siRNA pathway. Dicer-2 expression

Sequencing lane	Injec.	Flies genotype	Seq ID	Total number of reads	Trimming surviving	Second demultiplexing surviving	21-22-23 miRNAs (1M)	Normalization factor
#1	DCV	dicer-2null	LCTD1	26527174	24898601	23185167	200296	0,42
		GFP::Dicer-2WT	LCTD2	29891872	27910383	26709027	188134	0,39
		GFP::Dicer-2G31R	LCTD3	26605916	25939925	24407357	71466	0,15
		GFP::Dicer-2F225G	LCTD4	35699189	32857285	31054310	482808	1,00
	VSV	dicer-2null	LCTD5	29343201	28908123	27307952	656284	1,37
		GFP::Dicer-2G31R	LCTD6	33809535	32330252	30536682	428543	0,89
		GFP::Dicer-2F225G	LCTD7	26536639	24717197	23793555	1011780	2,10
	TRIS	dicer-2null	LCTD8	27951880	27199873	25863092	550929	1,15
		GFP::Dicer-2WT	LCTD9	38821718	37836892	36294157	639885	1,33
		GFP::Dicer-2G31R	LCTD10	37533278	36536315	34427865	362401	0,75
		GFP::Dicer-2F225G	LCTD11	28422420	27712894	26541170	695267	1,45
	#2	DCV	dicer-2null	LCTD12	32620608	31921800	30361142	254644
GFP::Dicer-2WT			LCTD13	35490477	34283929	32042631	244093	0,54
GFP::Dicer-2G31R			LCTD14	25406980	24770133	23559054	147496	0,33
GFP::Dicer-2F225G			LCTD15	32342262	31605983	29873166	493428	1,09
VSV		dicer-2null	LCTD16	35393665	34688725	32443143	289032	0,64
		GFP::Dicer-2WT	LCTD17	37612710	36488553	34403302	205026	0,45
		GFP::Dicer-2F225G	LCTD18	32262226	30937943	29087737	812396	1,80
TRIS		dicer-2null	LCTD19	37975555	36696101	34479932	896684	1,98
		GFP::Dicer-2WT	LCTD20	40320156	39067069	36442465	729231	1,61
		GFP::Dicer-2G31R	LCTD21	41178470	38910186	36937206	570611	1,26
GFP::Dicer-2F225G	LCTD22	25937497	24922236	23017467	326849	0,72		

Mean#1 =	480708,5
Mean#2 =	451771,8

Table 3 Summary table of the small RNA HTS performed on libraries constructed from DCV,VSV or TRIS injected adult flies extracted RNAs.

Experimental duplicates were split between two sequencing lanes.

Total number of reads: the number of all small RNA reads in the fastq file delivered by the sequencing platform. This number is obtained after demultiplexing all the reads from the sequencing lane according to the 3' adaptor barcode. **Trimming surviving:** all the reads that were not identified as primers dimers or 2S ribosomal RNA. **Second demultiplexing surviving:** all the reads sequenced with the 5' adaptor barcode expected from this specific library reads. **21-22-23 miRNAs (1M):** alignment of the surviving reads on all miRNA hairpin sequences was done with 1 mismatch allowed. The miRNA identified reads of 21-23nt long were used for our normalization strategy (see Materials & Methods). **Normalization factor:** factor applied to each individual library of the sequencing lane in order to make comparisons between them possible.

Figure 13: Characterization of flies used for small RNA HTS.

A) Crossing strategy used to generate the flies. Males containing the different *dicer-2* transgenes were provided by BestGene. Df refers to Df(2R)BSC45 described in Kemp et al, 2013. **B)** Eye color of the different fly genotypes according to their gender. **C)** Relative mRNA level of Dicer-2 in comparison to the house-keeping gene *RP49* measured by RT-qPCR (n=2, biological triplicates, all samples were normalized to the *dicer-2^{rescue}* condition, statistical test: one-way ANOVA). **D)** Western blot showing Dicer-2 protein level in the flies. Stain free exposition of the blot is used for Dicer-2 protein quantity normalization (n=1).

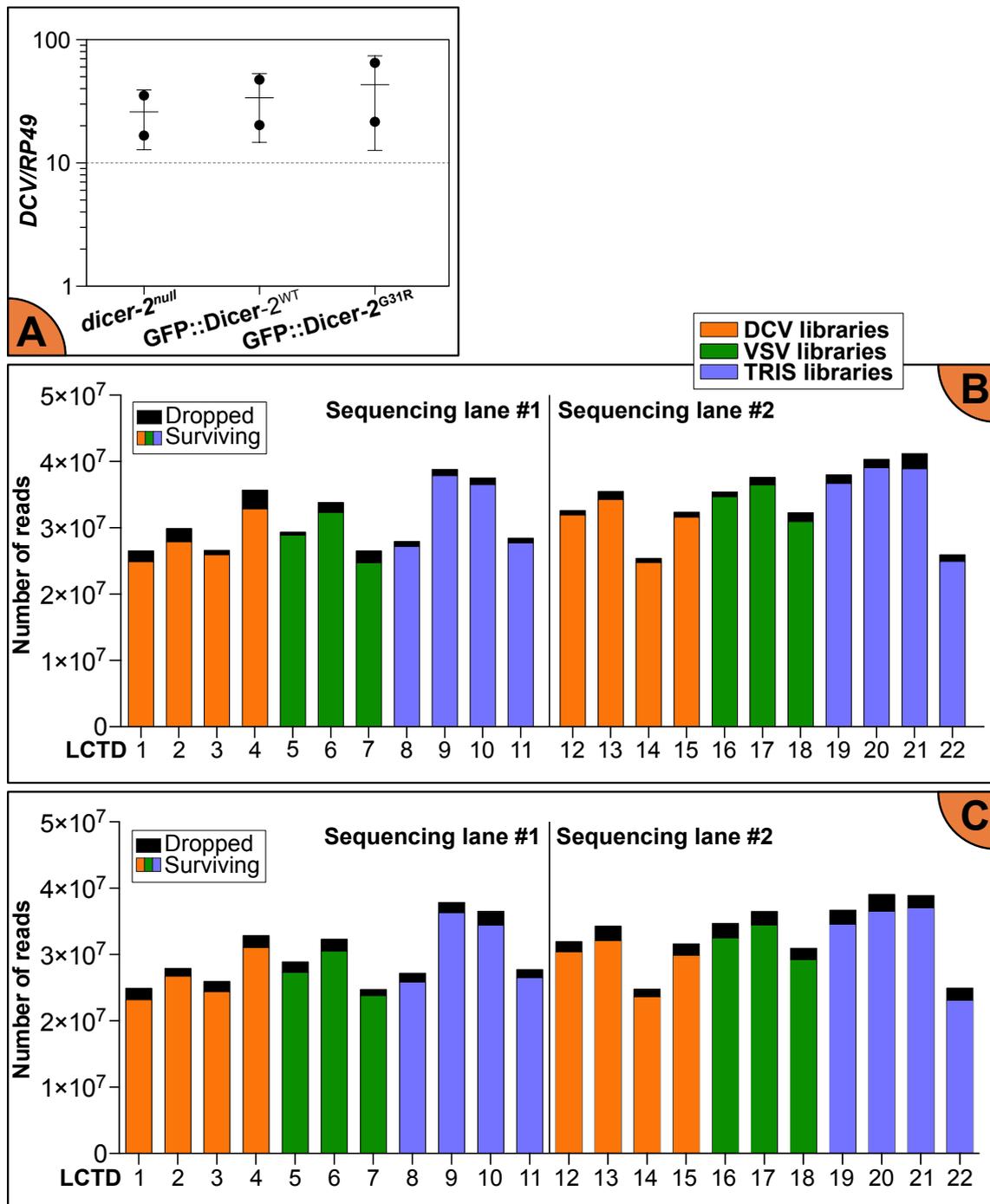


Figure 14: Preparation of libraries from DCV, VSV or TRIS injected flies.

A) Relative level of DCV RNA in comparison to house-keeping gene *RP49* measured by RT-qPCR ($n=1$; experimental duplicate). **B) Trimming:** reads corresponding to adapters dimers or drosophila 2S ribosomal RNA were removed from the data. **C) Second demultiplexing:** reads with either mismatches in the 5' internal label or internal label from another library were removed from the data.

LCTDXX: library identification number

LCTDI-4 and I2-15: libraries constructed from **DCV** infected flies extracted RNAs.

LCTD5-7 and I6-18: libraries constructed from **VSV** infected flies extracted RNAs.

LCTD8-11 and I9-22: libraries constructed from **TRIS** infected flies extracted RNAs.

in all flies was checked by RT-qPCR and western blot (**Figure 13C e³ D**). As expected, *dicer-2* mRNA was observed at a very low level in the *dicer-2^{null}* flies in comparison to the other flies. We can notice that Dicer-2 complemented flies present a higher level of *dicer-2* mRNA than *dicer-2^{rescue}* flies. This can be explained by the presence of the polyubiquitin promoter which is stronger than the endogenous *dicer-2* promoter. This difference in term of mRNA is not representative of the quantity of Dicer-2 protein observed which is apparently the same between complemented and *dicer-2^{rescue}* flies (**Figure 13D**). Of note, this experiment was only performed once with all these genotypes together but several times with the individual genotypes. Still, it should be repeated to allow precise quantification of Dicer-2 protein quantity in each fly genotype. The size difference between endogenous and complemented Dicer-2 is due to the 27kDa GFP tag. Finally, the data obtained using the different flies can be compared in the following experiments and should not reflect a difference in Dicer-2 protein quantity.

In order to prepare samples for the small RNA HTS, flies' infection was performed in duplicate by direct injection of the viral particles in the hemolymph. This method allows us to control the amount of DCV injected in each fly (500 PFU) but bypasses the gut barrier that is an important layer of protection of the fly. The infection was carried on for 3 days in order to let multiple viral cycles occur and to increase the number of cells infected. The infection state of the flies was checked by RT-qPCR and showed no significant difference between the fly genotypes used (**Figure 14A**). This result is surprising and will be further discussed in Chapter III. Briefly, it reflects a DCV infection that reached a plateau in term of viral load. The RNAs extracted from 3 males and 3 females 3d pi were used to build small RNAs libraries (**Table 3**).

Trimming and second demultiplexing of the libraries

We used the RNAs extracted from DCV infected flies 3d pi to construct the small RNAs libraries (**Table 3**). The same double indexing method used for small RNA sequencing in cells was used to reduce sample bleeding. Importantly, a blocking primer was added during the hybridization step of the RT primer in order to reduce the ribosomal 2S contamination of the libraries (Wickersheim and Blumenstiel, 2013). The experimental duplicates were split in two sequencing lanes that also contained TRIS and Vesicular Stomatitis Virus injected flies' samples. Another genotype (*w^{IR}*; *dicer-2^{L811fsX}*/*Df*; GFP::*Dicer-2^{F225G}*) was also sequenced along with the others. All these other conditions will be analyzed later in Chapter III. However, libraries contained in the same sequencing

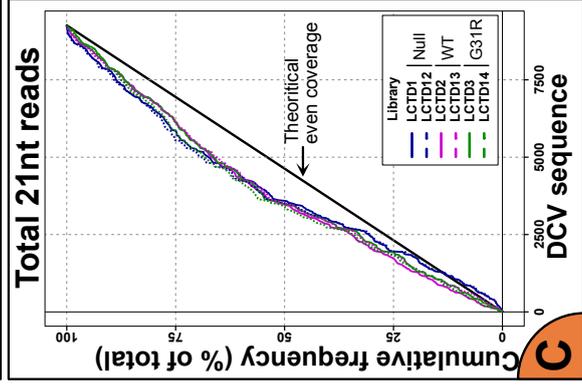
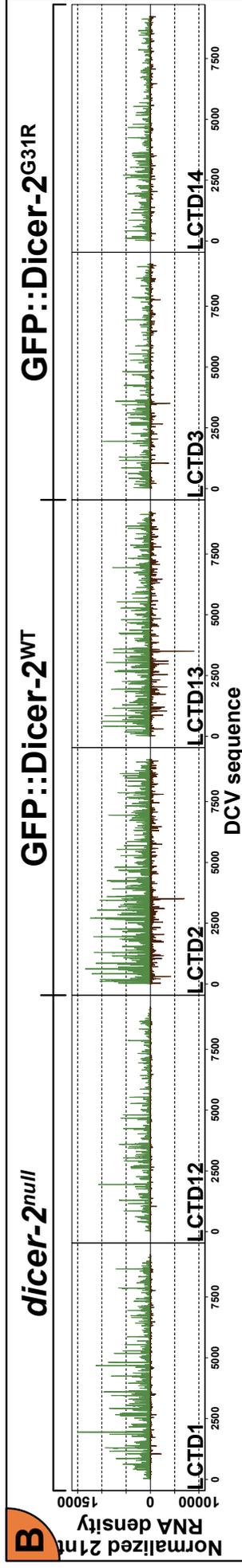
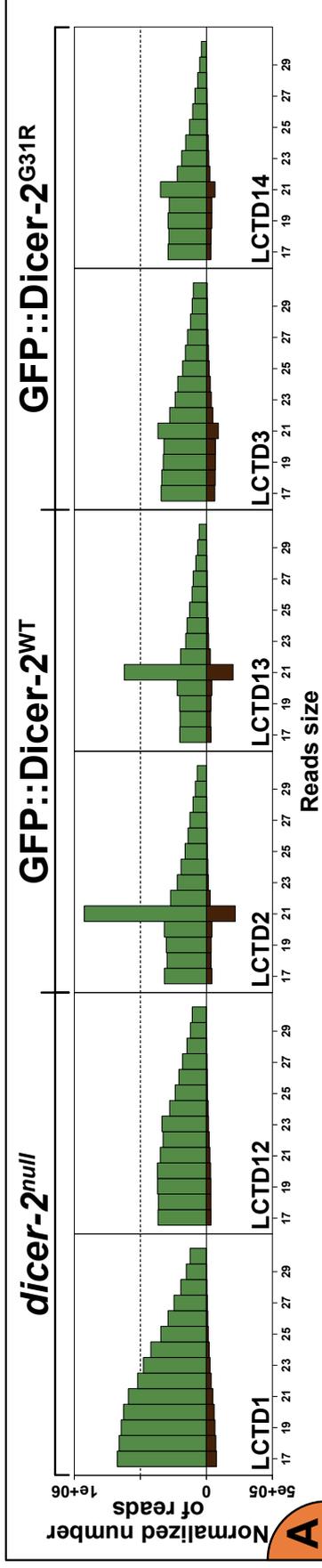


Figure 15: Small RNA sequencing of DCV infected flies 3d pi reflects a steady state of the infection.

A) Normalized DCV mapping reads size distribution. One mismatch was allowed for the alignment. **B)** Normalized DCV sequence coverage by 21nt long reads. **C)** Cumulative frequency of DCV coverage by 21nt long reads. The black line represents what would be a perfectly even coverage of the sequence.

Figure 16: Detection of phased siRNA duplexes.

A) Phasing analysis of 3 libraries representative of their corresponding genotypes. **B)** Offset analysis of 3 libraries representative of their corresponding genotypes.

lane were still treated as a whole regarding the trimming, second demultiplexing and normalization steps.

In comparison to the sequencing done in S2 cells, a high proportion of the reads survived the first trimming (*Figure 14B*). This emphasizes the importance of blocking the 2S ribosomal RNA during the preparation of the libraries. Second demultiplexing was done using the internal barcodes (*Figure 14C*) and the resulting datasets were used for further analysis.

b. Dicer-2 has a precise entry point on viral dsRNA of DCV 5' region

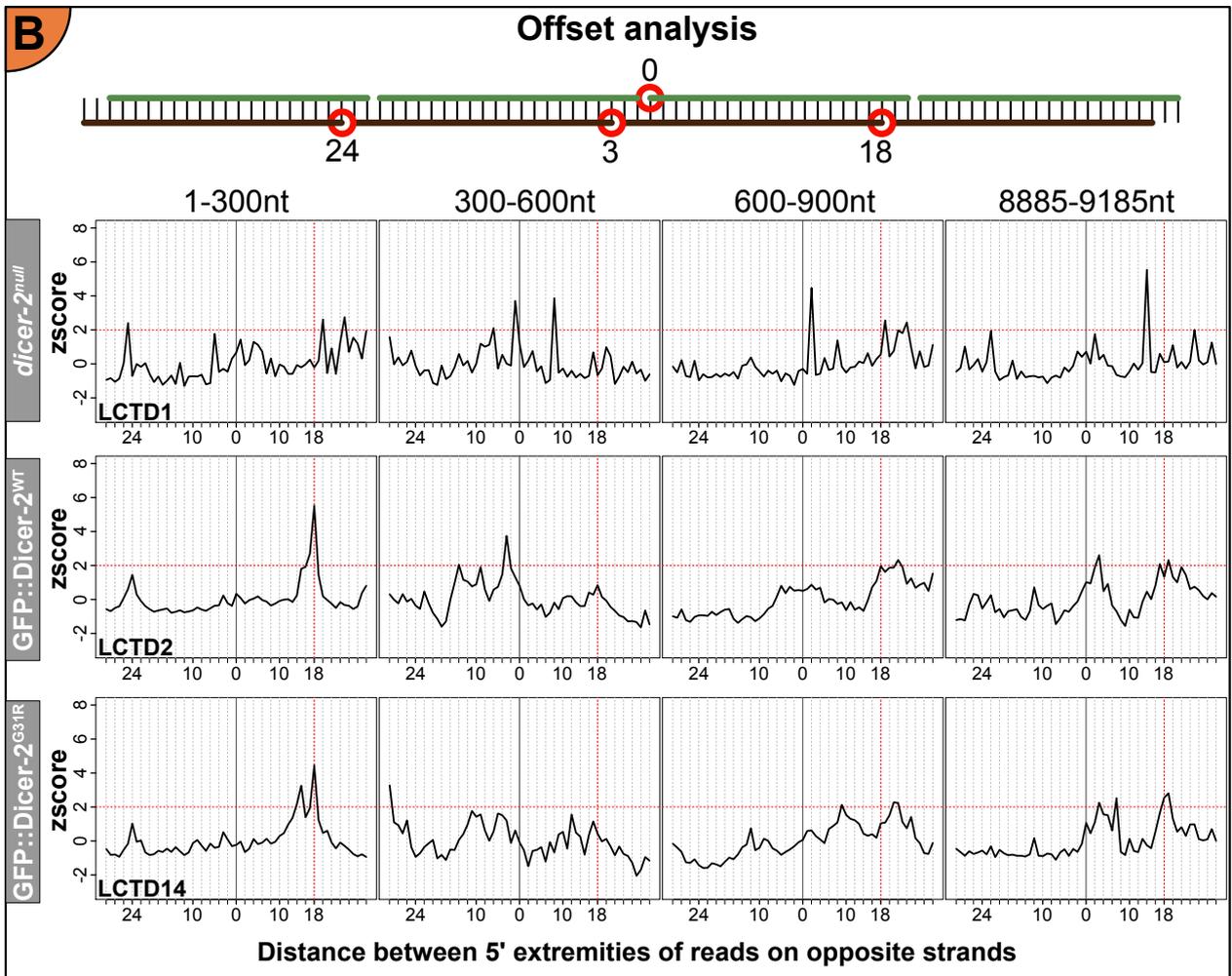
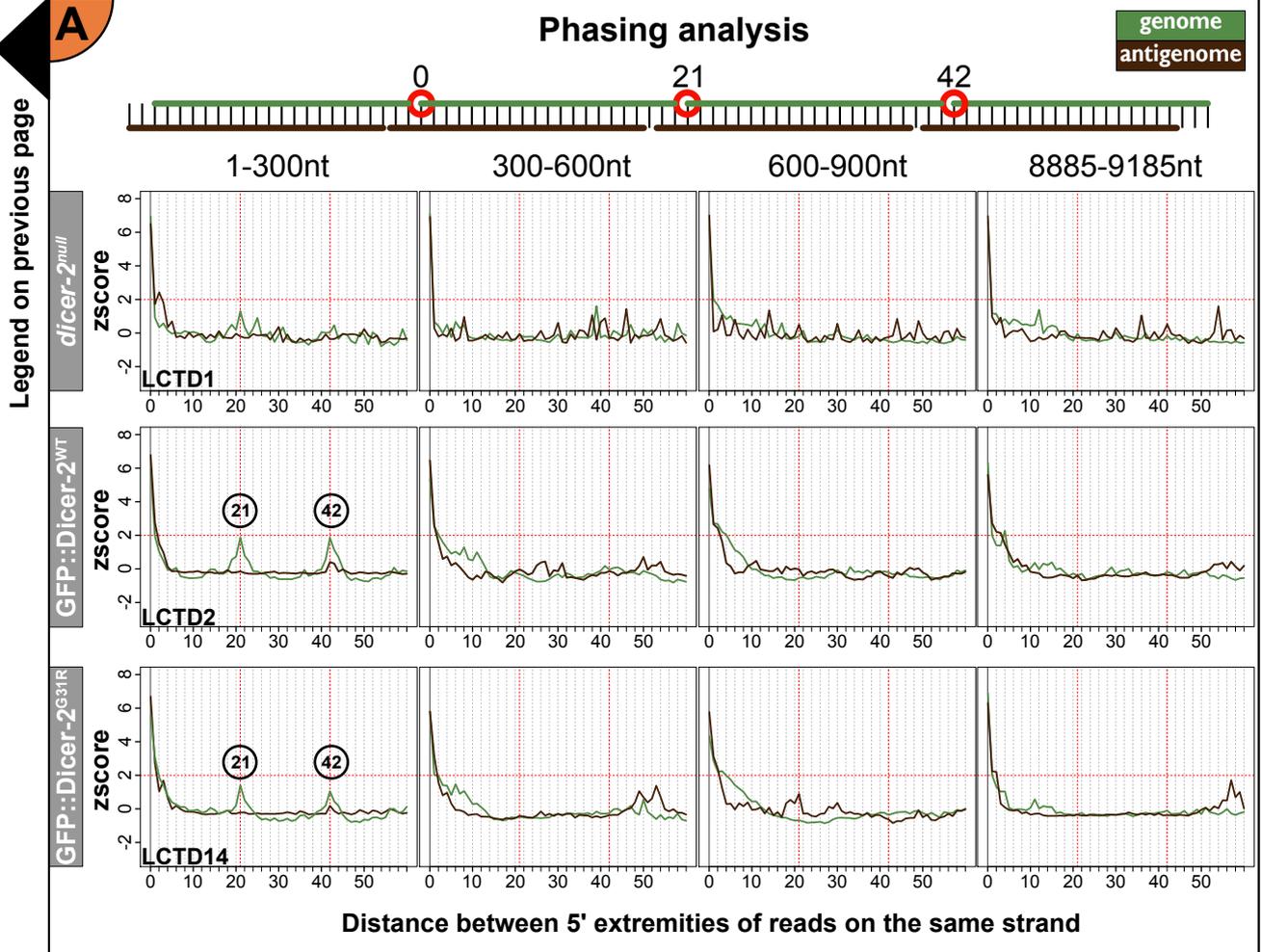
Dicer-2 produces a strong siRNA signature in the 5' region of DCV

The alignment of the reads on DCV sequence was performed allowing 1 mismatch. Plotting the size distribution of the reads shows us a clear and asymmetric peak at 21nt in GFP::Dicer-2^{WT} condition (*Figure 15A*). This peak can also be observed to a lesser extent in the GFP::Dicer-2^{G31R} condition but is absent in the *dicer-2^{null}* condition which shows its Dicer-2 dependency. This result is also in accordance with the fact that GFP::Dicer-2^{G31R} flies have red eyes and thus, probably have a less efficient siRNA pathway. Thus, Dicer-2 helicase-mediated ATP hydrolysis is required to produce a wild-type quantity of DCV derived siRNAs.

Importantly, a significant number of longer and shorter reads are also mapping on the genomic strand of DCV. These products are Dicer-2 independent (because also present in *dicer-2^{null}* condition) and were qualified as “degradation products”. These reads could also be observed in the S2 cells small RNA HTS but in a smaller amount (*Figure 5A*). Degradation products' origin will be discussed later but their presence in the analysis has to be reminded as they most likely interfered with the Dicer-2-dependent siRNA signature.

Distribution of all the 21nt long reads on DCV sequence shows no specific region enrichment regardless of the genotype we look at (*Figure 15B e³ C*). This was expected as the system must have reached steady-state for a long time after 3 days of infection. In comparison, in S2 cells, an even coverage of DCV sequence was already reached 12h pi. The even distribution of the 21nt long reads in *dicer-2^{null}* condition highlights how difficult it will be to discriminate degradation products from siRNA reads. Still, we searched for

Figure 16



Legend on previous page

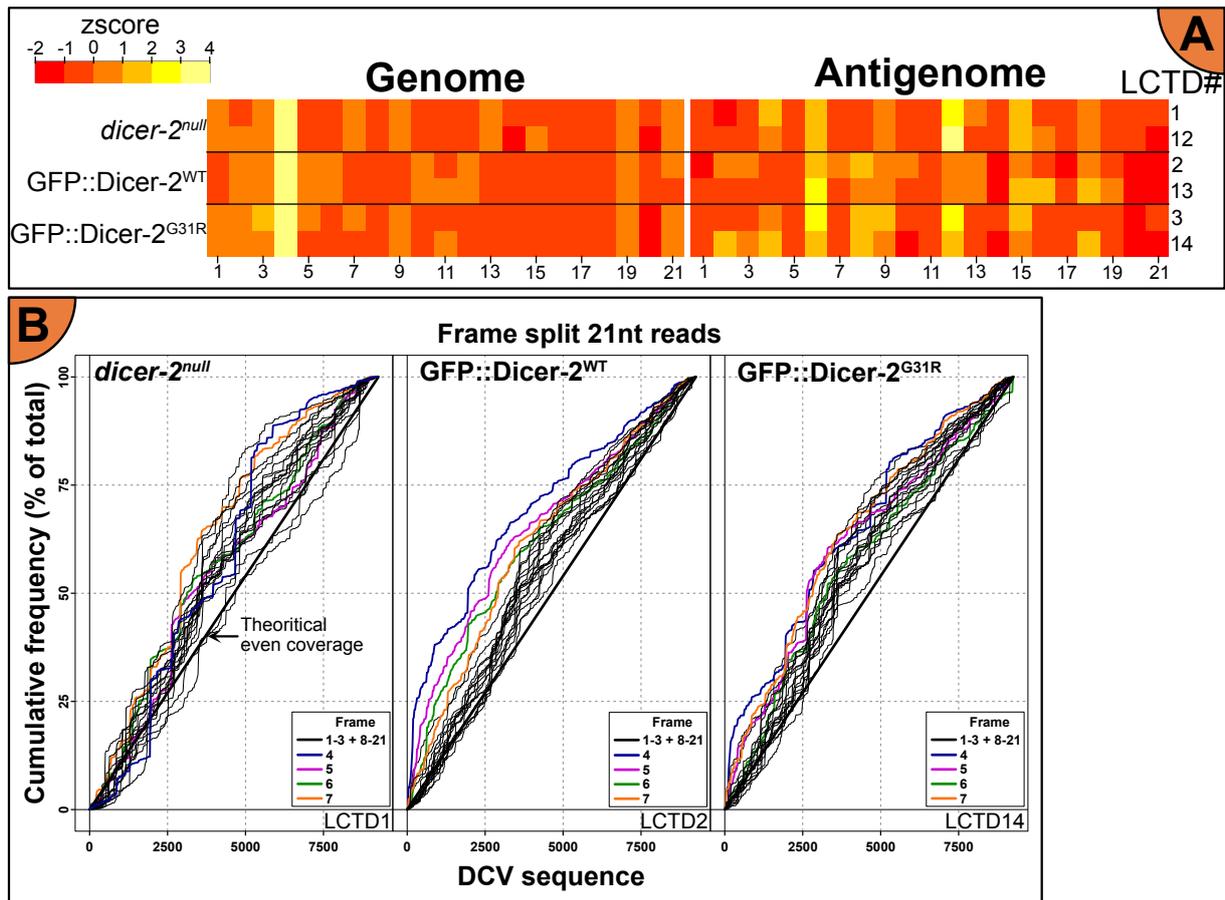


Figure 17: Frame-based analyses highlights a precise entry point of Dicer-2 in the 5' region of DCV.

A) Frame enrichment calculated individually for DCV genome and antigenome mapping 21nt long reads. **B)** Cumulative frequency of DCV coverage by independent frames. The three showed libraries are representative of their respective genotypes. The black line represents what would be a perfectly even coverage of the sequence.

siRNA signature by performing phasing and offset analyses. Because of the amount of 21nt long reads mapping on DCV sequence, analysis windows were reduced to 300nt. When looking at the most 5' 300 nucleotides, phasing analysis presents 2 peaks at 21 and 42nt mostly on the genome in GFP::Dicer-2^{WT} and GFP::Dicer-2^{G31R} conditions (**Figure 16A**). Moreover, a strong peak at 18nt can be observed in the same conditions in the offset analysis of the first 300 nucleotides (**Figure 16B**). This siRNA duplex signature cannot be identified by looking either at the *dicer-2^{null}* condition or further downstream the DCV sequence.

In conclusion, these analyses, as previously observed in cells, depict a strong Dicer-2 signature in the very 5' of the virus that quickly dims the further we go from the 5' extremity. Presence of this signature in the GFP::Dicer-2^{G31R} condition shows that the ATPase activity of Dicer-2 helicase is not mandatory to generate virus derived siRNA duplexes. However, as observed in **Figure 15A**, the number of 21nt long reads mapping DCV generated by GFP::Dicer-2^{G31R} is much lower than the ones generated by GFP::Dicer-2^{WT}. As further discussed in Chapter III, we can also noticed that GFP::Dicer-2^{G31R} expressing flies present much less 21nt long reads mapping on *w^{IR}* transgene in comparison to GFP::Dicer-2^{WT} (**Figure 36A**). This explains the red eyes of GFP::Dicer-2^{G31R} flies (**Figure 13B**). Thus, ATPase activity of Dicer-2 is not mandatory to generate some siRNA duplexes but is required to activate an efficient siRNA pathway in general, suggesting a threshold effect in quantity of siRNAs.

Dicer-2 has a precise entry point on viral dsRNA in the 5' region of DCV

Analysis of the frames' enrichment was done on the 21nt long reads in order to test whether Dicer-2 has a precise entry point on DCV. In a similar way to what has been seen in S2 cells, the reads in frame 4 mapping on the genome are highly enriched in the GFP::Dicer-2^{WT} and GFP::Dicer-2^{G31R} conditions (**Figure 17A**). However, no enrichment for frame 2 reads could be observed on the antigenome. One of the possible explanations for this is the very low amount of antigenome mapping reads identified in the libraries. From this analysis, it is quite surprising to see 21nt long reads in frame 4 mapping on the genome also being enriched in *dicer-2^{null}* condition. To determine whether this enrichment is real or due to the abundant presence of certain degradation products, we plotted the cumulative coverage frequency of DCV with all the frames split (**Figure 17B**). The GFP::Dicer-2^{WT} condition shows a similar distribution of the reads as the one observed in S2 cells: a 5' enrichment of frame 4/2 reads followed by frame 5/3 and 6/4 while moving

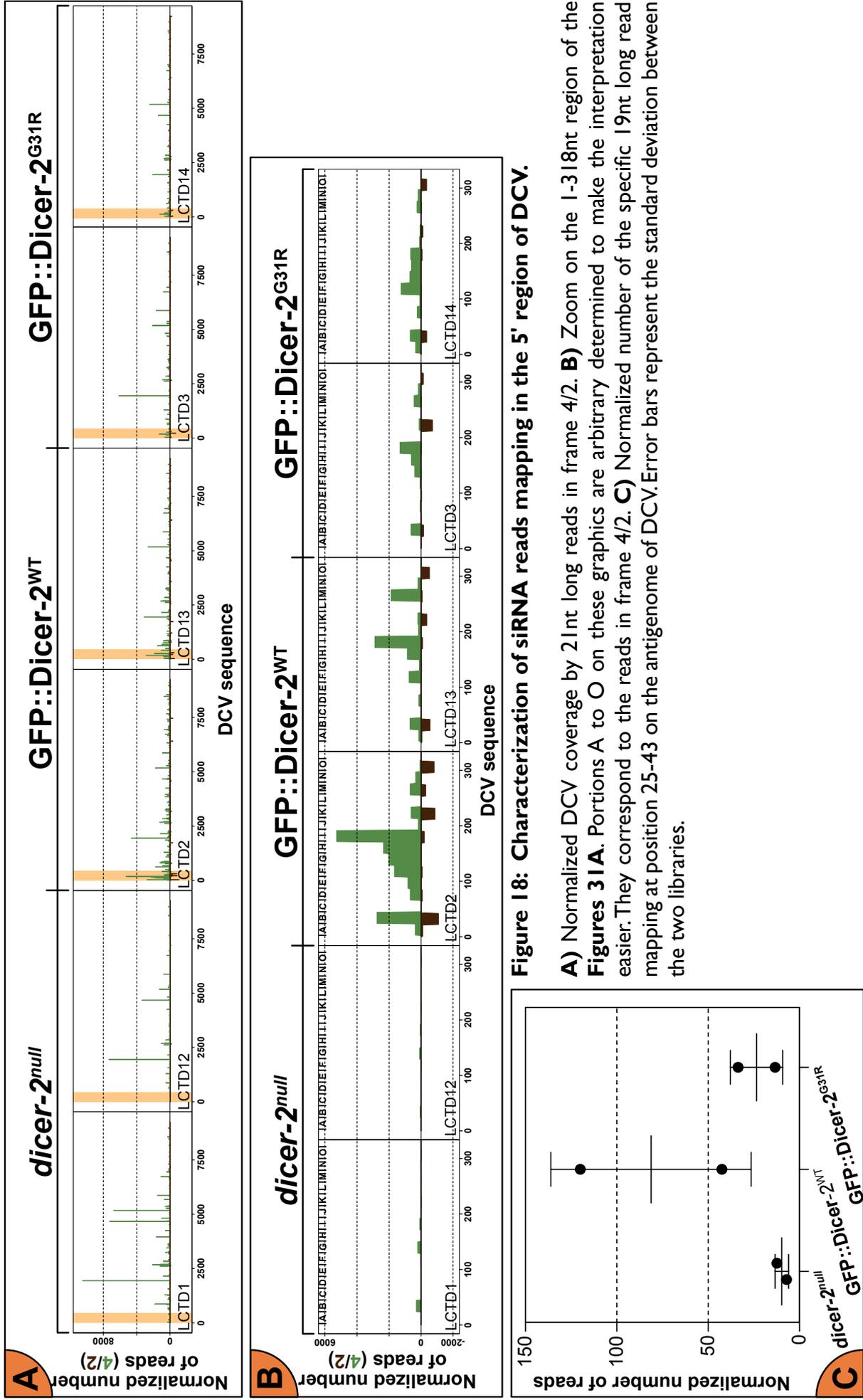


Figure 18: Characterization of siRNA reads mapping in the 5' region of DCV.

A) Normalized DCV coverage by 21nt long reads in frame 4/2. **B)** Zoom on the 1-318nt region of the **Figures 31A**. Portions A to O on these graphics are arbitrary determined to make the interpretation easier. They correspond to the reads in frame 4/2. **C)** Normalized number of the specific 19nt long read mapping at position 25-43 on the antenome of DCV. Error bars represent the standard deviation between the two libraries.

downstream DCV sequence. GFP::Dicer-2^{G31R} condition shows a similar trend for frame 4/2 to be enriched in the 5' region. However, after ~1000 nucleotides, an even coverage of DCV sequence can be observed regardless of the frame. Finally, no frames are enriched in a specific region of DCV sequence in *dicer-2^{null}* condition. We can thus conclude that the enrichment of genomic reads in frame 4 in *dicer-2^{null}* flies seems to be an unlucky coincidence and is most likely only coming from degradation products. Altogether, these analyses point toward a precise entry of Dicer-2 in an ATP-independent manner on the dsRNA corresponding to the 5' region of DCV.

Following the same procedure as before, we looked at the distribution of 21nt long reads in frame 4/2 on DCV sequence and zoomed in on the 5' region (*Figure 18A e' B*). Interestingly, distribution of the reads in GFP::Dicer-2^{WT} and GFP::Dicer-2^{G31R} conditions looked really similar to the one observed in S2 cells in *Figure 8B – left*. The only difference between the two genotypes lies in the number of reads observed at each position emphasizing the importance of ATP hydrolysis to generate many siRNAs. Finally, nothing similar could be observed in the *dicer-2^{null}* condition. Once again, the data obtained from sequencing of DCV infected flies correlate with the ones obtained in DCV infected S2 cells. We thus decided to investigate for the 21_{genome}/19_{antigenome} signature that was previously observed at position B. The number of 19nt long reads mapping at position 25-43 on DCV antigenome was determined in each library (*Figure 18C*). Nearly no reads corresponding to this characteristics could be observed in *dicer-2^{null}* or GFP::Dicer-2^{G31R} conditions. A higher number of reads were identified in the GFP::Dicer-2^{WT} condition. This result has to be tempered by the high variability that can be observed between the libraries and by the low number of reads observed. Nonetheless, being able to find this specific 19nt long read in flies libraries suggests its relevance. In conclusion, the small RNA HTS of DCV infected adult flies points toward a precise entry point of Dicer-2 on the 5' region of the virus dsRNA in an ATP-independent manner. The origin and specificities of this entry point will be further investigated and discussed in Chapter II.

Conclusions – Discussions – Perspectives

During my Ph.D, I conducted complementary small RNA HTS of infected S2 cells and drosophila flies in an attempt to understand how Dicer-2 is able to sense and process protected viral RNA. The kinetic study of the apparition of siRNAs done in S2 cells allowed to highlight what appears to be a specific Dicer-2 entry point on DCV and CrPV, two closely related dicistroviruses. Initial entry of Dicer-2 on both viruses was identified to be done on the dsRNA corresponding to the 5' region but not the 5' extremity. Importantly, a Dicer-2 siRNA signature could be identified as early as 3h pi in this region suggesting a rapid detection and processing of viral dsRNA by Dicer-2. The study conducted using different Dicer-2 variants expressing flies allowed further characterization and validation of this entry point. Because of the similar results obtained in DCV infected GFP::Dicer-2^{G31R} flies, I could propose an ATP-independent entry of Dicer-2 on its viral substrate. Despite strong tendencies precisely pointing toward the described entry point, several details remain obscure and need to be discussed.

a. Did I find Dicer-2 entry point?

Asymmetrical distribution of reads

In DCV libraries from flies and to a lesser extend from cells, I could consistently observe a bias in favor of genomic strand derived small RNAs (*Figures 5A e' 15A*). The asymmetrical distribution of the 21nt long virus-derived reads was already previously observed in DCV infected cells and FHV infected flies and is accompanied by abundant smaller and longer genomic RNAs (Han et al., 2011; Sabin et al., 2013). This information is contrary to the idea of Dicer-2 attacking a dsRNA molecule and subsequently generating small RNAs of 21nt long with a 1:1 ratio between genomic and antigenomic strands. Interestingly, such a canonical behavior is observed when sequencing the small RNAs of CrPV or VSV infected flies or cells (*Figure 10B*, Chapter III – *Figure 40C* and Sabin et al., 2013). One of the major differences between DCV, CrPV, FHV and VSV lies in the presence and mechanisms of viral suppressors of RNAi (VSRs). Indeed, while DCV and FHV both encode a VSR (1A and B2 respectively) able to bind dsRNA and protect it from degradation, CrPV produces a VSR inhibiting the RNAi pathway at Ago2 level (1A) and VSV has no known VSR (Lu et al., 2005; Nayak et al., 2018; van Rij et al., 2006). Interestingly, it was shown that flies expressing a FHV replicon lacking B2 only accumulate 21nt long virus-derived reads in a 1:1 ratio between genomic and antigenomic

strands (Han et al., 2011; Martins et al., 2019). Two hypotheses emerged from these observations: DCV and FHV VSRs efficiently protect viral dsRNA intermediate of replication and accumulation of genomic small RNAs is due (1) to the action of other nucleases acting on the overly abundant genomic strand or (2) to the action of Dicer-2 on genomic ssRNA secondary structures. As further discussed in Chapter II, the establishment of an infective DCV replicon should be done to allow a better characterization of the DCV 1A VSR.

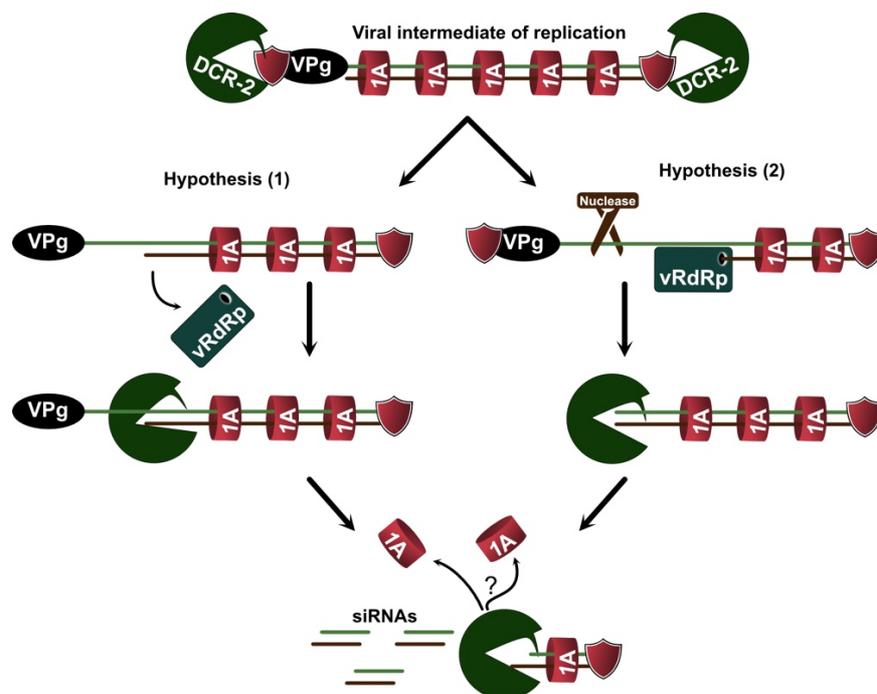
Sequencing of DCV infected *dicer-2^{null}* flies reveals that these non-21nt long reads are Dicer-2 independent (**Figure 15A**). Moreover, in-house small RNA sequencing of Ago2 immunoprecipitations after DCV infection of S2 cells revealed a single symmetrical peak of virus derived small RNAs of 21nt long (unpublished data from our lab.). Altogether, these data suggest that viruses encoding for a dsRNA-binding VSR are subject to both precise Dicer-2 processing of dsRNA in 21nt long siRNAs and random nuclease degradation of abundant (+) strand. It could be interesting to further investigate the pathway involved in the generation of these degradation products and decipher whether they have an impact on the antiviral response. Such a study could be done in the *dicer-2^{null}* flies used here or alternatively, in *dicer-2* knock-out S2 cell line (Kunzelmann et al., 2016). Different levels of degradation products could be observed between S2 cells (low level – early time points of infection) and flies (high level – late time points of infection) suggesting that they accumulate over time. Importantly, the bioinformatic analyses of both models' small RNA HTS were consistently pointing toward the same entry point of Dicer-2 on viral dsRNA.

How can Dicer-2 access an internal entry site?

The analyses of sequencing data allowed to narrow-down the research zone for Dicer-2 entry point in the 5' regions of DCV and CrPV. In DCV data more than in CrPV ones, a very peculiar pattern of reads distribution in this 5' region could be observed and is conserved between cells and flies of different genotypes (**Figure 8B e³ 18B**). This pattern consistently starts with what appears to be a canonical siRNA signature at position 23-45 (25-45 on genomic strand and 23-43 on antigenomic strand) and a 1:1 ratio between strands. No siRNA signature or any small RNA reads accumulation could be observed upstream of this one, which suggests that it could be the entry point of Dicer-2 on DCV dsRNA. Two hypotheses could be proposed to explain this internal entry of Dicer-2 and both rely on the generation of alternative dsRNA extremities:

- (1) The RdRP of RNA viruses is highly processive but also very relaxed, which leads to a high mutation rate and potentially, dissociation from the RNA substrate (Sanjuán, 2012; Sanjuán et al., 2010). Indeed, as later explained in Chapter II, the identified entry point of Dicer-2 in DCV and CrPV is located in a very structured region of the 5'UTR, which might be a difficult template for the viral polymerase. Additionally, recombination events are also frequent in RNA viruses and give rise to defective viral RNAs that lack portions of their genome (Pathak and Nagy, 2009; Routh et al., 2012). Such defective particles could be present in our virus stock or generated in the time course of the infection and be the primary targets of Dicer-2.
- (2) In its dsRNA form, DCV is coated by its VSR 1A and protected from degradation. Its ssRNA genome, however, is apparently only protected at both extremities and could eventually be targeted by a host endonuclease. Involvement of Dicer-2 itself in this ssRNA targeting will be investigated in Chapter II by performing cleavage assays using recombinant proteins or embryonic extracts. The newly generated 5' extremity could be made double stranded by the concomitant action of the viral RdRP and would become a perfect template for Dicer-2 to enter.

These hypotheses would lead to the creation of unprotected viral dsRNA extremities.



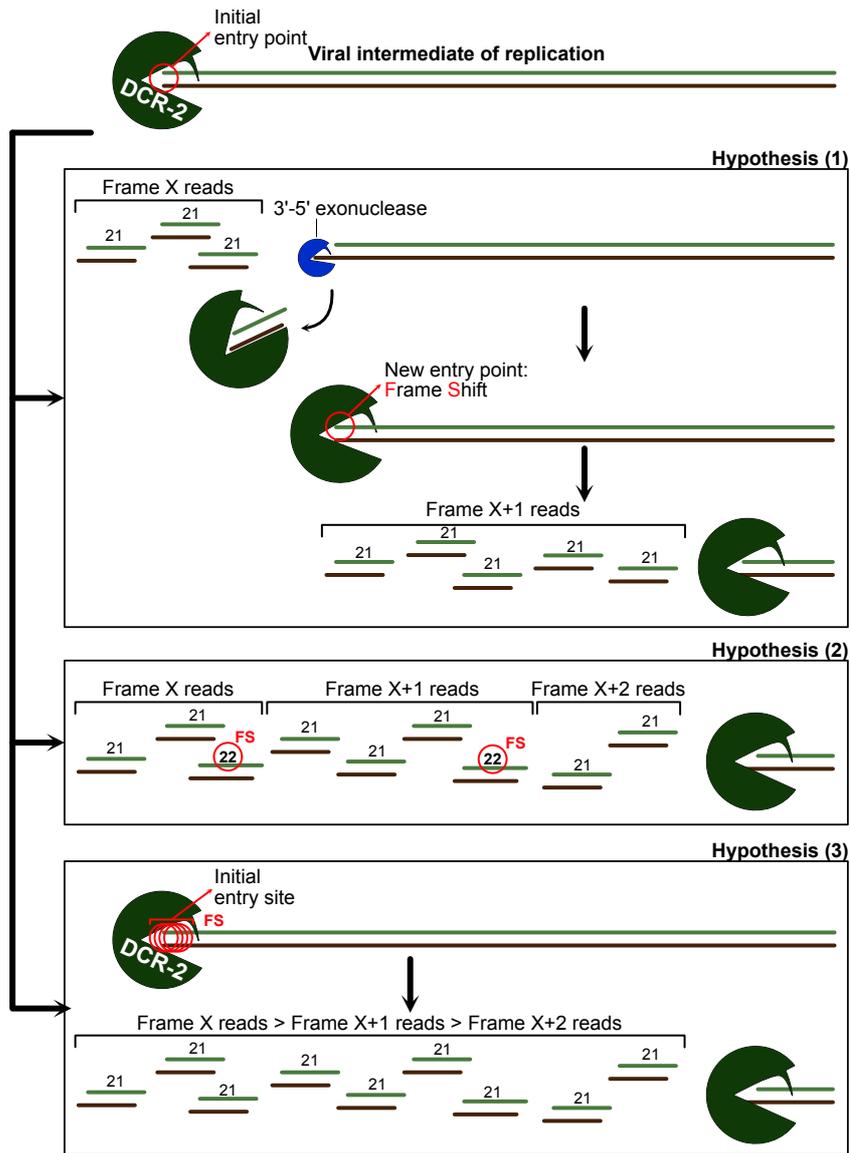
Attempts at identifying this internal entry point of Dicer-2 on DCV will be further explained in Chapter II. However, these small RNA HTS data are not sufficient to validate the exact nature of Dicer-2 entry point and many questions remain: (1) How can I reconcile the simultaneous presence of what appear to be blunt dsRNA extremity (21/19 siRNA) and 2nt 3' overhang dsRNA extremity (21/21 siRNA) at the position of the very first siRNA signature? (2) Is the proposed entry point the only one or the strongest/first one? (3) How can I explain the peculiar distribution pattern in the 5' region and what does it tell about Dicer-2 mode of action? (4) Is Dicer-2 able to displace bound 1A protein upon processing of the viral dsRNA? Alternatively, is 1A a limiting protein during the replication cycle that would lead to some unprotected portions of dsRNA intermediates?

Some of the answers we are looking for could be obtained in the future by performing PAR-CLIP experiments. Indeed, such a study was previously conducted in human cells and in *C. elegans* and highlighted the unanticipated wide diversity of Dicer substrates in both these organisms (Rybak-Wolf et al., 2014). Combination of this technic with the usage of catalytically inactive Dicer-2 (mutations in RNaseIII domains) and RNA sequencing should allow us to precisely identify the initial binding sites of Dicer-2 on viral (and endogenous) targets.

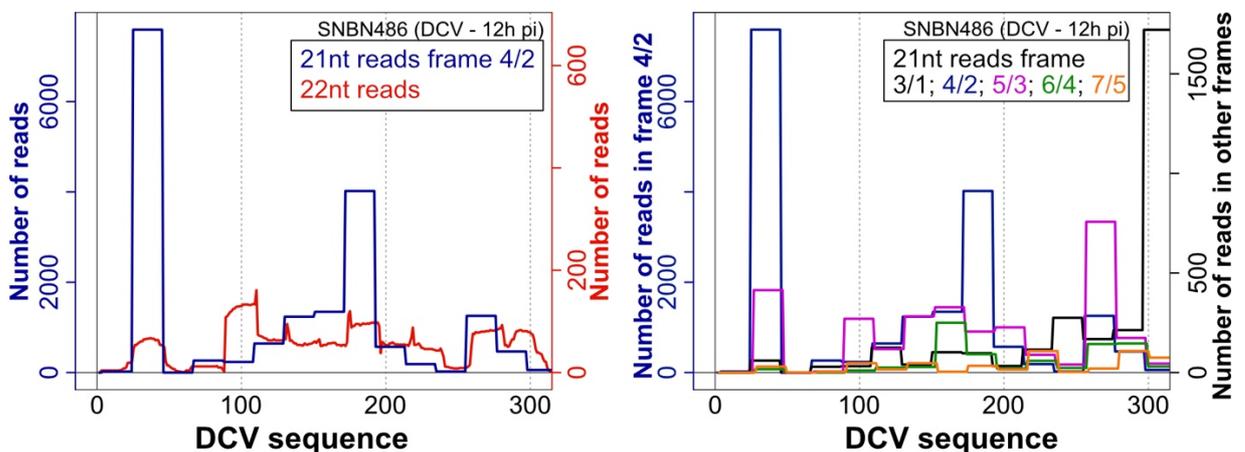
Decrease of siRNA signature across viral sequences

Another observation that could be made and that goes in the direction of a precise entry point of Dicer-2 in the 5' region of both viruses is the dimming of siRNA signature from 5' to 3'. This information was obtained by looking at the phasing and offset of 21nt long reads as well as at the cumulative frequency coverage of viral sequences with every frame split (**Figure 6, 7C, 11, 12B, 16 e³ 17B**). Thus, in both DCV infected cells and flies, a strong siRNA signature could be seen starting at frame 4/2 in the 5' region and quickly diminishing. Interestingly, this enrichment in frame 4/2 is followed by subsequent enrichments in frame 5/3, 6/4 and 7/5. A similar but less striking phenomenon could be observed in CrPV infected cells with an initial enrichment observed in frame 5/3 followed by an enrichment in frame 6/4. Three hypotheses were raised to explain this rapid shift in dominant frame:

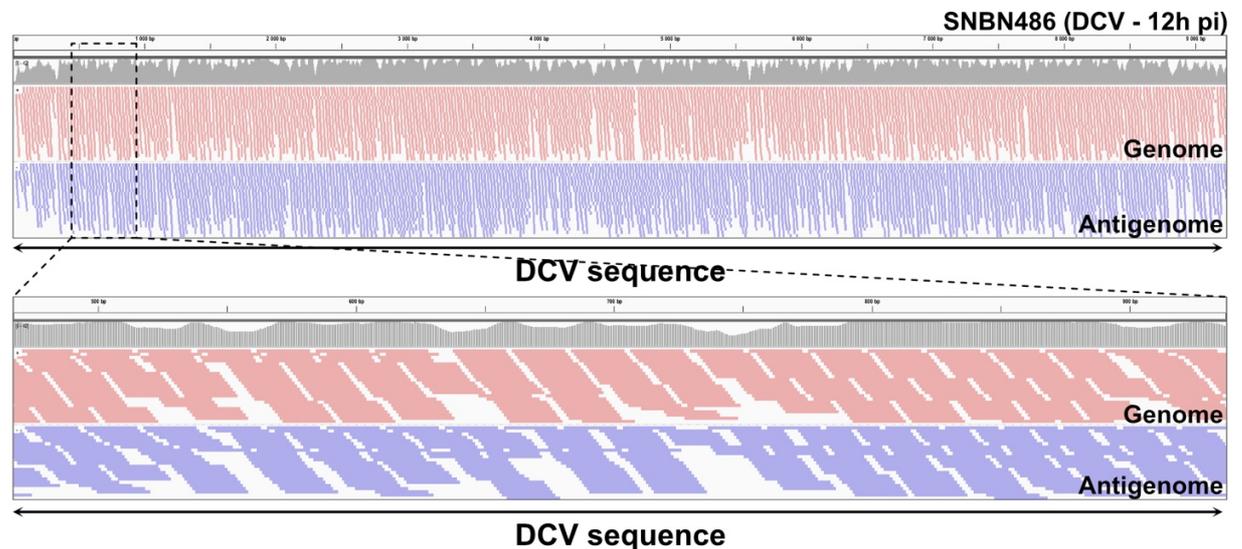
- (1) Dicer-2, after an initial precise entry, could fall from its substrate, leaving a 2nt 3' overhang substrate susceptible to 3'-5' degradation. Subsequent sensing by another Dicer-2 enzyme would result in a shifted frame. This hypothesis is supported by *in vitro* studies showing that Dicer-2 can process its substrate in a distributive manner and dissociating after every cut (Cenik et al., 2011; Sinha et al., 2015; Welker et al., 2011). Moreover, in DCV libraries, a small number of antigenomic 19nt long reads could be consistently observed mapping in front of 21nt long genomic reads in frame 4/2 (*Figure 9D*). Preliminary studies in CrPV libraries reveal a similar phenomenon in frame 5/3.
- (2) Dicer-2 can randomly produce 22nt long reads when processing viral dsRNA. This is supported by the fact that some endogenous Dicer-2 products like *esi-1* and *esi-2* derived endo-siRNAs are 22nt long (see Chapter III and Marques et al., 2010). Even if not predominant by looking at reads size distribution, a small number of 22nt long reads could be enough to generate a heterogeneity dampening the phasing signal of viral siRNAs.
- (3) The initial entry of Dicer-2 on its dsRNA substrate might not be as precise as we expect it to be and could be subject to little variations in the precise nucleotide it is starting from. Thus, the supposed entry site in frame 4/2 might not be the only one but the most occurring one.



Interestingly, hotspots of 22nt long reads could be observed at the same positions as enriched 21nt long reads in the 5' region and in frame 4/2. Moreover, 21nt long reads in other frames can be seen following a pattern roughly similar to the frame 4/2 reads.



Then, in order to fully appreciate the diversity of reads in my sequencing data without being influenced by their relative abundance, I concatenated all similar 21nt long reads (same mapping position) as single ones. Observation of these concatenated reads highlighted the fact that virtually every possible starting position is occupied by at least one read.



Thus, it is more than likely that the final answer of Dicer-2 entry and mechanism lies in a mix of the hypotheses explained above. These lastly presented analyses are preliminary work done on a subset of libraries and would require further investigation to quantify and characterize the observed phenomenon. In order to precisely monitor this sliding effect of Dicer-2 on its dsRNA substrate, it could also be interesting to couple *in vitro* dicing experiments with small RNA HTS. Indeed, to my knowledge, the only output of these methods remains the size of the small RNA products without any information on the sequence of these. The very sensitive sequencing method would allow the detection of non-canonical and ill-detected subpopulations of RNAs and could help us answer some questions such as: (1) What are the sizes of the siRNA duplex strands generated from a dsRNA with blunt extremities (21/19 or 23/21)? (2) Is the sliding effect measurable and sequence dependent?

b. Technical comments

Detection of viral antigenome at 0h pi time point

When preparing small RNA libraries from infected S2 cells, I checked for the absolute number of genomic and antigenomic strands at each time point to show that a productive infection was occurring. An intriguing observation was the presence of antigenomic

strands at 0h pi (*Figure 3B e³ 10A*). I cannot exclude the fact that the viral stocks, even if purified on sucrose gradient, could have a minimal antigenomic contamination that would be detected by the very sensitive qPCR method. A strand-specific qPCR should be done on the viral stock to check for the presence of antigenomic strands. Alternatively, the presence of these RNAs could reflect the fact that the infection started earlier than expected and that a first round of antigenome synthesis occurred in some infected cells.

Viral loads of infected flies

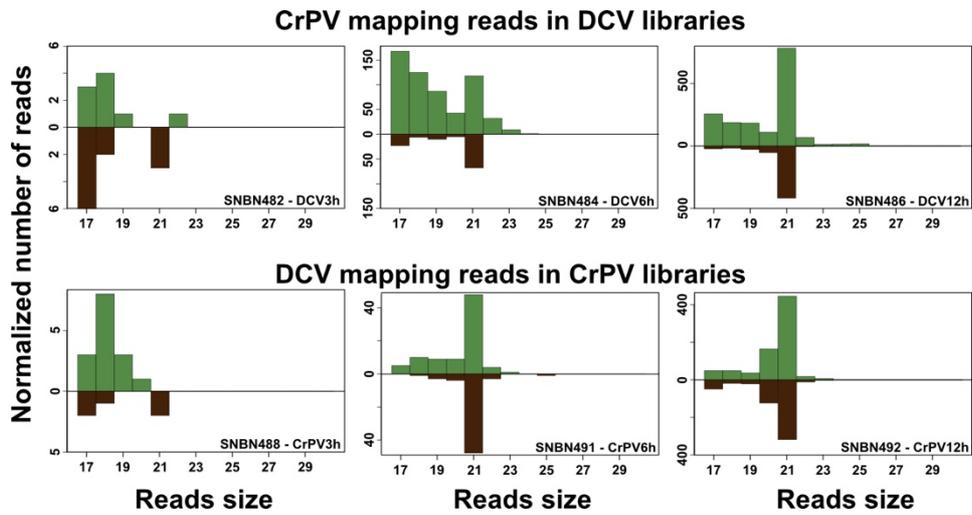
Intriguingly, when checking for viral loads in infected flies, I could not detect any significant difference between *dicer-2^{null}* flies and GFP::Dicer-2^{WT} flies for instance (*Figure 14A*). Explanation about this phenotype will be given in Chapter III. Briefly, it appears that 2d pi, these fast replicating viruses already reach a threshold of infection. In this case, small RNA libraries were prepared from RNA samples extracted from 3d pi flies, explaining the similar viral loads between all genotypes.

Preparation of the libraries

The efficiency of the 2S ribosomal RNA blocking step was proven when preparing libraries from infected flies that contained a low number of reads mapping to this abundant RNA (*Figure 14B*). At the moment of the preparation of small RNA libraries from S2 infected cells, no 2S RNA blocking primer was added. This resulted in contaminated libraries with most of the reads mapping on 2S (*Figure 4A*). Thus, in this case, addition of the blocking primer could have greatly enhanced the number of virus mapping reads, especially at 3h pi. However, the number of virus mapping reads were sufficient at 6 and 12h pi to draw my conclusions.

In addition to the 2S ribosomal RNA contamination, my data clearly emphasized the presence and importance of a sample bleeding phenomenon in small RNA HTS. This bias induced by the multiplexing of libraries is a major concern and can lead to false biological interpretations, especially when looking at small populations of RNAs. One obvious but economically impossible solution would be to sequence only one library per sequencing lane. However, this solution would not take into account possible reagent contamination of the sequencing machine. The method I decided to use was the construction of double-indexed libraries allowing a second demultiplexing. This second sorting of the reads is supposed to reduce the error rate to less than 0,01%. As a proof of concept, I detected

virtually no CrPV specific reads in DCV libraries nor DCV specific reads in CrPV libraries.



Importantly here, reads detected as mapping on the opposite viral sequence (CrPV for DCV libraries and DCV for CrPV libraries) are located in regions encompassing identical sequences between the two closely related viruses.

Altogether, I would recommend the usage of a 2S blocking primer as well as library specific 5' adaptors when performing small RNA HTS of drosophila cells or flies. In addition, precise information about the entirety of the sequencing lane samples should be given when uploading RNA sequencing data online for publication and extra care should be taken when dealing with small number of reads.

Introduction

In the previous chapter, thorough investigation of the apparition of siRNAs in DCV and CrPV infected conditions was conducted. The small RNA HTS performed simultaneously in flies and in S2 cells pointed toward a precise entry point of Dicer-2 on the viral dsRNA, without us being able to identify it. Still, strong evidences suggest that this entry point is located in the 5' untranslated region (UTR) of both studied viruses. In dicistroviruses, this region encompasses a wide diversity of RNA structures of uttermost importance for their replication, translation, virulence and interaction with host proteins.

As a reminder, dicistroviruses are bicistronic viruses with translation of ORF1 under the control of the 5'IRES (located in the 5'UTR) and translation of ORF2 under the control of the internal IRES (IGR – Chapter I - *Figure 1*). Because it requires no translation initiation factor, the dicistrovirus' IGR is among the simplest 40S ribosome recruiting structure and is well conserved between these viruses (Jan and Sarnow, 2002; Pestova, 2003). On the contrary, 5' IRES structural organization and initiation factors requirement is much more variable from one virus to another, which could reflect distinct adaptation strategies used to hijack the host ribosome. For instance, DCV and CrPV require a ribosomal protein called Rack1 to control their 5'IRES-mediated translation (Majzoub et al., 2014).

A recent study aiming at better understanding the 5'IRES-mediated translation highlighted the interaction between this RNA structure and the eukaryotic initiation factor 3 subunits (Gross et al., 2017). Dicistroviruses' 5' UTR encompasses both this 5'IRES structure and an additional structured domain located upstream: the domain I. According to our small RNA sequencing data, Dicer-2 entry point would be located in this domain I region. In the frame of the previously cited study, the 2D structures of both these domains were determined for CrPV. The domain I model revealed a structure encompassing 5 simple stem-loops separated by short ssRNA linker regions. However, further characterization of this region using a DMS/CMCT chemical probing method revealed a structure originally described in poliovirus as a cloverleaf (unpublished data from Dr. Franck Martin's team, IBMC-UPR9002, Strasbourg – *Figure 25*). In picornaviruses, this region was shown to interact with the viral RdRP and with host proteins to promote synthesis of negative and positive strands (Andino et al., 1990; Barton

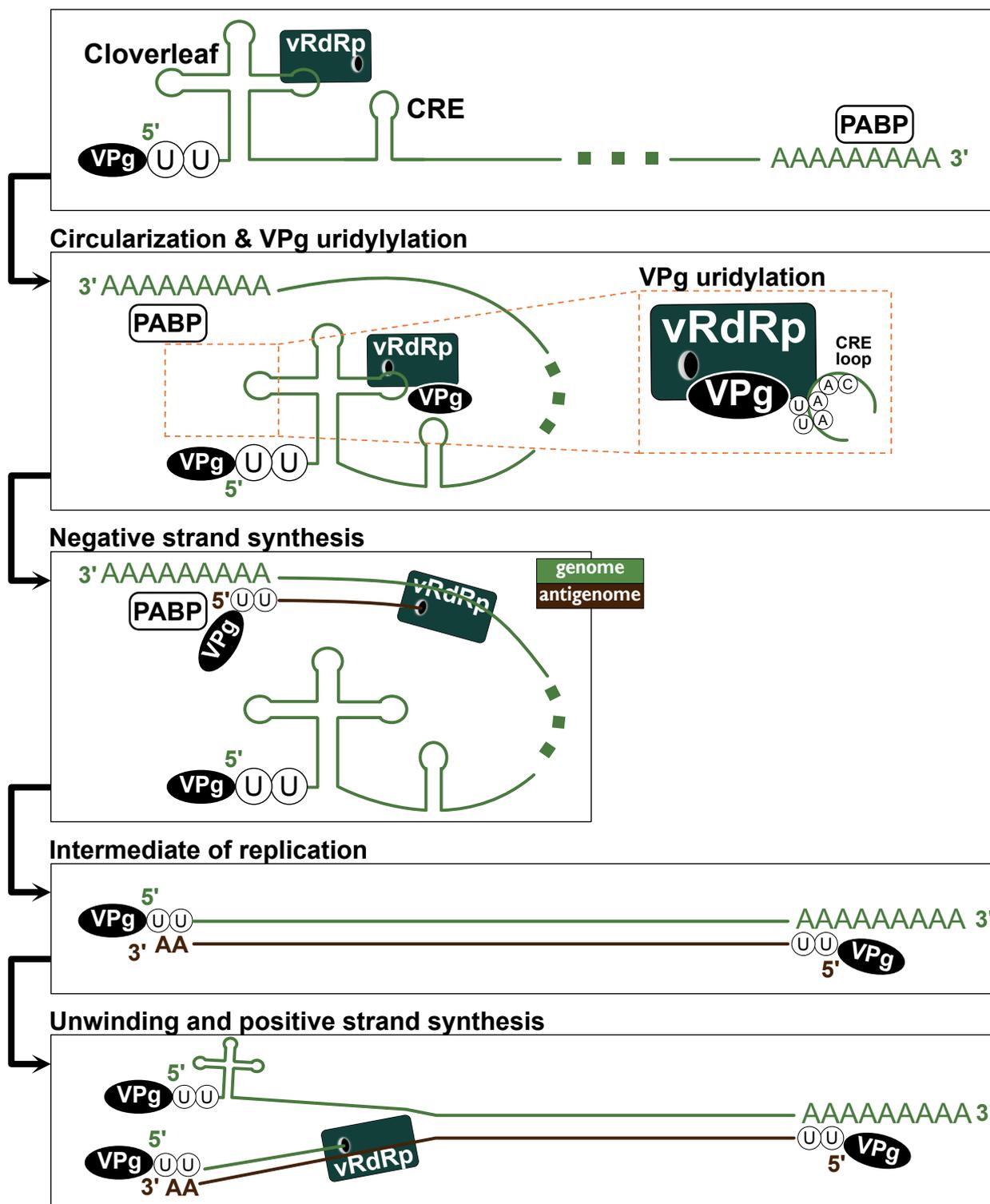


Figure 19: Model of picornaviruses replication strategy.

Concomitant binding of host Poly A Binding Protein (PABP) to viral genome poly A tail and binding of viral RdRP to the 5' UTR cloverleaf structure promotes circularization of the genome. In addition, the vRdRp promotes the uridylylation of a newly translated VPg by using a picornaviruses-conserved **AAAY** (with Y a pyrimidine) motif found in a Cis-active RNA element (CRE). This initiates the protein-primed synthesis of the negative strand (antigenome). Thus, the resulting intermediate of replication is protected at both extremities by a VPg. It is unclear how unwinding of this dsRNA starts to initiate the protein-primed synthesis of the genome strand. One hypothesis lies in the refolding of the genomic cloverleaf structure and subsequent interaction of the VPg-U-U with 3' terminal AA nucleic acids of the antigenome.

et al., 2001; Gamarnik and Andino, 1997; Herold and Andino, 2001; Vogt and Andino, 2010). Moreover, it was shown that interaction between the cloverleaf structure, vRdRP and VPg is responsible for the initiation of protein-primed picornavirus RNA synthesis (Lyons et al., 2001; Rieder et al., 2000). The current proposed model for negative and positive strands synthesis in picornaviruses is presented in *Figure 19*.

In picornaviruses and in the predicted CrPV 5'UTR RNA structure, this cloverleaf is located in the very 5' region of the viral genome, which contains the hypothetical entry point of Dicer-2. Importantly, (1) we do not understand how Dicer-2 could have access to an unprotected dsRNA extremity different from the viral 5' extremity and (2), we do not know the 2D structure of DCV domain I, which also seems to encompass Dicer-2 entry point. Thus, further characterization of this region of DCV genome is required and was performed during my Ph.D. To begin with, a model of DCV domain I 2D RNA structure was obtained by performing *in vitro* chemical probing. Then, with the aim of finding the precise entry point of Dicer-2, characterization of the cleavage sensitivity of this ssRNA was determined *in vitro* using either recombinant Dicer-2 protein or embryonic extracts of flies. Finally, an attempt at determining 5' extremities produced in infected conditions was done using a 5' Rapid Amplification of cDNA Ends (RACE) method. Altogether, these techniques were used to gain further insights in this uncharacterized region of DCV.

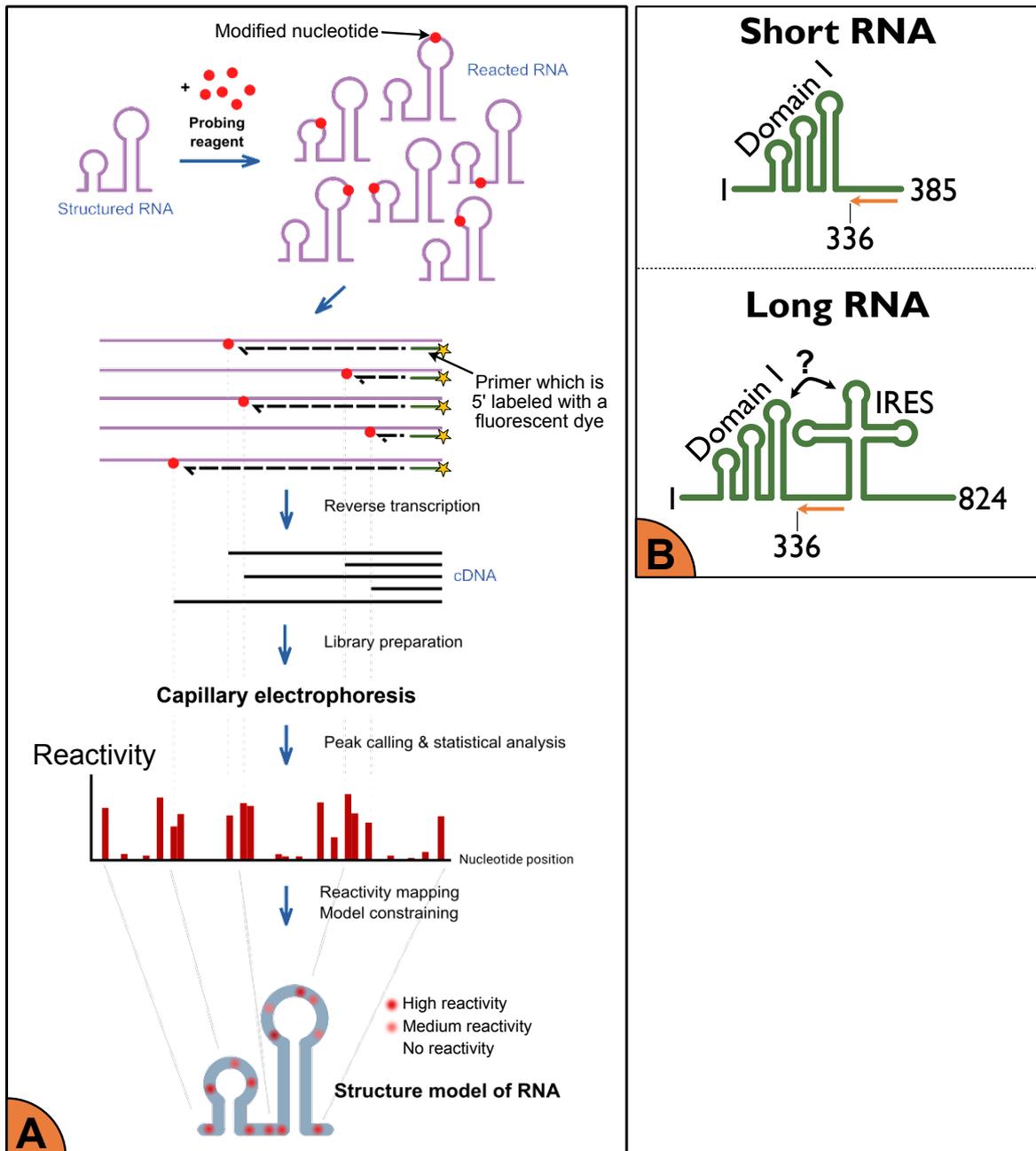


Figure 20: Experimental setup for the determination of DCV domain I 2D structure.

A) Adapted from https://en.wikipedia.org/wiki/Nucleic_acid_structure_determination. Schematic explanation of the determination of secondary structures in ssRNA by using chemical probing. Briefly, the studied synthetic RNA is incubated with chemicals that will randomly react with non-base-paired nucleotides in the secondary structure. The modified nucleotides will block the following reverse transcription reaction conducted using fluorescent primers. The obtained cDNA molecules are run in an capillary electrophoresis apparatus to precisely measure their length. This information allows us to attribute a reactivity score to each nucleotide or a value representing how accessible it is. **B)** Two different ssRNA substrates were used for chemical probing. It has to be noted that the same RT primer was used (orange arrow) meaning that only the domain I region was studied.

Results

I. Analysis of the domain I structures of DCV and CrPV

a. Determination of DCV domain I 2D structure by chemical probing

Experimental setup

We used chemical probing to determine the 2D structure of the domain I of DCV (1-318nt). Briefly, this method relies on the random chemical modification of non-base-paired nucleotides in the RNA structure (*Figure 20A*). These RNAs are used for reverse transcription with fluorescent primers that will be stopped by the chemical modifications. The lengths of obtained cDNA products are measured by capillary electrophoresis. Finally, a “reactivity” score is attributed to each nucleotide. This score reflects the number of reverse transcription events that stopped at this specific nucleotide because of its chemical modification. Thus, the higher the reactivity of a nucleotide is, the lower the chances are that it is base paired in the RNA secondary structure. This experiment was performed in collaboration with the team of Dr. Franck Martin.

Chemical probing of DCV domain I RNA was done using two complementary chemical modification methods:

- Dimethyl sulfate (DMS), which reacts mainly with N1 of Adenosine and N3 of Cytidine
- 1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate (CMCT), which reacts mainly with N3 of Uridine and N1 of Guanine.

Moreover, chemical probing was performed on two versions of the DCV 5'UTR RNA: one encompassing the DCV domain I and the 5' IRES (**long RNA** – 1-824nt) and one composed of the DCV domain I alone (**short RNA** – 1-385nt). By doing this, we wanted (1) to maximize the chances of having a synthetic RNA folding as closely as possible to the *in vivo* secondary structure and (2) to check for possible interactions between the domain I and the 5' IRES (*Figure 20B*).

Long RNA

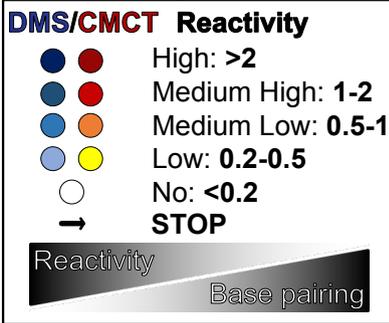
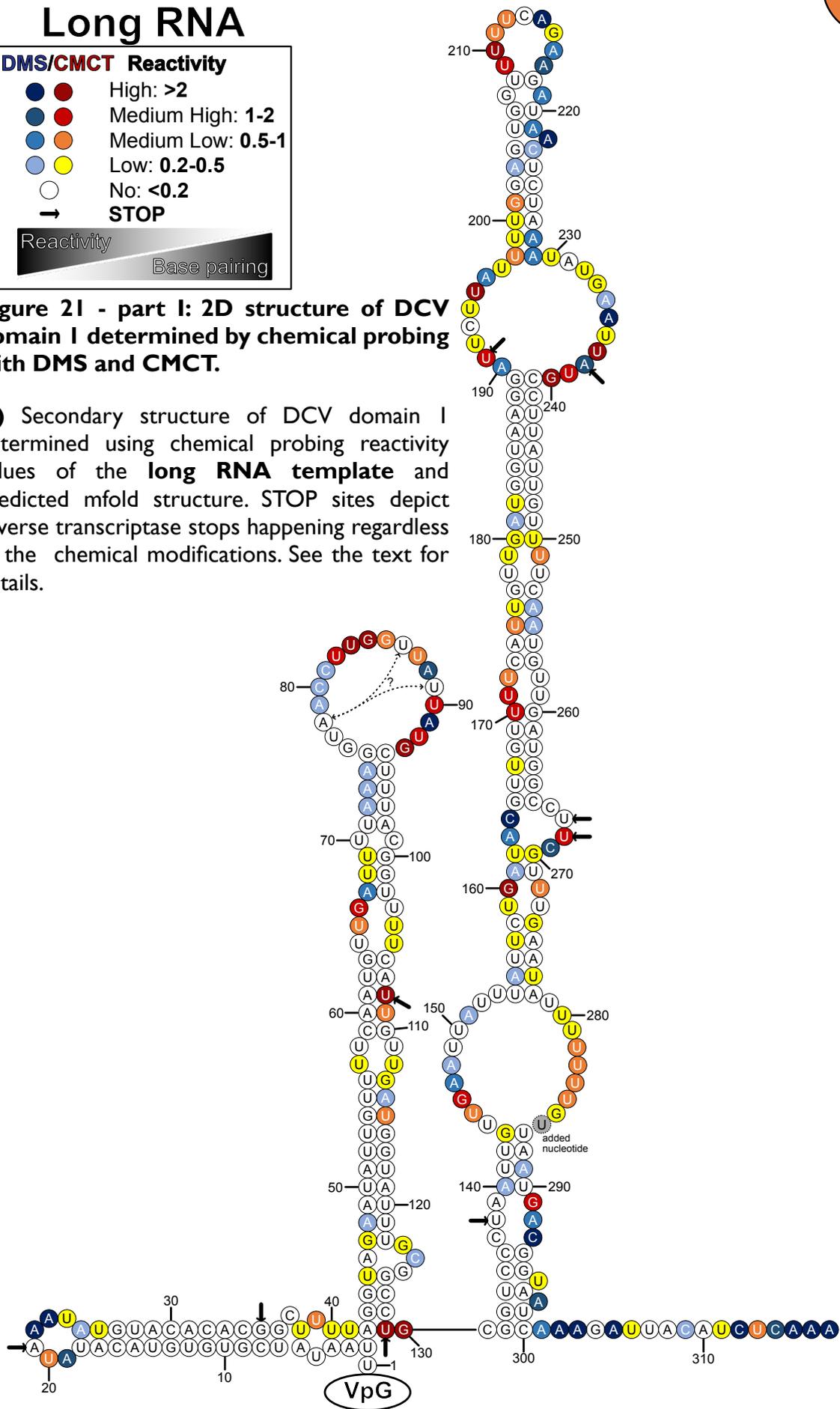


Figure 2I - part I: 2D structure of DCV domain I determined by chemical probing with DMS and CMCT.

A) Secondary structure of DCV domain I determined using chemical probing reactivity values of the **long RNA template** and predicted mfold structure. STOP sites depict reverse transcriptase stops happening regardless of the chemical modifications. See the text for details.



Short RNA

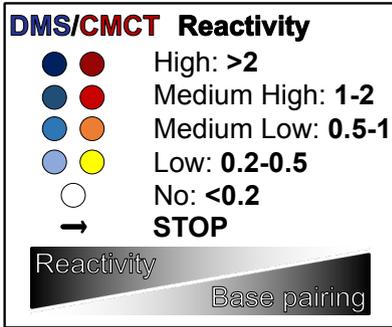
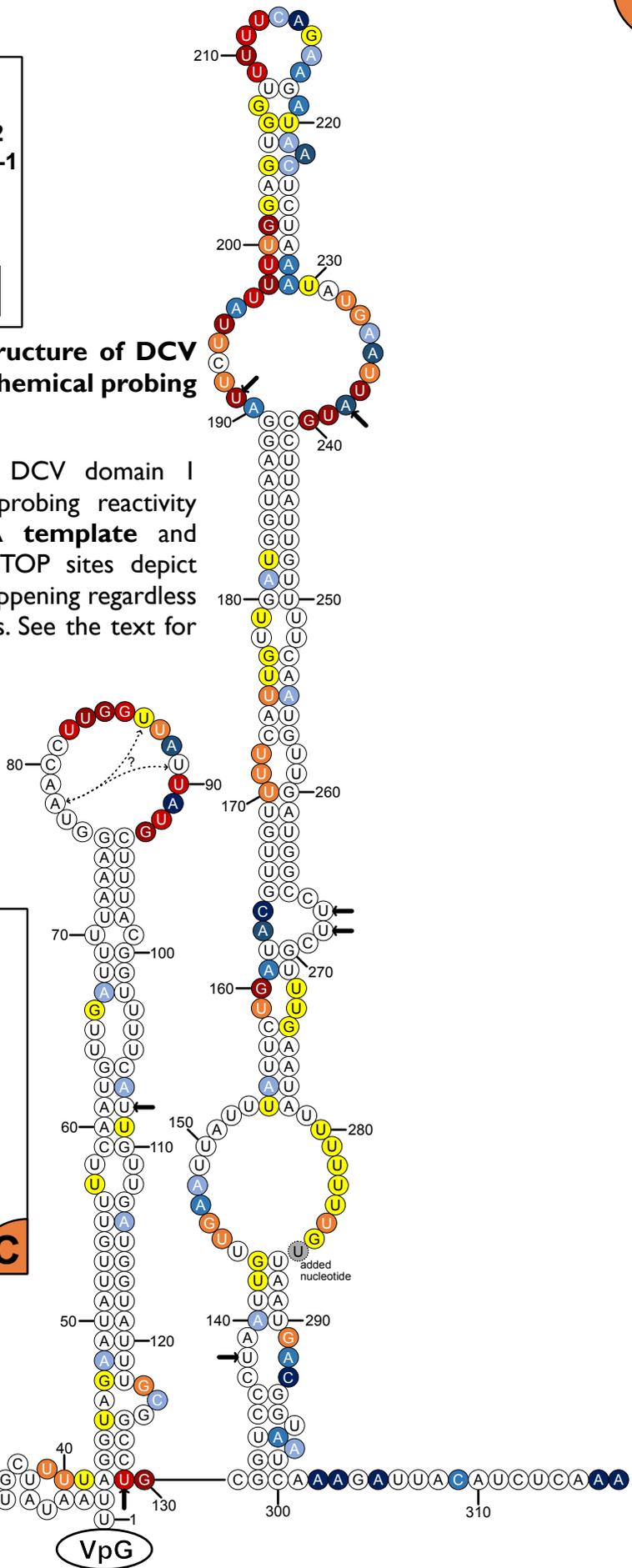
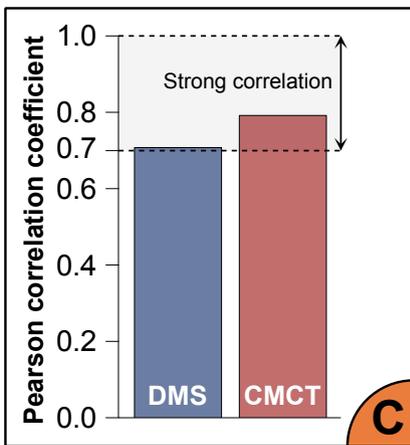


Figure 2I - part II: 2D structure of DCV domain I determined by chemical probing with DMS and CMCT.

B) Secondary structure of DCV domain I determined using chemical probing reactivity values of the **short RNA template** and predicted mfold structure. STOP sites depict reverse transcriptase stops happening regardless of the chemical modifications. See the text for details.

C) Pearson correlation for DMS and CMCT reactivities between long and short RNAs. A $R_{Pearson} > 0.7$ suggests no statistical difference between the two RNAs used.



B

C

Characterization of DCV domain I 2D structure

Chemical probing of short and long RNAs were performed in triplicates to determine the average reactivity value of each nucleotide. These data were used together with the help of mfold (<http://unafold.rna.albany.edu/?q=mfold>) predictions of the RNA structure to draw a model of the 2D structure of DCV domain I (**Figure 21A e³ B**). Before going in the detail of the structure, we wanted to check whether we could detect an interaction between domain I and 5'IRES using this method. Calculation of Pearson correlation coefficients for DMS and CMCT profiles revealed highly similar reactivities ($R_{Pearson} > 0.7$) between the long and the short RNAs (**Figure 21C**). This information indicates that the domain I and the 5'IRES fold independently and thus, do not interact *in vitro*.

The secondary structure of DCV domain I could be recapitulated in 3 stem loops:

- The first stem loop is predicted to be short, tightly associated and to start at the very 5' end of the DCV genome (nucleotides 2-42). Moreover, its mfold-predicted free energy is very low ($\Delta G = -16.10 \text{ kcal/mol}$) making its accessibility quite hard. In addition, no dangling tail that could be targeted by exonucleases is present upstream of this stem. Overall, the 5' extremity of the viral genome seems very protected and hard to access.
- The second predicted stem loop is bigger (43-129) and has a few wobbling sites. It is intriguing that nucleotides at position 75-81 present no or nearly no reactivity to both DMS and CMCT while they are present in a large loop. Nevertheless, we could not draw the model in a way that would accommodate the apparent base pairing of these nucleotides. The only hypothesis we could raise for this phenomenon is a possible intra loop interaction with uridines at position 86 and 89 that are no or very little reactive. Still, this second stem loop is tightly associated ($\Delta G = -17.80 \text{ kcal/mol}$) and hard to access.
- The third stem loop is much bigger (123-300), has a tiny terminal loop and four internal bulges. It has to be noted that a uridine has been added between positions 286 and 287. This nucleotide is present in our cloned sequence of DCV but is absent from the genome of DCV we used to perform the infections (determined by the small RNA HTS data). Because of its localization and the reactivity of the surrounding nucleotides, it is highly unlikely that this single nucleotide addition had an impact on the global structure of the DCV domain I model.

Sequence ID: Query_221105 Length: 9185 Number of Matches: 1			
Range 1: 1 to 9185 Graphics Full length		▼ Next Match ▲ Previous Match	
NW Score	Identities	Gaps	Strand
315	5974/9640(62%)	831/9640(8%)	Plus/Plus
Sequence ID: Query_228501 Length: 300 Number of Matches: 1			
Range 1: 1 to 300 Graphics First 300nt		▼ Next Match ▲ Previous Match	
NW Score	Identities	Gaps	Strand
-253	161/338(48%)	76/338(22%)	Plus/Plus

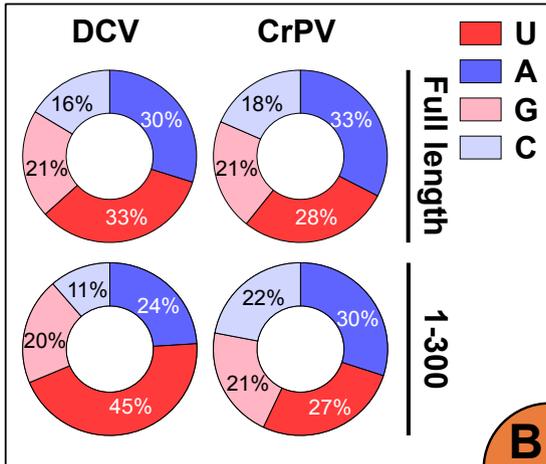


Figure 22: Comparison between DCV and CrPV sequences.

A) Top: Blast Global alignment (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) result of the full-length sequence of DCV against the full length sequence of CrPV. Bottom: Blast Global alignment result of the first 300nt of DCV against the first 300nt of CrPV. **B)** Base composition of either full length (**top**) or first 300nt (**bottom**) of DCV and CrPV sequences. DCV sequence is 9260nt long. CrPV sequence is 9185nt long.

Overall, the determined DCV domain I 2D structure is pretty simple with only three stem loops. It reveals a highly structured motif with only a few flexible ssRNA regions. According to this model, we can predict a really stable structure that could, in addition to the VPg, prevent 5'-3' exonuclease degradation.

b. DCV and CrPV domain I models do not compare

DCV and CrPV both belong to the *Cripavirus* genus of the *Dicistroviridae* family. When looking at their full-size genome sequence similarity using BLAST – Global Alignment tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), 62% of identity could be determined (**Figure 22A**). Because of their apparent similarity, one could also expect their 5' domain I to also present a high identity percentage. However, when using the first 300 nucleotides of both genomes only (which encompass the domain I structures), global alignment dropped to 48% of identity. Thus, the 5' region of these viruses seem to be a divergent part of their genome. Moreover, while base composition is quite similar when considering the full-length genomes of both viruses, a huge bias toward uridines in the first 300 nucleotides of DCV was not observed in the first 300 nucleotides of CrPV (**Figure 22B**). In conclusion, the 5' regions of DCV and CrPV represent a divergent point sequence-wise.

We then decided to compare our model of DCV domain I to the unpublished model of CrPV cloverleaf obtained by the team of Dr. Martin and mentioned in the introduction (**Figure 23**). This 2D structure was also obtained by performing DMS and CMCT chemical probing on two sizes of RNA encompassing either CrPV domain I alone or with the 5' IRES. It is quite obvious at a first glance that DCV and CrPV domain I models do not compare. Indeed, the domain I of CrPV is composed of a first short cloverleaf region similar to what has been described in polioviruses (1-70nt) and followed by 3 stem loops (81-178 / 182-223 / 224-263) as well as 2 linker regions (71-80 / 179-181). We can also notice a very accessible 5' extremity of CrPV cloverleaf with the first 8 nucleotides being non base-paired. Finally, the highly reactive nucleotides 264-300 represent the linker region between domain I and 5' IRES of CrPV. In comparison, DCV domain I is only composed of three stem loops of increasing sizes and does not present linker regions in-between. Moreover, no cloverleaf-like structure could be identified in this region of DCV.

Overall, the 2D structure models of DCV and CrPV 5' regions do not compare and could be one of the reasons for the differences in reads distribution observed in the small RNA

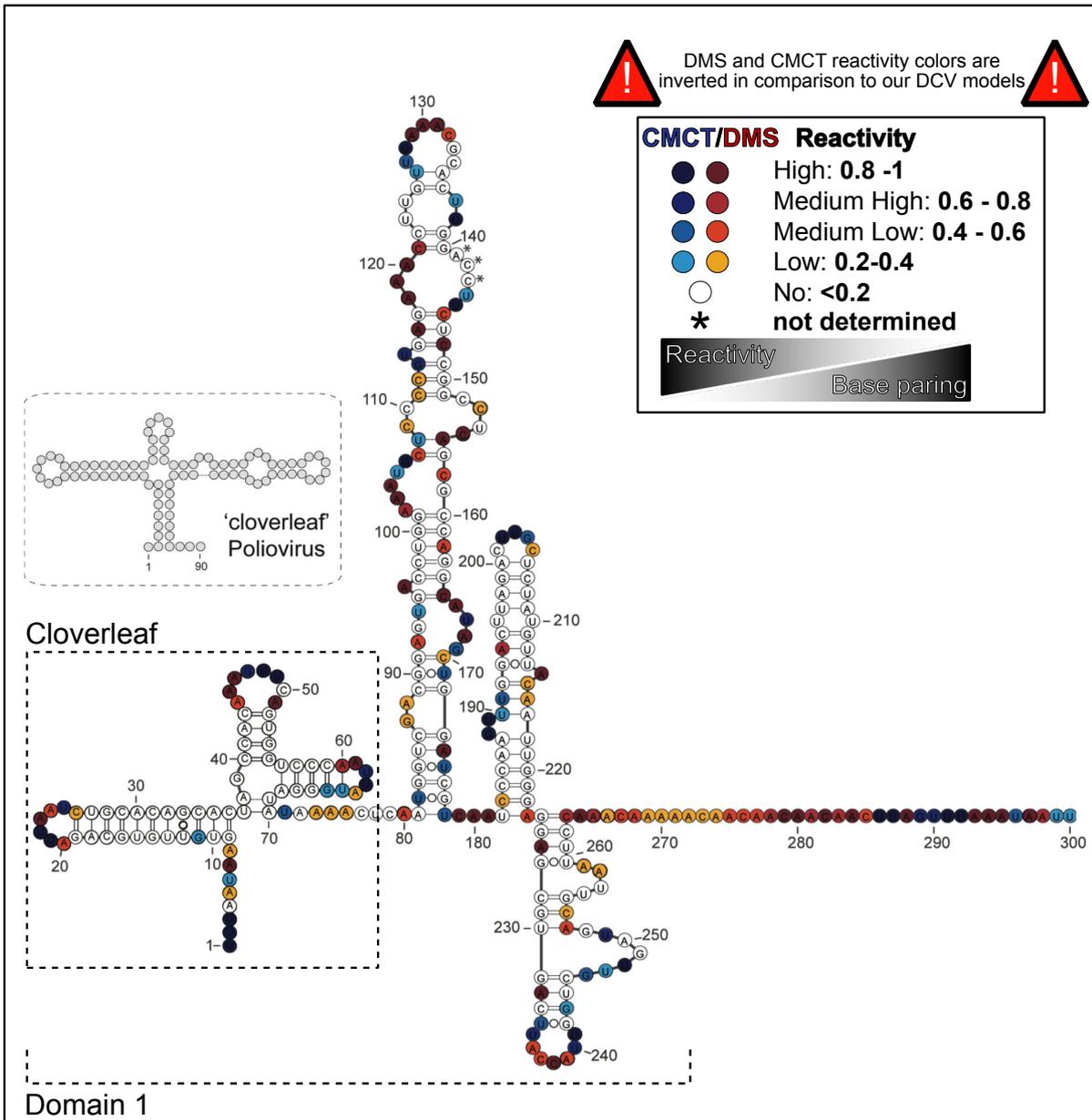


Figure 23: CrPV 5' region has a similar structure than poliovirus cloverleaf.

CrPV domain I structure model obtained by Dr. Franck Martin team. The domain I encompasses a cloverleaf structure similar to poliovirus cloverleaf and 3 stem loops.

HTS of infected S2 cells. Nonetheless, a precise entry point of Dicer-2 was identified in the dsRNA corresponding to this region for both viruses. The fact that this entry point is not located at the viruses 5' extremity is both expected, due to the presence of the protective VPg, and surprising as it suggests the creation of an alternative 5' extremity. That is why we decided to investigate the sensitivity to cleavage of DCV ssRNA 5' region in *in vitro* and *in vivo* conditions.

II. Characterization of DCV domain I sensitivity toward cleavage

The two complementary small RNA HTS methods that were used pointed toward an early entry point of Dicer-2 on the dsRNA corresponding to the 5' regions of DCV and CrPV. Interestingly, this entry point is not located at the 5' extremity but further downstream on the sequence (frame 4/2 for DCV and 5/3 for CrPV). However, from these data, we could not understand the precise mechanism by which Dicer-2 could have access to this internal point. That is why, we asked ourselves whether a new extremity could be created during the infection. Two main hypotheses were suggested to explain this:

- The RdRP could do a mistake during the replication. This hypothesis is supported by the fact that viruses can produce defective particles that lack part of their genome in the time frame of the infection. Moreover, the reverse transcription that was done for 2D structure determination of the DCV 5' region revealed some STOP sites independent of chemical modifications (*Figure 21A e³ B*). This depicts the fact that this RNA is highly structured and might be a difficult template for RNA synthesis in general, hence the generation of other 5' extremities.
- A cut could be happening in the ssRNA regions of the viral genome corresponding to the 5' region. Indeed, even if this region is highly structured, some regions including the loops are single stranded and could be targeted by endonucleases. These cuts would be mediated by a cellular protein and would leave a 5' extremity not protected by the VPg and thus, accessible to 5'-3' degradation for instance.

Both hypotheses require a concomitant action of the viral RdRP, which would reach the new 5' extremity, fall off and thus, generate a free unprotected dsRNA where Dicer-2 could enter. These hypotheses were investigated using diverse *in vitro* and *in vivo* methods.

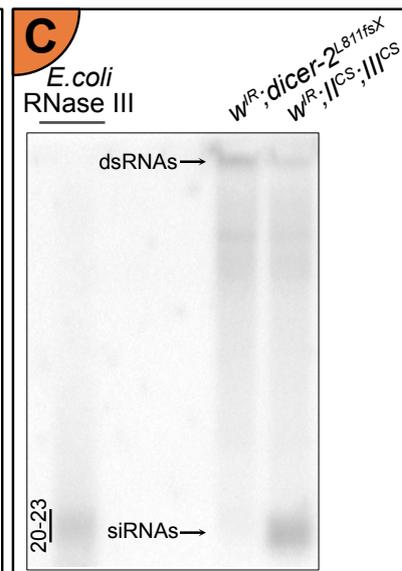
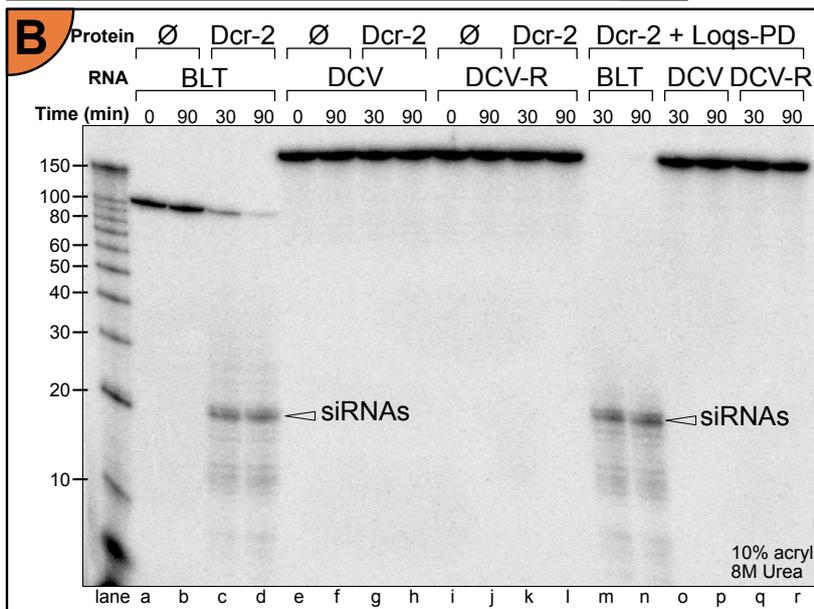
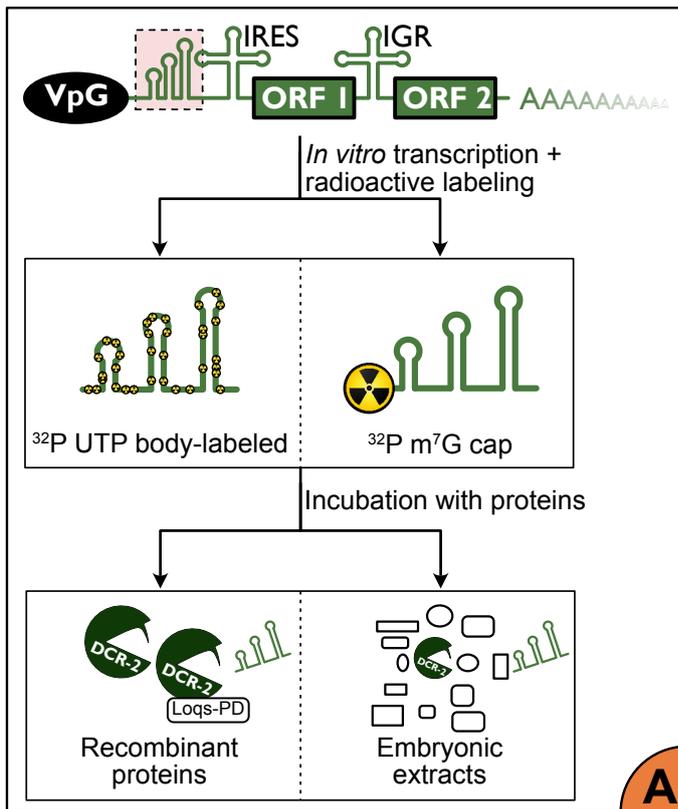


Figure 24 - part I: *In vitro* cleavage assays show that Dicer-2 is not able to cleave DCV domain I.

A) Schematic representation of the two approaches used to perform *in vitro* dicing assays. The RNA substrate was either *in vitro* transcribed using radioactive ^{32}P -UTP or subsequently capped with a radioactive m^7G . These RNAs are incubated with recombinant Dicer-2/Loqs-PD proteins or with a mix of proteins coming from embryonic extracts. **B&C)** Denaturing acrylamide gel used to visualize cleavage products of an *in vitro* cleavage assay using radioactive body-labeled RNA with **B)** recombinant Dicer-2 and Loqs-PD proteins; **C)** embryonic extracts from control or *dicer-2^{all}* mutant flies. Ø: buffer only condition; **BLT**: canonical blunt dsRNA; **DCV** and **DCV-R**: ssRNA domain I with and without heating refolding step respectively.

a. *In vitro* cleavage experiments reveal a Dicer-2 independent cutting pattern in DCV domain I ssRNA

Cleavage assay using Dicer-2 and Loqs-PD recombinant proteins

One of the candidates that was proposed as being able to mediate a cut in the ssRNA secondary structure of DCV domain I is Dicer-2 itself. This hypothesis is supported by the fact that Dicer-2 generates regulatory endo-siRNAs from ssRNA precursors with stem loop structures (Czech et al., 2008a; Okamura et al., 2008a). Moreover, *in vitro* and without its cofactor R2D2, Dicer-2 is able to cleave Dicer-1 substrates that are ssRNA hairpins: pre-miRNAs (Cenik et al., 2011). It is thus possible that Dicer-2 alone or with a specific cofactor might be able to perform such a cleavage in DCV domain I ssRNA.

In vitro dicing assays have been widely used in order to understand the molecular mechanisms of Dicer-2. Among other things, these experiments allowed us to decipher between different modes of action of Dicer-2 depending on the extremities and length of its dsRNA substrate but also on the presence or absence of its cofactors such as R2D2 and Loqs-PD. Even though *in vitro* dicing assays are far from recapitulating the *in vivo* complexity of a viral infection, we decided to use this method to test if Dicer-2 could be able to cleave the ssRNA secondary structure of DCV domain I. This study was performed in collaboration with the team of Pr. Brenda Bass (University of Utah - Salt Lake City). The principle of this experiment is to incubate the *in vitro* synthesized radioactively body-labeled DCV domain I ssRNA with purified Dicer-2 coupled or not to its cofactor Loqs-PD (**Figure 24A**). Because this region is very structured, two methods of RNA folding have been used to try to avoid misfolding of the RNA. After *in vitro* transcription, the RNA was either directly used or further heated and let to anneal at room temperature to ensure thermodynamic refolding. As positive control, a 152nt long blunt dsRNA was used to check for proper activity of Dicer-2 on a *bona fide* substrate. The incubations were performed, and the resulting cleavage products analyzed on an acrylamide gel.

First of all, no visible degradation of either control blunt dsRNA or DCV ssRNA can be observed over time when no proteins are added to the reaction (**Figure 24B - lanes a, b, e, f, i and j**) meaning that those substrates are stable and do not decay over time (0-90min). Then, we can see that the canonical blunt dsRNA substrate is efficiently diced in 21nt long RNAs by Dicer-2 alone (**Figure 24B - lanes c and d**) and that this cleavage is even

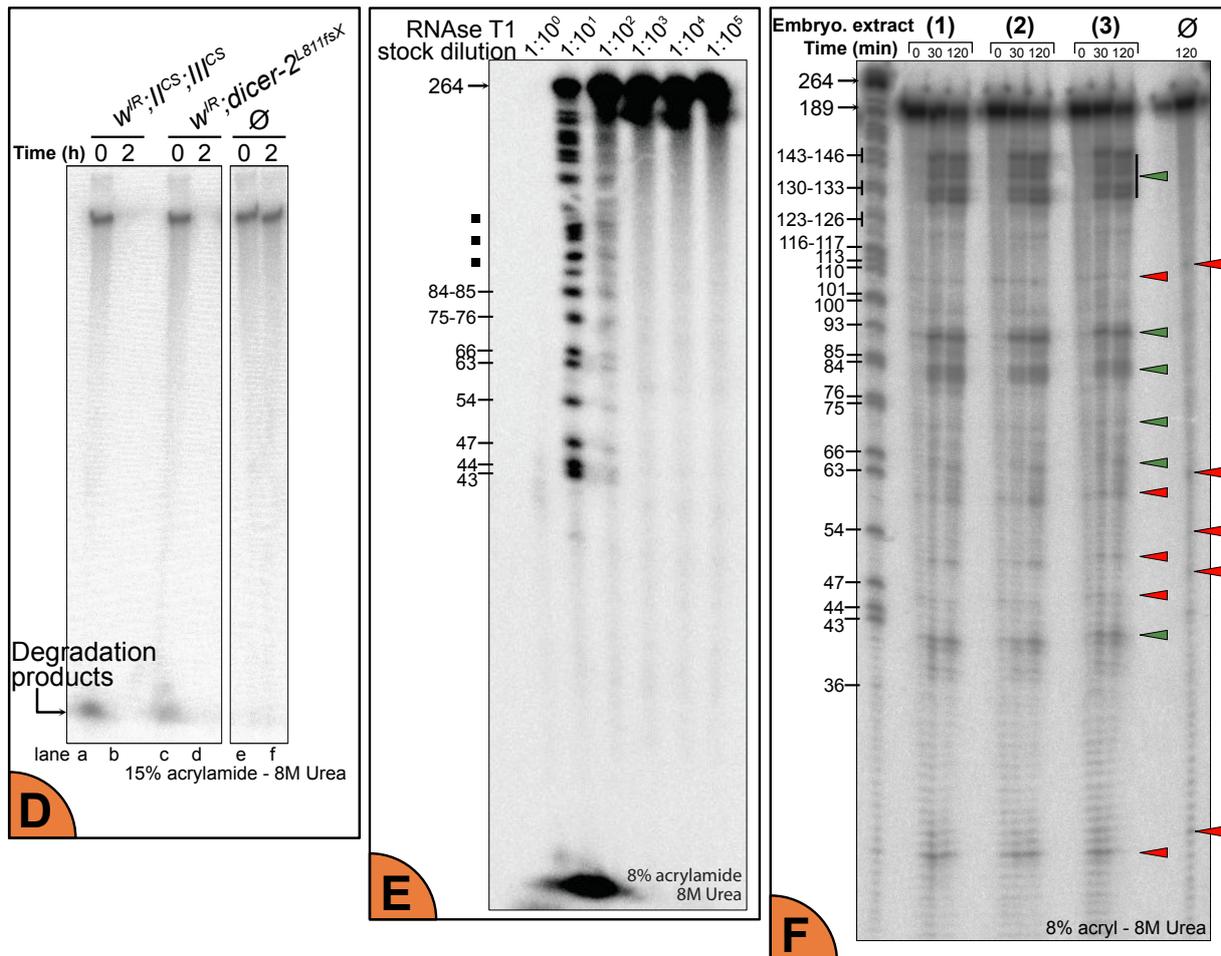


Figure 24 - part II: *In vitro* cleavage assays show that Dicer-2 is not able to cleave DCV domain I.

D) Denaturing acrylamide gel used to visualize cleavage products of the incubation of a 106nt blunt dsRNA with *E.coli* RNaseIII or embryonic extracts. **E)** Denaturing acrylamide gel to determine the best working quantity of RNase T1 to generate a T1 ladder. The initial quantity ($1:10^0$) of T1 enzyme is 1.10^{-3} Unit. Radioactive and capped RNA was incubated for 15' with the T1 enzyme dilutions. **F)** Representative denaturing acrylamide gel to visualize cleavage products of an *in vitro* dicing assay using embryonic extracts. **(1)** $w^R; III^{CS}; III^{CS}$; **(2)** $w^R; dicer-2^{L811fsX}; GFP::Dicer-2^{WT}$ **(3)** $w^R; dicer-2^{L811fsX}$. Red arrows (resp. green arrows) represent embryonic extracts-independent (-dependent) cleavage products of the assay.

more efficient when Loqs-PD is added to the reaction (*Figure 24B - lanes m and n*) as previously described in the literature (Sinha et al., 2015). However, no cleavage can be observed in the ssRNA domain I of DCV when incubated with Dicer-2 alone or with its cofactor and this, regardless of the RNA folding method used (*Figure 24B - lanes g, b, k, l, o – r*). Thus, in these conditions, Dicer-2 itself or in association with Loqs-PD is not able to perform a cut in the ssRNA secondary structure of DCV domain I.

Cleavage assays using embryonic extracts

In order to get closer to *in vivo* conditions, we decided to use embryonic extracts obtained from embryos with different genetic backgrounds instead of recombinant proteins (*Figure 24A*). In addition to Dicer-2 and its known cofactors, these mixes of proteins could contain an unknown cofactor of Dicer-2 or an unrelated ribonuclease able to sense and cleave the domain I structure of DCV. For these experiments, we used different embryos with the following genetic backgrounds:

Genotype	Simplified name
$w^{IR}; II^{CS}; III^{CS}$	$dicer-2^{WT}$
$w^{IR}; dicer-2^{L811fsX} / dicer-2^{L811fsX}$	$dicer-2^{null}$
$w^{IR}; dicer-2^{L811fsX} / dicer-2^{L811fsX}; GFP::Dicer-2^{WT}$	$GFP::Dicer-2^{WT}$

(CS: CantonS)

The proper activity of Dicer-2 in these embryonic extracts was checked by incubating them with a canonical blunt dsRNA (*Figure 24C*). Dicer-2-dependent products could be observed at the same size as *E. coli* RNaseIII digested dsRNA (20-23nt). From this experiment, we can conclude that the embryonic extracts are active and that Dicer-2 is still able to work on its canonical dsRNA substrate.

Then, we incubated these embryonic extracts with the same DCV domain I radioactively body-labeled ssRNA as the previous cleavage assay (*Figure 24D*). No degradation of the ssRNA could be observed when no embryonic extract was added to the reaction (*Figure 24D - lane f*). However, a complete degradation of the ssRNA could be observed both in $dicer-2^{WT}$ and $dicer-2^{null}$ conditions (*Figure 24D - lane b and d*). The lack of a cutting pattern as well as the accumulation of what seems to be a single nucleotide band at the bottom of the gel after 2h of incubation led us to think that this unprotected radioactive viral RNA was in fact the target of the very efficient 5' – 3' exonuclease activity of Pacman. Thus, we

decided to protect the synthesized viral RNA with a radioactive methyl-guanosine cap to mimic the 5' protective covalently-linked VPg of the virus (*Figure 24A*).

In order to be able to identify cuts at a single nucleotide resolution, a T1 ladder was created. Basically, the *in vitro* synthesized and capped radioactive RNA is incubated with the ribonuclease T1 that cuts ssRNAs 3' of guanine residues. The RNA used for this incubation was previously denatured to open the stem loop structures that would be protected from the enzyme. The optimal quantity of the T1 enzyme to be used for this has been experimentally determined by incubating various amount of the enzyme with the RNA and observing the cutting pattern (*Figure 24E*). While a condition with 1.10^{-5} Unit of T1 RNase seemed too strong (all the ssRNA molecules are entirely processed after 15min, *Figure 24E - lane a*), a 1.10^{-5} Unit or lower quantity is too weak to observe a clear cutting pattern (*Figure 24E - lanes c - f*). Thus, we chose to use 1.10^{-4} U of the enzyme to generate a reliable ladder for this *in vitro* dicing assay (*Figure 24E - lane b*).

Different embryonic extracts were incubated for various lengths of time with the radioactive and capped DCV ssRNA domain I and the resulting cutting products were visualized on a denaturing acrylamide gel (*Figure 24F*). As a general remark, it has to be noted that the technique was sensitive enough to have a single nucleotide resolution at some parts of the gel depending on the migration time and on the acrylamide percentage of the gel. Three different sizes of DCV domain I ssRNA (264, 189 and 120nt long) were incubated with the embryonic extracts. It has to be noted that these sizes were chosen based on predictions of DCV domain I ssRNA structure prior to its chemical probing. The same cutting pattern could be observed in the part of the sequence common to these RNAs. That is why only the results obtained using one of these RNAs (189nt long) are presented. The T1 ladder, however, was generated from the 264nt long DCV domain I RNA. The cutting pattern showed here is representative of several experiments using different migration time, acrylamide percentage of the gel and sizes of ssRNA DCV domain I substrate.

A control condition without embryonic extracts shows a few products smaller than full-size ssRNA (*Figure 24F - red arrows*). These products are also present at the 0min time point of every tested condition and probably depict a certain RNA instability upon storage but are not cleavage products. On the contrary, products identified with a green arrow are absent (or very weakly present) at 0min and increase over time suggesting that they

are embryonic extracts-dependent cleavage products (*Figure 24F – green arrows*). However, none of these cleavage products appear to be Dicer-2 dependent as the same cutting pattern can be observed in *dicer-2^{null}* embryonic extracts and in extracts containing a wild-type version of Dicer-2.

Still, these Dicer-2-independent products were consistently found between experiments and could reveal weak points of *in vivo* relevance in the DCV domain I structure. That is why we decided to check for the localization of these cuts on the 2D structure obtained by chemical probing and in relation with the theoretical frame 4 reads position (*Figure 25*). Interestingly, most of the cuts happened inside or in a very close proximity to a bulgy region of one of the stem loops. This information by itself already supports the DCV domain I model that has been drawn from the chemical probing experiment. Indeed, ssRNA structures are much more likely to be targeted by nucleases than dsRNA regions. Then, we noticed that most of the cuts (16/20) happened 5' or 3' of uridines. This last information could lead us to hypothesize the implication of a uridine specific endonuclease. However, as previously explained, the 5' region of DCV base composition shows a strong bias with uridines representing 45% of the nucleotides (*Figure 22B – bottom left*). The correlation between the observed cutting pattern and the U specific localization is thus most likely due to this bias. Finally, no identified cut could be directly correlated to a siRNA in frame 4 position.

To investigate if these cuts in the bulgy regions of DCV domain I can be detected *in vivo*, we decided to try to detect the genomic 5' extremities of DCV that appear in infected conditions.

b. Identification of *in vivo* relevant DCV genome 5' extremities by RACE

In order to identify those extremities, we decided to use a technic historically used to determine 5' extremities of mRNAs which is the 5' Rapid Amplification of cDNA Ends (5' RACE).

Experimental proof of concept

The principle of this method is explained in *Figure 26*. As a proof of concept, we wanted to test whether we could detect the full-length viral RNA starting from the stock of DCV.

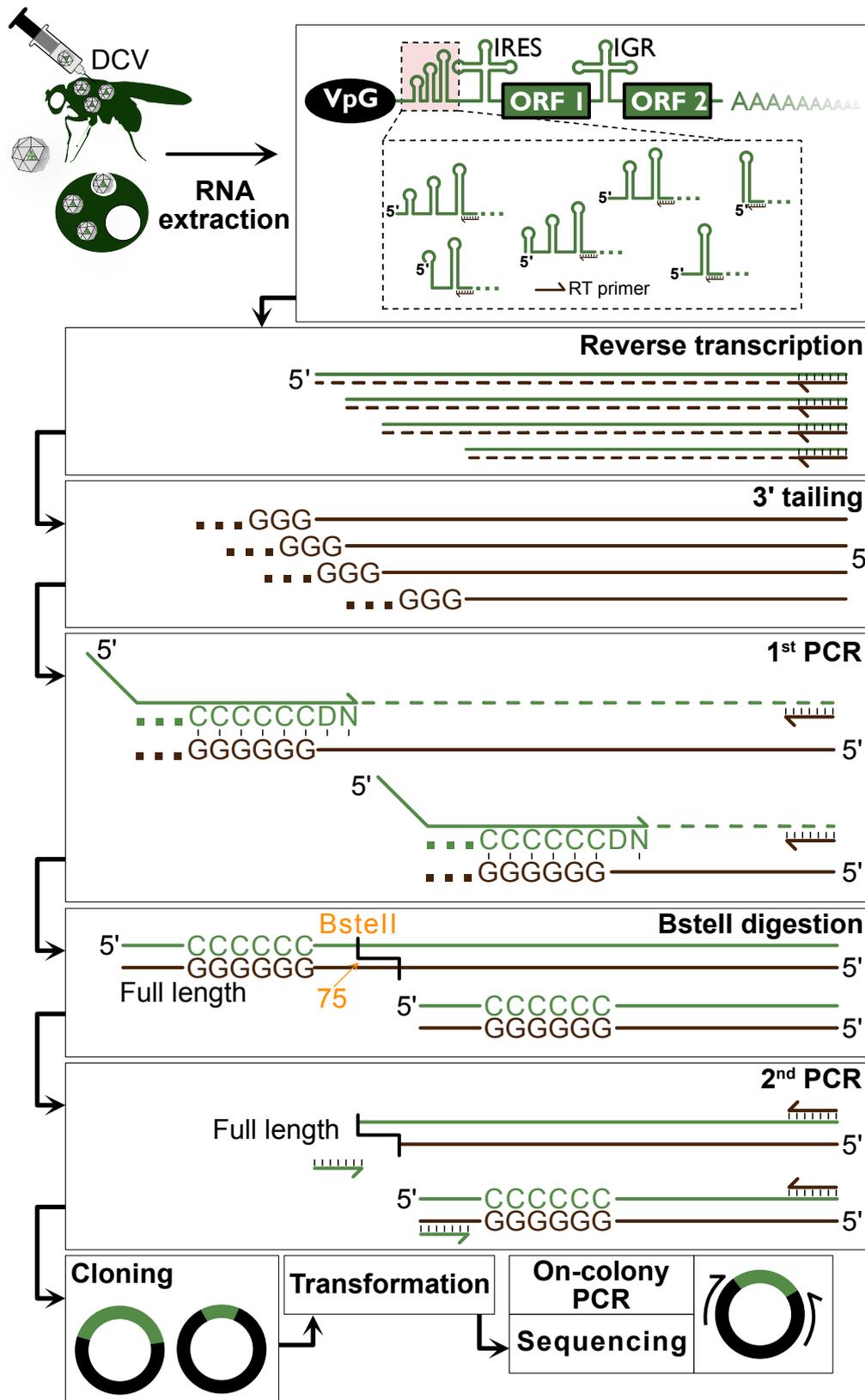


Figure 26: Schematic representation of the 5' RACE method.

Briefly, RNAs are extracted from a DCV infected sample. These RNAs can either be full length or being shorter. A reverse transcription specific to the 5' region of DCV will allow us to generate cDNAs of various length depending on the original size of that RNA. Guanine residues are added at the 3' of those cDNAs using a terminal transferase. A first PCR using anchored primers (D=A, T or G; N=A, T, G or C) will allow an amplification of these cDNAs. Then, a BstI restriction site located in the very 5' of the domain I sequence will specifically cut the full-length products that are of lesser interest for us. Finally, a second PCR will amplify the surviving molecules only and the resulting products are cloned into a pJET vector that will be sequenced after amplification in DH5alpha bacteria.

Thus, RNA extraction was performed from the purified stock of DCV used for the infections. The first three steps of the method (reverse transcription, 3' tailing and PCR1) were performed (*Figure 27A*). The agarose gel run after the first PCR of the protocol showed us that only the condition where reverse transcriptase as well as terminal transferase enzymes were added presents an amplification band corresponding to the size of the full-length DCV domain I (*Figure 27A - lane a*). Also, no nonspecific amplifications could be observed in the other control conditions, suggesting that the obtained products are dependent of reverse transcriptase and terminal transferase activities.

The final aim of this RACE experiment is to be able to detect non-full-length products that could reflect internal entry sites of Dicer-2. Thus, we needed a way to get rid of full-length reads that will most likely be over abundant. To do so, a specific step of BstEII digestion was added to the protocol designed to cut full length products. This restriction site is the most 5' located and unique site we could find in the domain I sequence. However, its restriction site is 75bp away from the 5' extremity of the DCV genome. Thus, if new genomic 5' extremities are located between nucleotides 1 and 75, they will also be subject to BstEII digestion and will not be detected using this method (*Figure 26*).

We tested this digestion step on the PCR1 obtained starting from DCV stock. We did not expect to see any surviving products as all RNAs in viral stock should be full-length. BstEII digestion of PCR1 products for 15' shows the expected restriction profile with apparently no full-length products remaining (*Figure 27B*). We performed PCR2 on both digested and non-digested products as a control. Several digestion times and number of PCR cycles were tested. As expected, when no BstEII digestion was performed on the PCR1 products, only a major band corresponding to the full-length domain I could be seen (*Figure 27C - lane a*). However, this band could also be observed in the BstEII digested conditions, regardless of the time of digestion or the number of second PCR cycles (*Figure 27C - lanes b - f*). This puzzling observation might suggest that either an entire digestion of full-length products is hard to reach in our conditions or that single nucleotides modifications in the sequence of a subset of DCV particles disrupt the BstEII restriction site. In the end, we were not able to understand where these full products could come from and how could they be so predominant.

In order to check for the identity of the amplified products, we randomly cloned them (products in black boxes – *Figure 27C*) in a pJET vector and transformed the ligation in

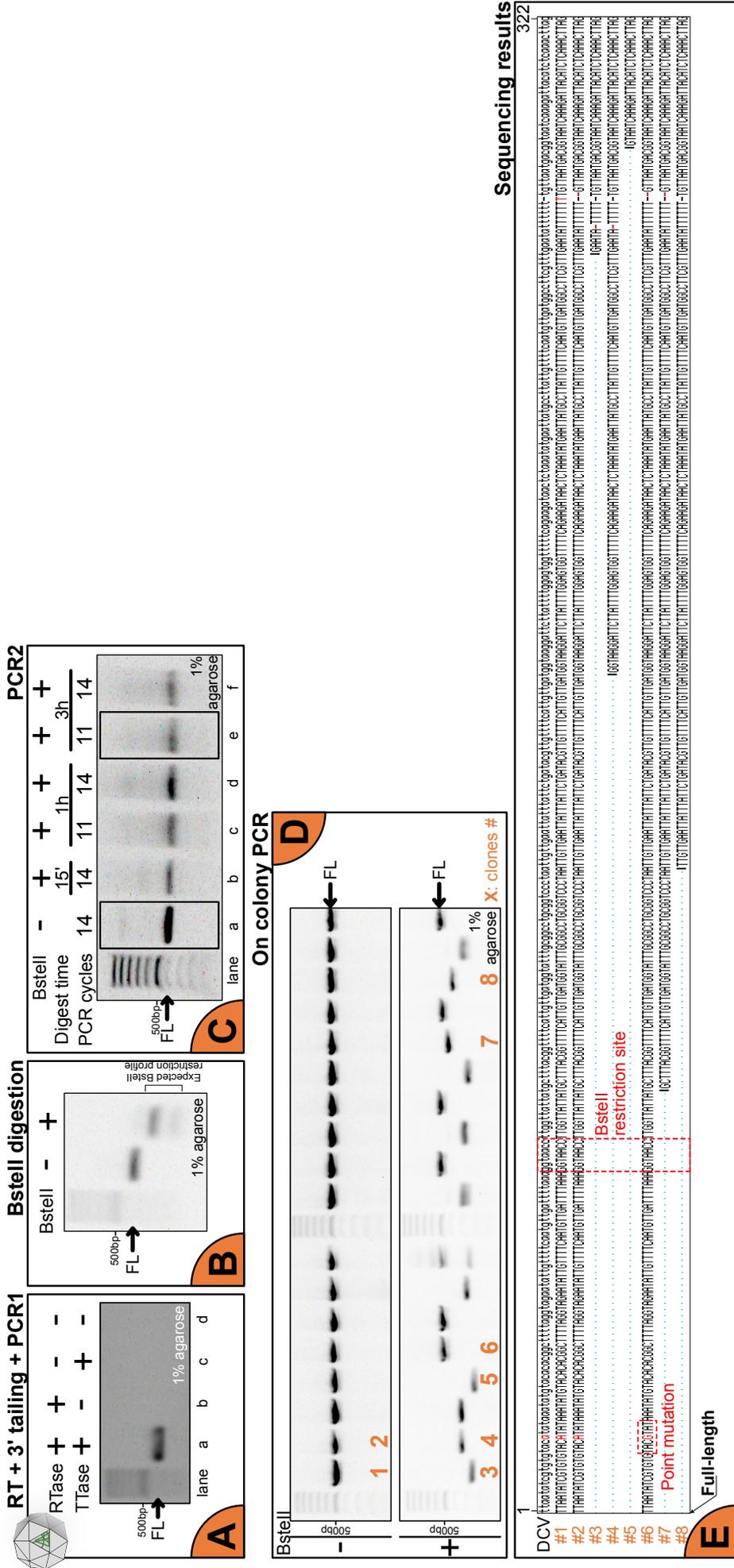


Figure 27: Proof of concept of 5' Rapid Amplification of cDNA Ends to determine viral 5' extremities.

A) Agarose gel showing the products of PCR1. Four experimental conditions were tested including presence/absence of reverse transcriptase/terminal transferase enzymes in the reactions. **B)** Agarose gel showing digestion of PCR products for 15min by BstI. **C)** Agarose gel showing the products of PCR2. Different incubation times with the restriction enzyme as well as various numbers of PCR cycles were tested. The black boxes show the conditions that were used to continue the experiment. **D)** Agarose gel showing on colony PCR products. **Black arrows** show the expected size for a full-length (FL) product. It has to be noted that this size varies during the steps of the experiment as adapters are gradually added at both extremities of the products. **E)** Alignment of the sequences obtained after sequencing of clones #1 to #8.

DH5 α bacteria. By doing so, we expect to see an incorporation of approximately all the different sizes of RACE products. Growing colonies were randomly picked to perform on colony PCR (*Figure 27D*) and sequencing (*Figure 27E*) of the transformed plasmids. In the non BstEII digested condition, all the colonies picked for on colony PCR showed a band corresponding to the full-length insert (*Figure 27D-top*). When looking at the BstEII digested condition, we could observe a mix of what appeared to be full-length inserts (such as clone #6) and other inserts with a high diversity of sizes (*Figure 27D – bottom – clones #3-5 & 7-8*). No on-colony PCR products higher than the expected full-size insert could be observed.

Sequencing of the 8 clones annotated on *Figure 27D* all yielded DCV sequences (*Figure 27E*). Thus, the shorter products identified by on-colony PCR are not unrelated contaminants of the RACE experiment. Still, it was unexpected to witness such products in this control condition with only virus stock. Two hypotheses were raised to explain these shorter products:

- they could be coming from errors in the reverse transcriptase used to generate the cDNAs. Indeed, because this region is highly structured, we could be observing event where the enzyme got stuck and fell off its substrate.
- they could reflect errors of the viral RdRP. Indeed, it was previously shown that during replication of some viruses, especially at high MOI, defective particles with missing parts of the genome can be detected. Thus, the DCV stock used to do this experiment could contain such defective particles.

Then, clones #1, #2 and #6 could be identified as containing full-length inserts (*Figure 27E*). This shows that (1) without BstEII digestion, the main viral RNA that we can find in the DCV stock is full-length and (2) some full-length products are resistant to BstEII digestion. The resistance of these products to BstEII digestion cannot be explained by disruption of the restriction site as shown by sequencing of clone #6 in *Figure 27E*.

These results together prove that the method is sensitive enough to detect full-length DCV genome. However, the presence of what appears to be non-full-length genomes already in the DCV stock will increase the difficulty to discriminate the 5' extremities generated during the infection. We still decided to give it a try using *in vivo* samples with the hope that a significant biological cut would be strong enough to be enriched in the final data.

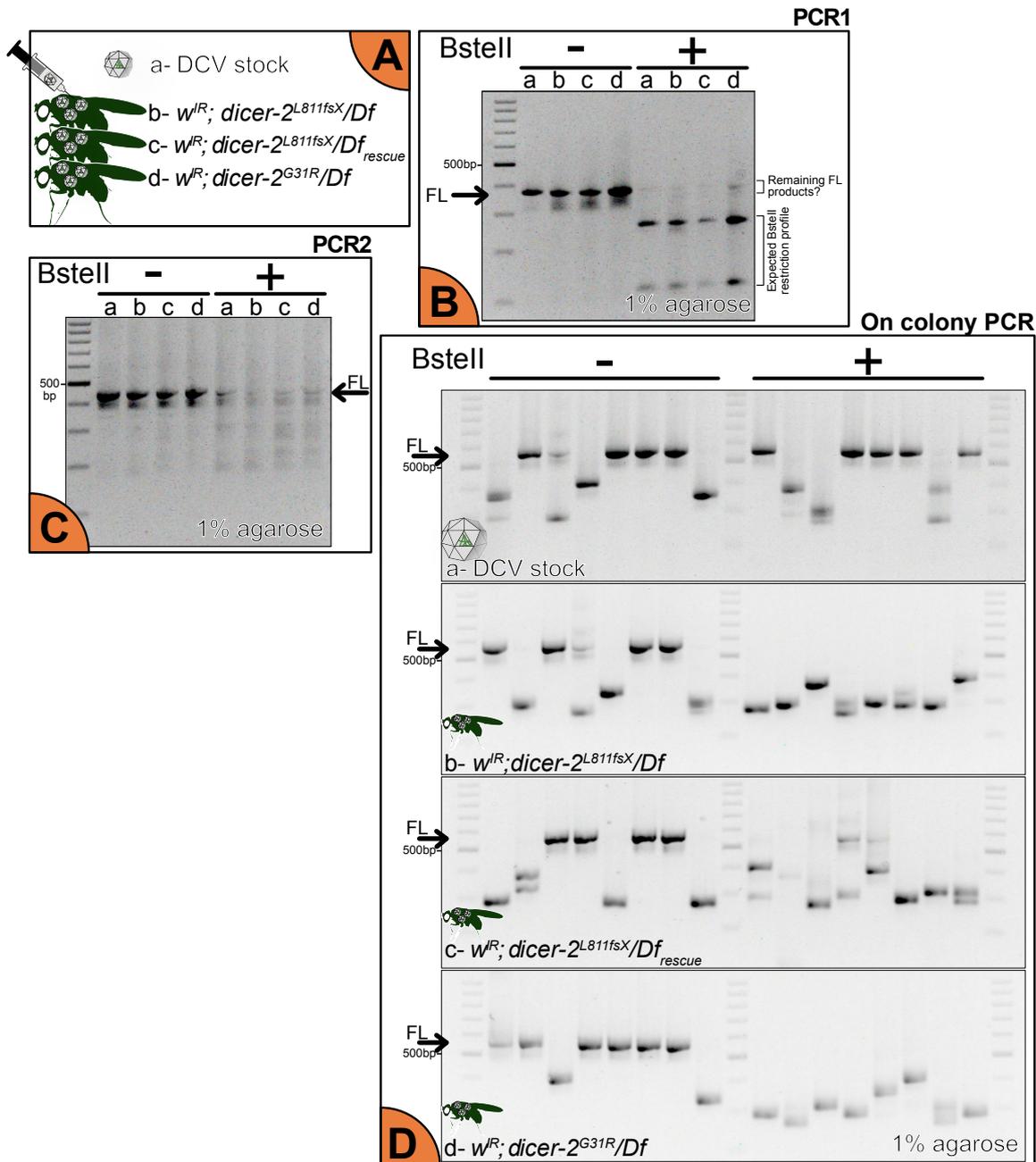


Figure 28: 5' RACE experiment on DCV infected flies.

A) Schematic representation of the samples tested in this experiment. RNA was extracted from either a pure DCV stock or DCV infected flies of the indicated genotype (500PFU - 3d pi). **B & C)** Agarose gel showing the PCR1 and PCR2 products with and without BstI. The BstI digestion condition (3h) as well as the number of cycles for PCR2 (11X) were kept similar in this experiment. **D)** Agarose gel showing on colony PCR products. **Black arrows** show the expected size for a full-length product. It has to be noted that this size varies during the steps of the experiment as adapters are gradually added at both extremities of the products.

Determination of DCV 5' extremities by RACE in infected flies

Once the proof of concept of the 5' RACE was done, we decided to try to detect the 5' extremities of DCV genomic RNA that we could find in an *in vivo* condition. Thus, flies of different genotypes were infected with DCV, their total RNAs were extracted 3d pi and the viral stock was used as internal control (*Figure 28A*). The experiment was conducted as before until PCR2 and revealed no differences between flies and viral stock samples nor between the different fly genotypes (*Figure 28B e3 C*).

Still, the experiment was continued and PCR2 products cloned and transformed in bacteria. On colony PCR results are shown in *Figure 28D*. When looking at the DCV stock control, we could not observe the clear-cut result of the proof of concept experiment (*Figure 27*). Indeed, not all the clones from the non-digested condition were full-length and a high proportion of the clones from the digested condition were full-length. Furthermore, the *in vivo* samples result also shows a high variability in sizes of on colony PCR products. It has to be noted that this experiment is representative of at least 8 independent experiments with slight protocol modifications in each. Thus, a wide variety of starting material was used (DCV stock, infected flies and infected S2 cells), the protocol has undergone many optimizations and dozens of sequencing were done but no significant pattern or enrichment in a specific shortened RNA could ever be detected.

From all these experiments, we can conclude that even though the method is working (we are able to detect full-length viral RNA), its specificity is not high enough to detect subtle events as the one we are looking for. Indeed, the method is subject to a lot of background noise from different origins:

- The presence of shorter products in the DCV stock controls suggests either the presence of viral RdRP-dependent defective particles or of a reverse transcriptase-dependent bias of the experiment.
- The small RNA HTS performed in flies raised the evidence for viral degradation products in the specific case of DCV. The origin of these products was not determined but could be interfering with the 5' RACE output.
- Antiviral RNAi machinery is active. Indeed, more than the initial entry of Dicer-2 on its substrate, it is the entire RNAi machinery that fights against the virus and generates cleavage in its genome and antigenome.

That is why it has been decided to drop this technic that, in the state of the art, was not sensitive enough to detect what we were looking for and would need a high throughput implementation to maybe have the chance to observe patterns emerging from the noise. Other technics that could have been used will be discussed later.

In vitro cleavage assays did not identify Dicer-2 as a good candidate to perform the cuts observed when incubating DCV domain I RNA with embryonic extracts. Thus, nucleases present in the mix of proteins must be responsible for these cuts. We hypothesized that if these cuts can be observed *in vitro*, they might have *in vivo* relevance and could explain the internal entry site of Dicer-2 on viral dsRNA. That is why we decided to perform a screen of candidate proteins that could be responsible for these cuts.

III. RNAi-based screen in S2 cells to identify new antiviral and proviral proteins

a. Experimental setup of the screen

Screen design

The rationale for this screen lies in the description of a precise and internal entry point of Dicer-2 on the viral dsRNAs of DCV and CrPV (see Chapter I). One of the possible explanations for this behavior includes the action of an enzyme able to sense and cleave the domain I ssRNA structure of these viruses. This newly generated 5' extremity would eventually be made double stranded by the action of the viral RdRP and would become a perfect template for Dicer-2 to enter.

However, we lack biochemical methods sensitive enough to detect the precise entry point of Dicer-2 on its substrate and that would be scalable to a high throughput screen. We hypothesized that the lack of the domain I sensing enzyme should have a strong impact on the host defense as Dicer-2 would not be able to access the viral dsRNA anymore. That is why, the output that was chosen for this screen was simply the viral load measured by RT-qPCR. We decided to conduct this screen in S2 cells for three reasons:

List of candidates								
Adar	CG2972	CG6724	DNApol-e255	mei-9	PGAP1	Sid	Xrcc2	CG2910
AIMP2	CG2990	CG6744	DNasell	mei-W68	PNPase	slx1	ZC3H3	CG3496
Arc1	CG30343	CG8038	Drep4	mldr	Pop2	Smg5	zuc	CG7358
Asph	CG31812	CG8343	drosha	mle	r	Smg6	DRP1	CG6745
aub	CG3308	CG8360	Edc3	mms4	r2d2	Snm1	Drice	CG3045
beg	CG33082	CG8366	egl	mre11	Rad1	Snp	Tao	CG3709
bsf	CG33260	CG8475	eIF2Be	mRpl44	rad50	spn-A	ZN72D	CG7849
Cand1	CG3358	CG9125	EndoG	mus201	Rad51D	spn-B	CG3800	CG34140
CG10214	CG34317	CG9272	Ercc1	mus312	Rad9	spn-D	lost	CG3333
CG10943	CG42389	Cklalpha	FASN3	mus81	Rat1	SRPK	Hel25E	CG4159
CG11269	CG42668	Clp	Fen1	Nbr	Rga	stau	fandango	CG31719
CG11486	CG44242	DCP1	Gen	Nnp-1	RNASEK	Stip1	tailor	CG10692
CG13690	CG4611	Dcr-1	hdm	Non3	RNaseX25	stnB	blanks	
CG14057	CG4825	Dcr-2	lre1	Not1	rmh1	tam	caz	
CG15526	CG4847	Ddx1	Jhl-1	Not3	Rpp20	tos	CG2199	
CG15784	CG5285	DIP1	l(1)G0020	Ns1	Rpp30	Trax	vir-1	Controls
CG16790	CG5316	Dis3	l(1)G0045	Ogg1	RpS3	trsn	CG4947	Ago2
CG17514	CG5626	Dis3l2	Larp4B	Pabp2	Rrp1	Tudor-SN	CG6315	Rack1
CG2145	CG5641	DNApol-a180	ldbr	PAN2	Rrp6	twin	CG5933	LacZ
CG2246	CG6171	DNApol-d	loqs	pcm	RtcB	WRNexo	CG7818	Diap1

Table 4: List of the candidates used for the RNAi-based screen in S2 cells.

Plate A layout												
<>	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		LacZ	CG3709	Tgt	Rpp30	Hop	Thread	Drp1	Mett14	CG8475	Rat1	
C		Cand1	aub	zuc	bsf	AGO2	tos	Nbr	CG3045	CNBP	LacZ	
D		rad50	Nop60B	Thread	DCP1	CG3358	CG8343	nito	empty	drosha	PAN2	
E		Not1	EndoG	Arc1	Ercc1	SRPK	CG4847	stau	Dis3l2	Tudor-SN	CG2199	
F		CG11486	empty	Gen	Rack1	ldbr	RNaseX25	Pop2	AIMP2	Thread	Zn72D	
G		beg	Rad9	Ddx1	mRpl44	Snm1	CG5641	CG6171	Trax	DNasell	Non3	
H												

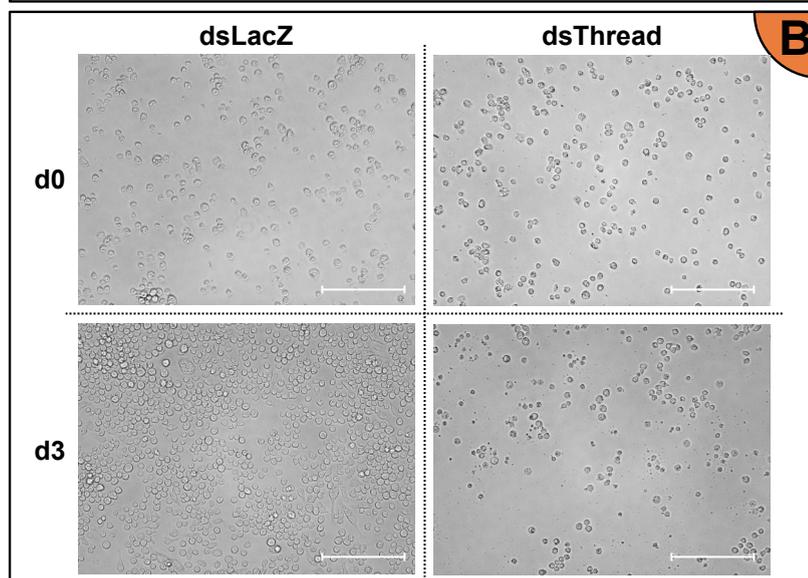


Figure 29: Experimental setup of the RNAi-based screen in S2 cells.

A) Example of disposition of samples and controls on one of the six plates obtained from Harvard Medical School (plate A). Positions of the controls are indicated in orange. Grey wells were filled with PBS to prevent evaporation effects. **B)** Inverted microscope images depicting dsThread effect on S2 cells. **Top:** d0; **Bottom:** d3 post soaking; **Left:** S2 cells soaked 1d in dsLacZ; **Right:** S2 cells soaked 1d in dsThread. Scale bar: 100um.

- The kinetic study using small RNA HTS of DCV infected S2 cells revealed similar profiles to small RNA HTS of DCV infected adult flies. Thus, all the components required for the entry of Dicer-2 at a specific position are present in S2 cells.
- We can easily knock down candidate genes expression using the RNAi machinery of the cells by soaking them in corresponding long dsRNAs
- In term of handling procedures, S2 cells is a model much easier to implement in a high throughput screen compared to flies (Mohr, 2014; Perrimon and Mathey-Prevot, 2007).

Thus, the screen we setup to try to identify a protein able to sense and cleave DCV domain I ssRNA is a RNAi-based screen in DCV infected S2 cells with DCV viral load as output.

Selection of the candidates

The selection of the candidates was made by using data mining resources as well as in-house experimental results. We first searched for proteins described as “nuclease” in the Drosophila database Flybase. It was decided not to restrict the research for ribonucleases because for some of these proteins, the attribution of GO terms is solely based on domain homology. Thus, an initially identified DNA nuclease could very well be able to cleave ssRNA for instance. 111 candidates could be identified using this methodology. Then, an in-house protein-protein interactome of Dicer-2 in DCV infected condition was performed by another Ph.D student in the lab (Claire Rousseau) and yielded 53 candidate proteins. Also, the team of Dr. Franck Martin generated RNA-protein interactomes of the cloverleaf and 5’IRES regions of CrPV. Because we have evidences that suggest a similar mode of action of Dicer-2 on DCV and CrPV, 4 candidates preferentially enriched on CrPV cloverleaf were picked. Finally, 20 candidates were cherry picked from literature and added to the screen list. Based on all this, a list of 188 candidates was established.

Due to the high number of dsRNAs to design and produce, it has been decided to outsource this part to the Drosophila RNAi Screening Center at Harvard Medical School. Among the original list of 188 candidates, they were able to select at least one suitable dsRNA for 172 of them (*Table 4*). The dsRNAs were selected according to their previously tested efficiency of knock-down and their ability to target all the isoforms of the selected gene. When possible, two different dsRNAs were selected. Control conditions were also added to the experimental setup:

- **dsLacZ**: this dsRNA does not have a target in the cells and shows the impact of the activation of the RNAi machinery on its own.
- **dsAgo2**: Ago2 is the final effector protein of the RNAi pathway. Thus, knock down of *ago2* should result in an increased viral load.
- **dsRack1**: Rack1 was previously identified as a dicistrovirus proviral protein (Majzoub et al., 2014). Knock down of *rack1* should result in a decreased viral load.
- **dsThread**: this dsRNA targets drosophila Diap1, a known anti-apoptotic protein. This condition is used as visual output for the efficiency of RNAi knock down. Indeed, wells containing this dsRNA should show a high percentage of cell death.
- **Empty**: a condition without dsRNA at all that should not have an impact on viral load.

The resulting 305 dsRNAs were delivered in 6 plates with identical layout regarding candidates/controls disposition (*Figure 29A*). This random distribution of the candidates dsRNAs was kept to perform the screen. Briefly, cells were soaked for 3 days with long dsRNAs mapping candidate genes and subsequently infected with DCV at a MOI of 0.01. DCV viral load was checked by RT-qPCR 20h post infection.

b. Identification of pro and antiviral proteins

First, **dsThread** controls showed a high amount of cell death 3 days post soaking, which indicates an efficient knock down of *diap1* (*Figure 29B*). From this, it was inferred that the knock down of candidate genes was efficient. However, as it will be discussed later, proper decrease of interesting candidate genes mRNA and protein levels will need to be checked to validate them.

The analysis of the data was done taking into account the row, column and plate effects that could be statistically identified and modeled (details in Materials and methods). The resulting global distribution of the data is shown in *Figure 30*. From this plot, we can see that the **dsLacZ**, **dsRack1** and **empty** controls behave as expected:

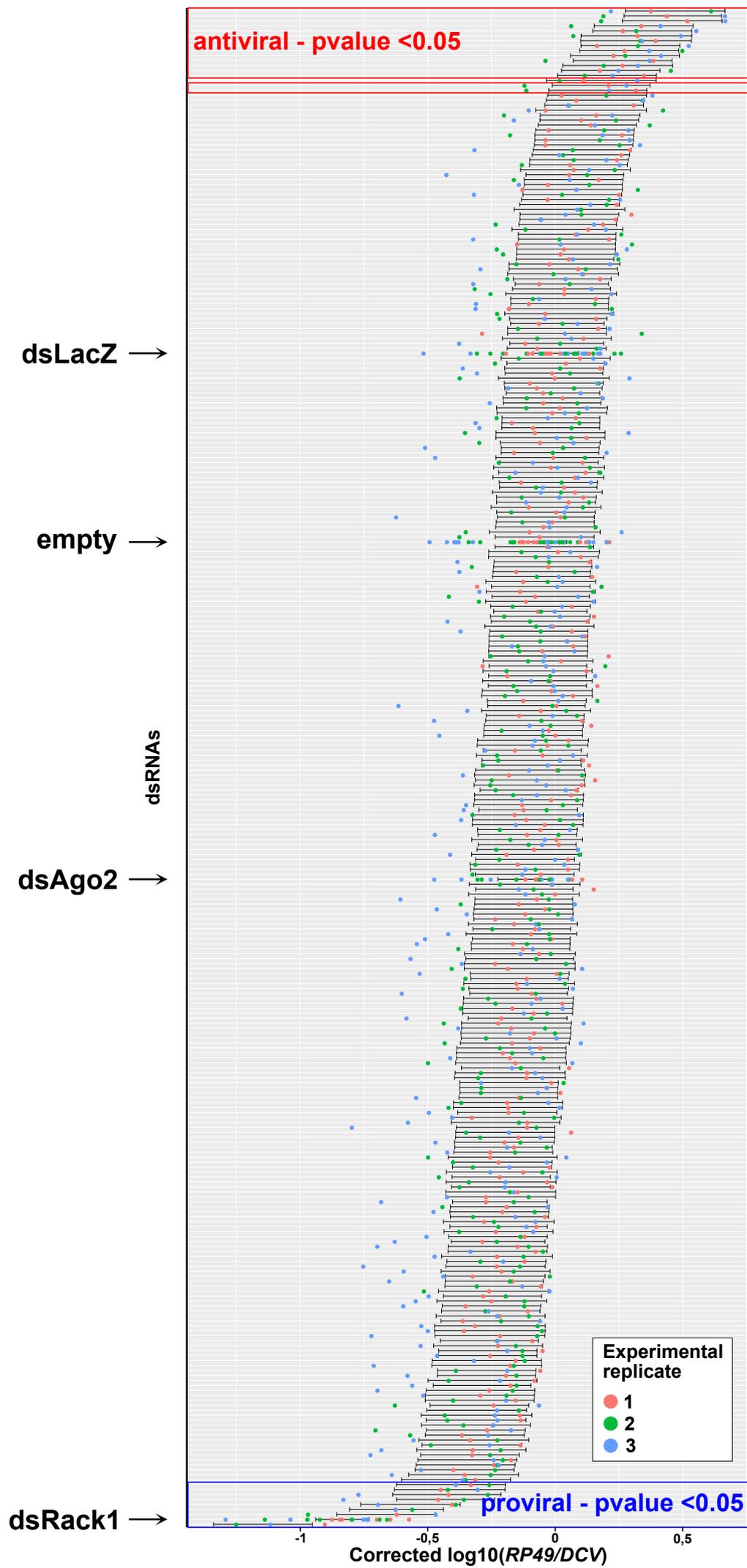


Figure 30: Global distribution of the screen results.

For the sake of visibility, only control dsRNAs were annotated on this plot. The values displayed are the values corrected by the statistical model taking into account the row, column and plate effects identified. Error bars represent the 95% confidence interval estimated by the model. Each point is an unique RT-qPCR value. Statistical test: one-way ANOVA against the modeled effects.

- **dsLacZ** and **empty** controls are close together in the middle of the dataset suggesting that activation of the RNAi machinery by itself does not have a significant impact on DCV viral load
- **dsRack1** shows a statistically significant decrease in the viral load of DCV infected cells. Its effect is ranked at position #2 in the pro viral dsRNAs which highlights its importance for the DCV viral replication.

Only **dsAgo2** does not show the expected increased viral load phenotype. This can be explained by the fact that in this case, we are using the RNAi machinery against itself. Thus, at the time of the infection, the knock down of *ago2* is not efficient anymore which results in a normal amount of *ago2* mRNA and no effect on DCV viral load. Likewise, *dicer-2* knock-down did not result in a significantly higher viral load.

Statistical analysis highlighted 24 dsRNAs (corresponding to 20 genes) that have an effect statistically different from the mean of the dsRNA effects on DCV viral load (*Figure 31 e³ Table 5*). Among the 20 candidate genes identified, 13 are described or predicted nucleases, 5 are Dicer-2 interactants in DCV infected conditions, one is a CrPV cloverleaf interactant and one was cherry picked from literature. In total, this screen allowed the identification of 14 candidate genes with an antiviral role and 6 with a proviral role on DCV replication. Explanation about their previously described role and possible implication in an antiviral response will be done in the Discussion section of this chapter.

An important information to take into account when performing such an RNAi based screen is the toxicity of the dsRNAs-mediated knock down of genes on the cells. Indeed, a dsRNA which has a too strong impact on global cell metabolism will randomly favor or inhibit viral replication independently of its possible interaction with the virus. That is why we tested the toxicity induced by the knock down of the 20 candidate genes by looking at the mitochondrial activity of dsRNA-treated cells 3d after soaking. In addition, a few randomly picked genes were included in the tested candidates (*Figure 32*). First, **dsLacZ** treated cells show around 100% of mitochondrial activity compared to untreated cells which is what is expected from this control. Then, the **dsThread** control shows a mitochondrial activity close to 0% which is expected as well as nearly all cells in this control are dead from apoptosis. Among the other 40 tested dsRNAs, 15 showed a significant decreased mitochondrial activity in comparison to LacZ treated cells. This

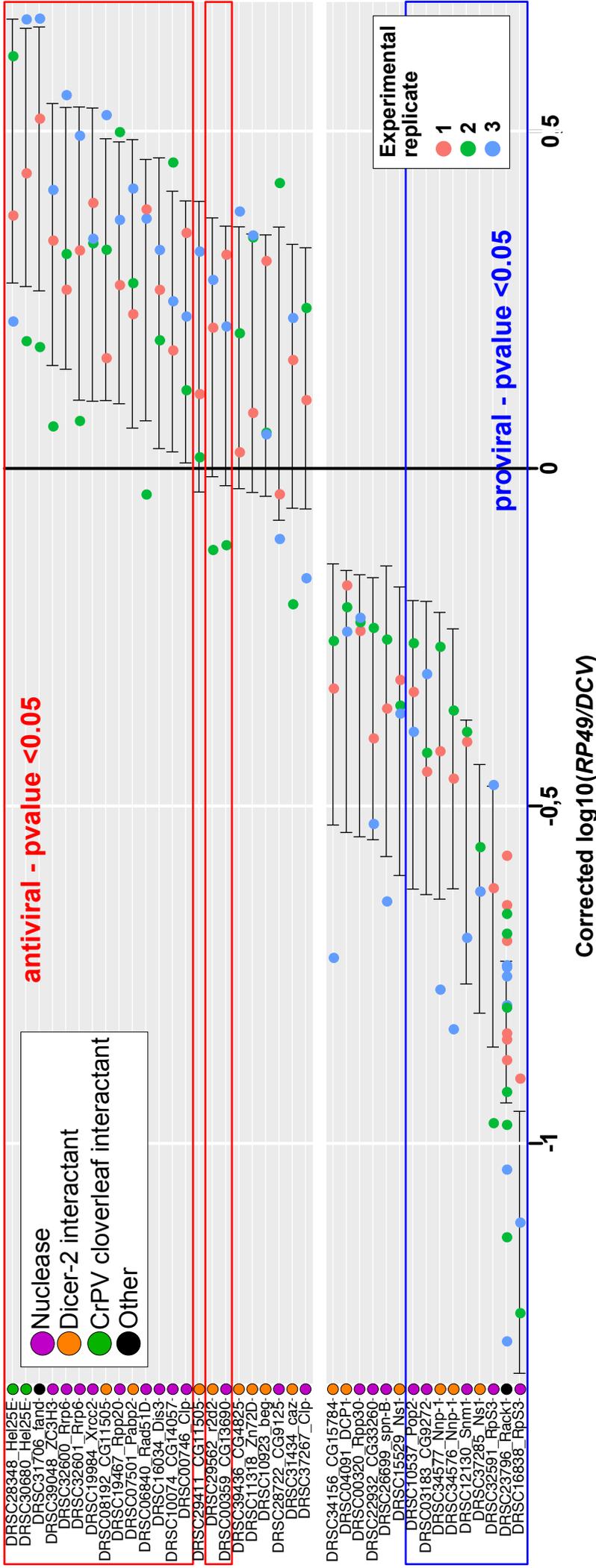


Figure 31: Zoom in on graphic from Figure 30.

# of dsRNA with significant impact on viral load	
1/1	2/2
Rpp20	Hel25E
Rad51D	Rrp6
CG14057	Rps3
Snm1	Nnp-1
CG9272	CG11505
	Pabp2
	Dis3
	Cip
	r2d2
	CG13690
	Ns1
	Pop2

Table 5: Summary table of genes with significant impact on viral load upon knock down.

Nuclease
Dicer-2 interactant
CrPV cloverleaf interactant
Other

Selected candidate genes were knocked-down using either one or two dsRNAs depending on the availability at the Drosophila RNAi screening Center at Harvard Medical School. This table summarizes how many dsRNAs per selected gene had a significant impact on DCV viral load (increase or decrease).

Total dsRNAs	5	4	11	24
--------------	---	---	----	----

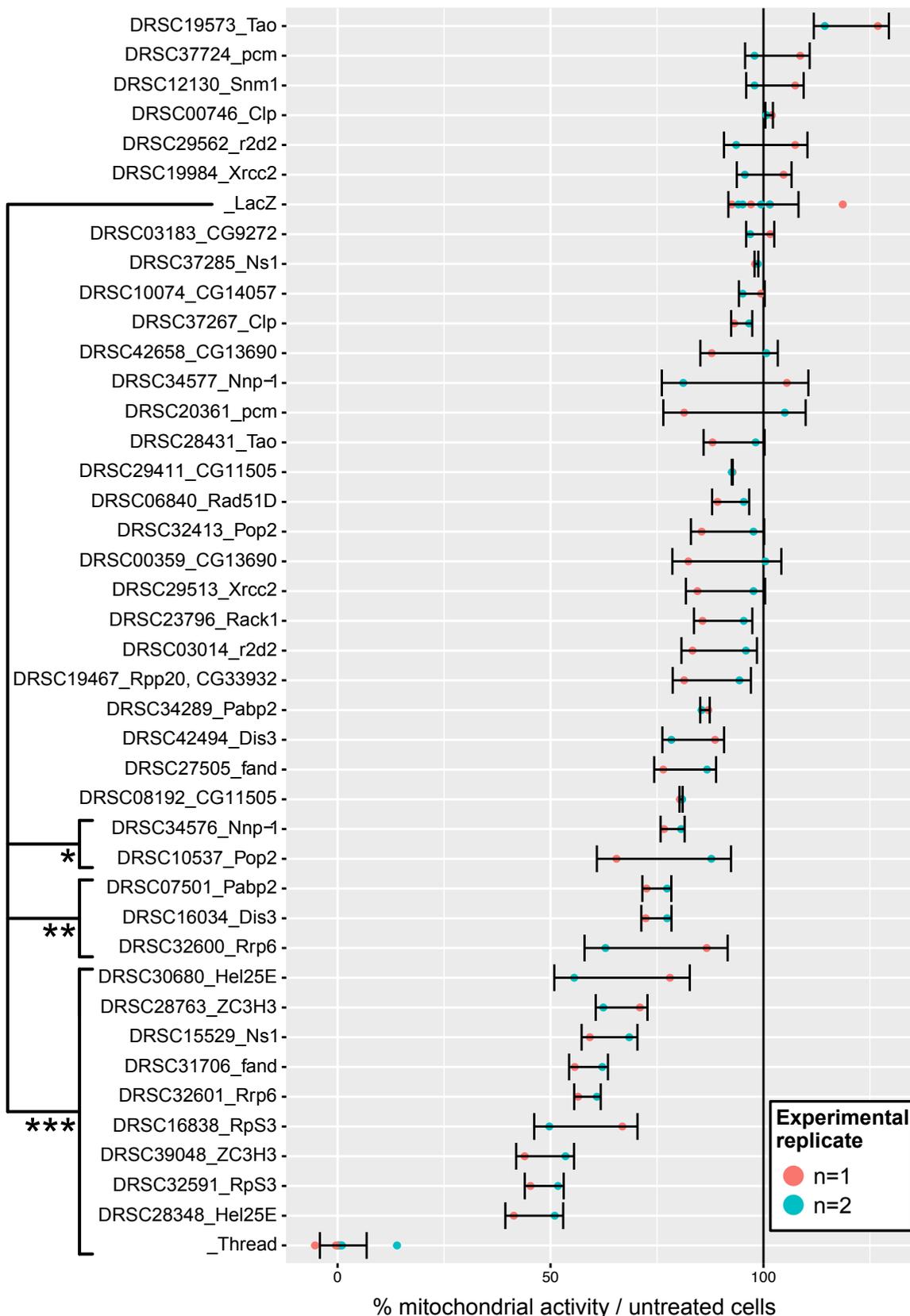


Figure 32: Determination of dsRNAs effect on mitochondrial activity of treated cells.

Cell toxicity of each dsRNA was tested by measuring the mitochondrial activity 3 days after soaking with each dsRNA. The mitochondrial activity of dsRNA treated cells was compared to the one of untreated cells (n=2, error bars represent the standard deviation between experimental duplicates, statistical test: one-way ANOVA).

information by itself does not invalidate their possible pro or antiviral activity against DCV but has to be kept in mind when later investigating their mechanism of action.

In conclusion, this screen allowed the identification of 20 genes which have a significant impact on DCV viral load when down regulated. Further investigation of their antiviral or proviral properties will be done first by repeating the experiment with viruses related or not to DCV. One interesting candidate to identify would be a protein with a dicistrovirus specific antiviral activity. Such a candidate could be further characterized in flies and tested for an eventual nuclease activity on DCV domain I ssRNA.

Conclusions – Discussions – Perspectives

In Chapter I, we saw that the bioinformatic analysis of small RNA sequencing data generated from infected flies or cells could help us to get hints on the mechanism of Dicer-2 on its viral substrate. A precise and internal entry point of Dicer-2 on the dsRNA corresponding to the 5'UTR of DCV and CrPV could be inferred from this. The distribution of siRNAs found in this region as early as 3h pi suggests a fast sensing and processing of viral dsRNA by Dicer-2. However, a question previously raised in the discussion of Chapter I remains: how does Dicer-2 manages to enter on another region than the very 5' extremity?

The described entry point is located in the domain I, a region mildly characterized in CrPV and unknown in DCV (Gross et al., 2017). Because the distribution of the reads was more consistent with the libraries generated from DCV infected samples, I decided to investigate on various aspects of the domain I of this virus. First, chemical probing of two sizes of DCV domain I RNA allowed me to draw a 2D model of it. Then, I wanted to test the hypothesis stating that the internal entry of Dicer-2 is dependent on an initial cleavage event of the viral ssRNA. To do so, I performed *in vitro* cleavage experiments revealing that this RNA structure is not sensitive to Dicer-2 processing but is subject to cleavages mediated by other cellular endonucleases. I unsuccessfully tried to validate these cuts *in vivo* by using a 5' RACE PCR method. Finally, as a first step toward the identification of proteins involved in antiviral immunity and eventually in the initial cleavage event of the DCV domain I ssRNA secondary structure, I performed an RNAi-based screen in S2 cells on candidate genes. Altogether, the results obtained in this chapter pave the way for future investigations and can be put in perspectives with other studies.

a. Comments on DCV domain I RNA secondary structure

The *in vitro* chemical probing of DCV domain I allowed me to draw a model that could fit nearly all identified reactivities. The organization in three stem loops upstream of the 5' IRES resembles structures previously identified in viruses belonging to the close picornaviruses family (summarized in Table 1 of Kloc et al., 2018).

Did I solve the secondary structure of DCV domain I?

Comparison between DCV and CrPV domain I revealed large differences in their 1D and 2D organization. These differences, even if not surprising considering the wide diversity of RNA structures described in the 5' UTR of closely related picornaviruses, made me wonder about the accuracy of my model. Indeed, while CrPV domain I RNA encompasses a poliovirus-like cloverleaf and 3 small stem loops, DCV domain I secondary structure is only composed of three stem loops (*Figure 21 e³ 23*). However, it has to be reminded that in the first paper describing CrPV domain I, a five stem loops organization was described (Gross et al., 2017). It is only by refining of the SHAPE-based model with DMS and CMCT chemical probing reactivity values that a cloverleaf-like structure could be identified. Thus, the model of DCV domain I obtained fit the reactivity values of DMS and CMCT probing but still requires complementary experiments to be validated. One way to validate the described secondary structures would be to mutate specific nucleotides and check for the differences in reactivity values after chemical probing.

Importantly, all these models were obtained from experiments using *in vitro* modifications of synthetic ssRNA and do not consider the interaction of these structures with host or viral proteins involved in the replication cycle of the viruses. Interestingly for the field, an increasing number of *in vivo* RNA structure determination methods appeared recently (reviewed in Bevilacqua et al., 2016). Because they are high throughput, these methods also bring the possibility to catch transient conformational changes of the structures that would not be detected with purified synthetic RNA. However, one of the major limitations from all these technics lies in the lack of chemical reagents known to modify double-stranded regions. Thus, only information about single-stranded regions can be directly gathered while double-stranded regions are indirectly inferred and can arise from both RNA secondary structures or RNA binding protein protective effect. In conclusion, it could be interesting to investigate DCV domain I structure *in vivo* but this information in itself will not be sufficient to explain the internal entry of Dicer-2 on DCV dsRNA.

VPg uridylylation motif

An interesting observation comes from the presence of an AAAY motif (with Y = pyrimidine) in DCV and CrPV domain I at positions 21-25 and 129-132 respectively (*Figure 21 e³ 23*). This sequence is conserved in *cis*-active RNA elements (CRE) involved in the replication of picornaviruses (reviewed in Steil and Barton, 2009). In poliovirus, it was shown that a AAAC motif located in a CRE is used by the viral RdRP as a template

for the uridylylation of the tyrosine residue of a newly translated VPg (Paul et al., 2003). This modified VPg is then used as a protein primer for viral replication (*Figure 19* and reviewed in Paul and Wimmer, 2015). Of note, the genomic position of CRE in picornaviruses varies a lot and can be located virtually anywhere. Thus, the identification of this uridylylation motif in the domain I of DCV and CrPV is purely speculative (60 and 82 occurrences of AAAC motif in DCV and CrPV sequences respectively). One way to test the implication of this motif in dicistroviruses replication could be by mutating it in a replicon. The need for replicons will be discussed shortly after. It could also be interesting to bioinformatically search which AAAC motifs in DCV and CrPV sequences are more likely to be present in loops of ssRNA hairpins (by mfold prediction for instance).

b. Characterization of DCV domain I sensitivity to cleavage

One of the hypotheses raised in the discussion of Chapter I to explain the internal entry of Dicer-2 on DCV relies on an initial cut of its domain I ssRNA secondary structure by an endonuclease. Therefore, I investigated DCV domain I sensitivity to cleavage and discovered that, while a Dicer-2 dependent cut of the RNA seems unlikely, it is subject to cuts in the predicted accessible (i.e. single-stranded) regions when incubated with embryonic extracts (*Figure 24*). Two aspects have to be taken into consideration to temper these results:

1. The RNA template is a synthetic one. Indeed, the RNA used for these cleavage assays is just a fraction of the DCV genomic RNA, lacks a VPg and is not being involved in viral replication. It is likely that the structure containing the hypothetical Dicer-2 entry point undergoes structural modification *in vivo* to accommodate biological processes.
2. Embryonic extracts represent *in vivo* conditions at a very precise moment. Even if sequencing data obtained in S2 cells (embryonic cell line) recapitulated the results obtained in flies, we cannot exclude the fact that the antiviral defense could be different at this developmental stage.

Still, this cleavage sensitivity of DCV ssRNA in bulgy regions builds confidence around my predicted model. Indeed, accessible ssRNA regions are more likely to be targeted by endonucleases.

I then asked myself whether these *in vitro* identified cuts could have an *in vivo* relevance. However, these cuts could not be validated using a 5' RACE PCR method despite extensive trials and optimization. Different comments could be done as a personal retrospective regarding this method. First, in the way I used it, the 5' RACE method is perfect to identify abundantly present RNAs but probably not sensitive enough to detect a cleavage event that might happen less than 1 time over 100 replication cycles. Attempts at adapting this method to high-throughput sequencing failed. Moreover, too many steps (e.g. PCR amplifications) were involved in the protocol not to induce experimental bias and artefacts. Thus, these technical issues added to the large background noise emerging from the ongoing infection (e.g. defective particles, degradation products, RNAi machinery...) make the method unsuitable to detect the hypothetical initial cleavage of DCV domain I ssRNA.

Being able to precisely identify the 5' extremities of DCV in the time course of an infection is relevant not only in the context of the host antiviral defense but also in the context of the ill-characterized replication cycle of dicistroviruses. One possible method that could be tried to shed light on these unknowns is the direct long RNA sequencing using Oxford Nanopore technology for instance. Indeed, this technology requires no amplification steps and allows the sequencing of full-size individual RNA molecule. It was previously successfully used to determine the exact sequences of FHV defective interfering particles for instance (Jaworski and Routh, 2017). Coupling of my small RNA sequencing data with these long RNA sequencing data would be a good method for the determination of DCV 5' extremities in a high-throughput and unbiased manner.

c. Comments on the RNAi-based screen in S2 cells

In an attempt to identify new drosophila antiviral proteins, I performed an RNAi-based screen in S2 cells. This study led to the identification of 20 genes with a potential impact on DCV replication upon knock-down (*Figure 31 e³ Table 5*). Before being able to conclude on the implication of these genes in the viral cycle, several things should be done:

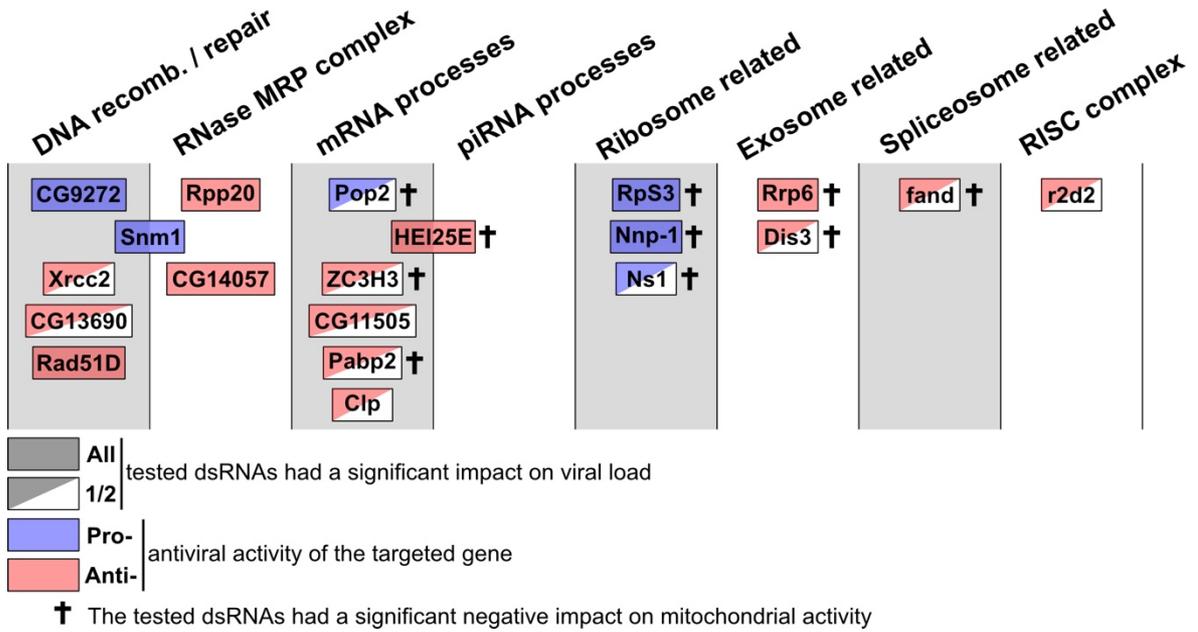
- Repeat the experiment outside of the screen context. Indeed, these large-scale handling of cells and plates can induce some bias in the final data. The analysis of the data was done taking into consideration these plate effects with the help of a

statistical model (see Materials and methods), but I cannot exclude the possibility that some false positives or negatives could have arisen from this.

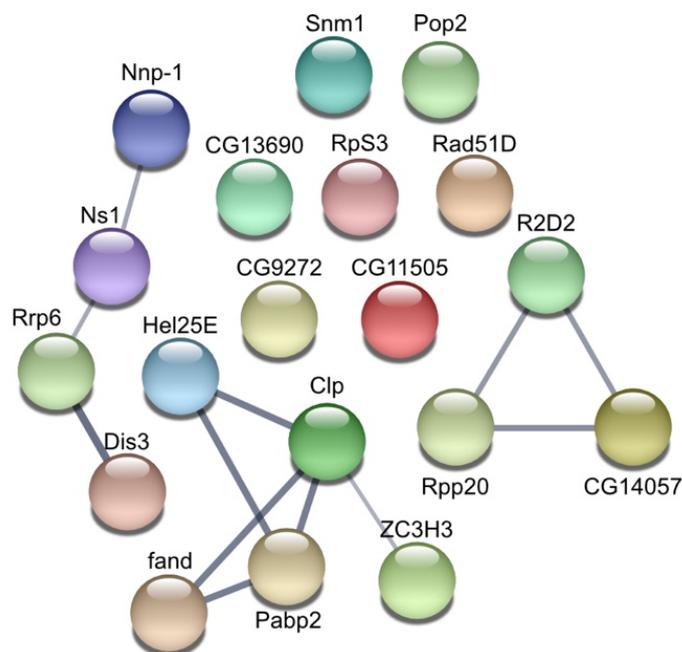
- Check for knock-down efficiency at RNA level (RT-qPCR). In the context of a large screen, I could not check the knock-down of each targeted gene. These dsRNAs used were picked because of their previously identified efficiency in other RNAi screens but it is still possible that in my conditions, their effect was dimed. The dsRNAs targeting *diap1* and *Rack1* had the expected effect on the cells and viral load. However, 11 out of 20 candidate genes were identified as pro or antiviral by only one of the two dsRNAs used, suggesting that not all dsRNAs are equally efficient (*Table 5*).
- Check for knock-down efficiency at protein level (western blot). If antibodies for the remaining candidates are available, the quantity of protein upon knock-down should be assessed. However, it has to be noted that a reduction in protein quantity might not be sufficient to prevent its action and even a small number of remaining proteins could be enough to perform its biological action. Thus, a knock-out of the gene in cells by CRISPR-Cas9 could be considered (Kunzelmann et al., 2016).
- Test for the effect of knock-down in flies. One of the reasons why drosophila is such a great model animal is because of the existence of enormous fly banks containing all the previously generated fly lines. Thus, if RNAi or knock-out fly lines were previously generated for my candidate genes, I could perform *in vivo* infection and check for the relevance of the data obtained in cells.
- Test for the virus specificity of the targeted gene. The first step of this screen was to test for the involvement of these proteins in the host response to DCV because it is the virus with which I had the clearest results in HTS experiments. However, dicistrovirus specificity of the candidate gene could be tested by performing a CrPV infection in addition to viruses outside of this family such as VSV ((-) ssRNA virus), FHV ((+) ssRNA virus) or IIV-6 (DNA virus).
- Manage to uncouple cell toxicity effect from pro or antiviral effect. Indeed, some of the tested genes had a strong impact on final viral load but also strongly diminished mitochondrial activity (*Figure 32*). Among the possible experiments to overcome this effect, I could (1) perform over expression experiments to check for possible opposite effect on final viral load or (2) check direct interaction between viral proteins/RNAs and host protein by microscopy and immunoprecipitation

methods. However, we cannot exclude that upregulation of candidate genes could also affect fitness of the cells and that microscopy techniques could miss an indirect interaction between virus and host protein.

In conclusion, the data of this screen should be carefully considered and are only a first step toward the long-term identification of new pro or antiviral proteins. Still, bibliography search allows to classify the candidate genes in different biological processes.



A string analysis only taking into account experimental data and databases mining revealed little interaction between the candidate genes.



As a general comment regarding this screen, the readout chosen might not have been the most optimal one. Indeed, more than finding new antiviral proteins, I would have liked to find a nuclease able to perform a cut in the DCV domain I ssRNA. The rationale behind the viral load output was that the lack of this hypothetical nuclease should have a strong impact on the global host response and thus, viral replication. However, I cannot exclude the fact that the inhibition of the first cut mediated in the ssRNA of DCV might not be sufficient to have a significant effect on viral load. Dicer-2 could find another way to target viral dsRNA which would still give rise to enough siRNAs to activate the antiviral pathway. Additionally, this screen could have been the opportunity to identify whether a host nuclease was responsible for the degradation products described in the sequencing data of Chapter I. Thus, once again, high-throughput sequencing could have been a valuable output choice for this screen. One of the major limitations of this technique remains its cost. However, new methods are emerging taking advantage of the double indexing method to multiply the number of multiplexed samples in a single sequencing lane up to 96 (Persson et al., 2017). Analysis of the DCV mapping reads size distribution could then have been a valuable information to answer our questions.

d. What is the role of DCV domain I?

One of the questions that has not been addressed so far is the following one: if this region represents such a weak point in DCV defense, how come adaptation of the virus did not modify it? Indeed, viruses are certainly the most adaptable parasites and always find ways to escape the host immune system. One of the reasons that could be proposed to explain this is given by the possible implication of this specific RNA domain I of DCV in its replication. Indeed, as explained in the introduction of this chapter, RNA structures located in the 5' UTR of picornaviruses were shown to be involved in various aspects of their cycle.

In order to test this, it could be interesting to generate a DCV replicon. Being able to clone the full sequence of DCV inside a plasmid recapitulating a virus infection upon transfection in cells or insertion in fly's genome would pave the way for several downstream applications. Such a strategy was already successfully used for CrPV or FHV in order to investigate translation mechanisms or antiviral RNAi response for instance (Khong et al., 2016; Li et al., 2002; Martins et al., 2019). Thus, a DCV replicon could be precisely mutagenized to search for involvement of domain I in replication for

instance. Additionally, characterization of the antiviral RNAi response in absence of the 1A VSR could bring answers to questions previously raised in the discussion of Chapter I. However, a big caveat of using replicons lies in the fact that it bypasses critical steps of the viral cycle, namely the binding, entry and uncoating of the virion. In addition, cloning of the DCV sequence might be difficult due to its size (~9,2kb). Several attempts were already done in the lab without any success. The solution to this could come from the discovery of new reverse transcriptase enzymes that form less abortive products from long RNA templates. As an example, a reverse transcriptase encoded by *Eubacterium rectale* (MarathonRT) was recently shown to be able to fully transcribe the highly structured genome of HCV (~9,6kb) with 93% full-length products and few apparent stops (Zhao et al., 2018). We are currently trying to use this enzyme to generate a DCV replicon.

Being able to recapitulate DCV replication system *in vitro* by using a replicon could also open the way toward high-throughput implementations by using microfluidics for instance. Indeed, we could develop a reporter system in which *in vitro* replication and translation of DCV replicon done in a micro droplet would yield a fluorescent signal for instance. Thus, impact of random mutations done in DCV RNA structures on viral replication could be assessed in a high-throughput manner. This method could also allow a fast refinement of our domain I 2D structure model and be used to investigate for a nuclease able to cut DCV domain I. Interestingly for us, the IBMC hosts a team successfully using microfluidics for ultrahigh-throughput screening and collaborative projects were already initiated (Dr. Michael Ryckelynck team – UPR9002).

Chapter III

Study of Dicer-2 helicase domain involvement in RNA sensing and processing



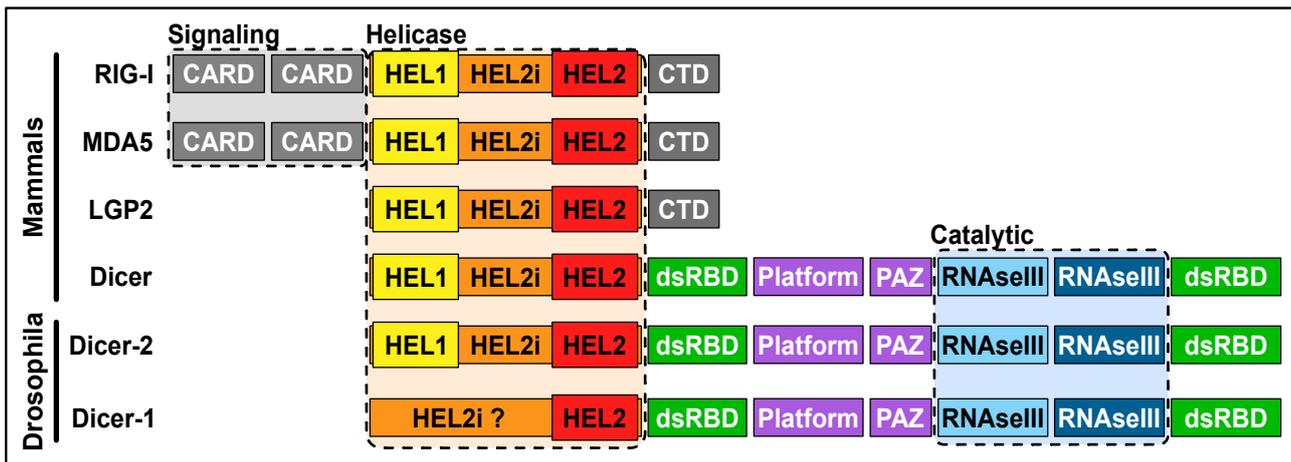


Figure 33: Domain organization of different proteins from the DRA family (adapted from Paro et al., 2015).

DRA-containing proteins of mammals and drosophila only are displayed. However, it has to be noted that other proteins involved in immunity but not discussed here also contain the peculiar helicase domain of DRAs (e.g DRH1-3 and Dicer-I proteins from *C. elegans*). All the presented proteins, with the exception of dmDicer-1 present the HEL1-HEL2i-HEL2 organization of their helicase domain. Thus, presence of dmDicer-1 in this category can be discussed. This helicase domain encompasses conserved motifs required for ATP-binding and hydrolysis. RIG-I and MDA5 both possess two additional C-terminal CARD domains that are sequestered when no RNA is sensed but exposed during oligomerization of these proteins on their dsRNA substrate. Interaction between CARD domains and MAVS is the initial step of a signaling cascade that ultimately leads to the production of ISGs. Dicer enzymes all possess additional domains including two dsRNA binding domains, the PAZ-platform domain used to bind the 5' phosphate extremity of dsRNA and act as physical "ruler" to determine the size of siRNAs and finally, two RNaseIII domains responsible for the dicing of the dsRNA.

Introduction

Helicases are ubiquitous proteins that exist throughout all forms of cellular life (Bleichert and Baserga, 2007; Linder and Jankowsky, 2011). Their involvement in virtually all facets of DNA and RNA metabolism is depicted by the numerous diseases caused by their deregulation proteins (Steimer and Klostermeier, 2012; Suhasini and Brosh, 2013). Helicases can be classified in 6 superfamilies (SF) based on their structures, functions and shared sequence motifs (Singleton et al., 2007). SF1 and SF2 enzymes contain a conserved helicase structure composed of two RecA-like domains. These domains encompass specific sequence motifs required for ATP hydrolysis and nucleic acid binding (Putnam and Jankowsky, 2013). Alignment of the core sequences of SF1 and SF2 helicases from *S. cerevisiae*, *E. coli* and some viruses led to the clustering of these two superfamilies in 12 families (9 in SF2 and 3 in SF1, Fairman-Williams et al., 2010).

Important players of mammalian antiviral immunity can be found in the Rig-I-like receptor (RLR) family of the SF2 superfamily, namely RIG-I, MDA5 and LGP2. Briefly, detection of viral RNA by RIG-I or MDA-5 lead to their oligomerization on their dsRNA substrate, signaling through their N-terminal CARD domain to MAVS, activation of the IFN pathway and subsequent expression of ISGs to mount an antiviral response (reviewed in Yoneyama et al., 2015). In the case of the CARD-less LGP2, additionally to a suggested role in modulation of RIG-I and MDA5 activity, it could be involved in the negative regulation of RNAi in mammals through a direct interaction with Dicer (Veen et al., 2018). RLRs have a peculiar helicase domain organization with a Hel2i domain inserted in between the two RecA-like domains. This peculiar domain organization is conserved between RLRs and Dicer enzymes and plays an important role in RNA sensing (Civril et al., 2011; Jiang et al., 2011; Kolakofsky et al., 2012; Luo et al., 2011). However, because the proper unwinding activity of these inferred helicases was never proven, they were renamed as Duplex RNA activated ATPases or Double stranded RNA-dependent ATPases (DRA, reviewed in Luo et al., 2013; Paro et al., 2015).

Dicer enzymes belong to the DRA family. The main differences between Dicer and RLR lies in the absence of N-terminal CARD domains and the presence of additional RNaseIII domains in Dicer (*Figure 33*). This difference in domain composition results in a catalytic mode of action of Dicer enzymes on their substrate, which explains why Dicer

enzymes are classified as catalytic (c) DRAs while RIG-I and MDA5 are signaling (s) DRAs (reviewed in Paro et al., 2015). However, it has to be noted that the strict distinction of both categories can be discussed. Indeed, direct antiviral effectors functions of RLRs by displacing viral proteins were already suggested (Sato et al., 2015; Weber et al., 2015; Yao et al., 2015). On the other hand, catalytic activity of Dicer-2 on viral dsRNAs (i.e dicing) is not sufficient to mount an antiviral RNAi response and requires the amplifying action of Ago2 (van Rij et al., 2006). Additionally, activation of antiviral genes by Dicer-2 such as *vago* was shown to restrict viral replication (Deddouche et al., 2008).

DRA enzymes were shown to have different substrate specificities and to present ATP-dependent and -independent activities. Thus, RIG-I was shown to recognize 5' di- or triphosphate moieties of dsRNAs while MDA5 adopts a stem-binding mode of long dsRNAs with no contact to the dsRNA end (Goubau et al., 2014; Lu et al., 2010; Schlee et al., 2009; Wang et al., 2010; Wu et al., 2013). Then, MDA5 oligomerization on its substrate is ATP-independent but ATP hydrolysis promotes its disassembly from short dsRNAs (Peisley et al., 2011, 2012). On the other hand, RIG-I monomers can bind dsRNA extremities in an ATP-independent manner but require ATP hydrolysis for their oligomerization (Goubau et al., 2014; Luo et al., 2011). Interestingly, these RLRs can present viral substrates preferences. In the case of picornaviruses that do not present a genomic 5' PPP, it is MDA5 and not Rig-I which is required to mount an efficient antiviral response (Feng et al., 2012). On the contrary, drosophila Dicer-2 was shown to be required for the defense against all tested viruses, regardless of the nature of their genome.

Similarly to RLRs, extensive study of drosophila Dicer-2 *in vitro* revealed ATP- and substrate-dependent distinct modes of action (Cenik et al., 2011; Sinha et al., 2015; Welker et al., 2011). Thus, blunt end dsRNA triggers an efficient DRA domain- and ATP-dependent processive activity of Dicer-2 whereby a single Dicer-2 protein will dice multiple times before dissociating. On the contrary, a dsRNA molecule with 3' overhang termini, promotes a slow, ATP-independent and distributive activity of Dicer-2 characterized by dissociation of Dicer-2 after each cleavage event. A recent cryo-electron microscopy-based study proposed two distinct sensing mechanisms of Dicer-2 depending on dsRNA termini (*Figure 34*, Sinha et al., 2018):

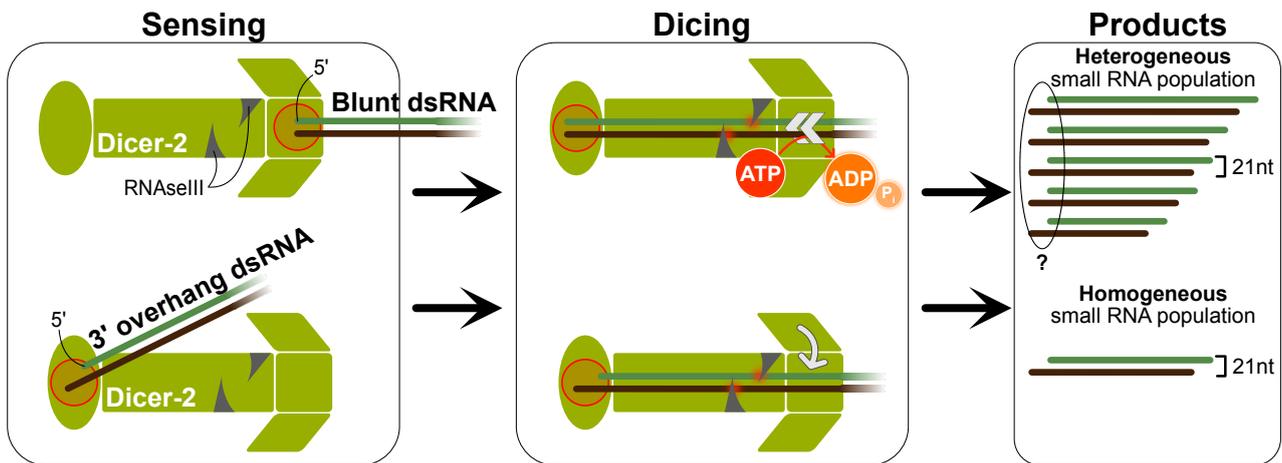


Figure 34: Model of action of Dicer-2 depending on the dsRNA substrate extremity.

Blunt dsRNA is sensed by the helicase domain of Dicer-2 and is threaded in an ATP-dependent manner until it reaches the PAZ-platform domain. This mode of action of Dicer-2 produces longer and shorter RNAs in addition to the canonical 21nt long siRNAs. Their disposition on this scheme following the 21nt 3' overhang rule is purely hypothetical. Occurrence of these other RNA products can be explained by the fact that while being threaded, the dsRNA molecule encounters the RNaseIII domains. This phenomenon most likely results in a random dicing of the moving dsRNA molecule.

21nt 3' overhang dsRNA is bound by the PAZ-platform domain (interaction between 5'P and phosphate binding pocket) and brought to a close proximity to RNaseIII domains in an ATP-independent manner. Because of this fixed anchoring point for the dsRNA molecule, a single canonical siRNA size can be generated (21nt), determined by the distance between the PAZ-platform domain and the RNaseIII domains.

- Blunt dsRNAs are first bound by the helicase domain of Dicer-2. This is followed by the ATP-dependent threading and unwinding of the dsRNA molecule until it reaches the PAZ-platform domain and is cut by the RNaseIII domains. It has to be noted that this threading mechanism can lead to the production of shorter or longer RNAs than canonical siRNAs (from 5 to ~30nt long). We do not know yet if this heterogeneity of Dicer-2 products is relevant *in vivo* or if binding cofactors prevent the random dicing of helicase-threaded dsRNAs.
- On the other hand, dsRNA with 3' overhang termini are directly 5' P bound by the PAZ-platform domain through the phosphate binding pocket (Kandasamy and Fukunaga, 2016). The dsRNA molecule is eventually brought in close proximity to the RNaseIII domains and cleaved. This mechanism is ATP-independent and only generates canonical siRNAs.

This study was made possible by the functional uncoupling of the two modes of action of Dicer-2. Thus, mutations in the PAZ-platform domain inhibited dicing of 3' overhang dsRNAs while maintaining processivity on blunt dsRNA. On the contrary, a mutation in the helicase domain of Dicer-2 altered its action on blunt dsRNA while leaving unchanged its distributive activity on 3' overhang dsRNAs. Interestingly, this last mutation was identified by searching for similarities between Dicer-2 sequence and C-terminal domain of RIG-I, which was suggested to be responsible for recognition of blunt dsRNA (Luo et al., 2011). One such region of Dicer-2 was identified in its helicase domain and a single mutation of a phenylalanine (F) 225 to a glycine (G) was performed (Dicer-2^{F225G}). Because of the proposed role of RIG-I C terminal domain in sensing, it was proposed that this mutation in Dicer-2 could have an impact on the processing of *bona fide* endogenous or viral Dicer-2 targets *in vivo*.

Altogether, this study proposes a model of drosophila Dicer-2 action on different synthetic dsRNAs. Importantly, in *in vitro* studies, the used dsRNA substrates sequences and extremities are arbitrary defined. Indeed, the exact characteristics of *bona fide* Dicer-2 dsRNA targets *in vivo* remain one of the outstanding unknowns in the field of RNAi. This is why, following this milestone publication, it was decided to study the impact of the *in vitro* characterized Dicer-2^{F225G} mutation in flies. Flies expressing this mutated version of Dicer-2 were generated alongside flies described in Chapter I. Characterization of the Dicer-2-dependent siRNA pathways, namely the endo-siRNA and antiviral siRNA

pathways, was done using regular RT-qPCR methods together with the powerful small RNA HTS method. Finally, comparison between data obtained from different fly genotypes allowed us to gain further insights in the involvement of Dicer-2 helicase in sensing and processing of its *bona fide* targets.

Results

I. Impact of two Dicer-2 helicase mutations on the endo-siRNA pathway

a. Flies generation and small RNA HTS of DCV, TRIS and VSV injected flies

To test the effect of the F225G mutation on the function of Dicer-2 *in vivo*, we generated GFP::Dicer-2^{F225G} expressing flies. The transgene used was inserted at the same genomic position as the ones used to generate transgenic flies from Chapter I in order to avoid any position effect. Also, the same poly-ubiquitin promoter controls the expression of GFP::Dicer-2^{F225G}. This allows the comparison between different variants of GFP::Dicer-2. Flies genotypes used during this study are summarized in this table:

Genotype	Simplified name
$w^{IR}; \text{dicer-2}^{L811f;X}/Df$	dicer-2^{null}
$w^{IR}; \text{dicer-2}^{L811f;X}/Df; \text{GFP::Dicer-2}^{WT}$	GFP::Dicer-2 ^{WT}
$w^{IR}; \text{dicer-2}^{L811f;X}/Df; \text{GFP::Dicer-2}^{G31R}$	GFP::Dicer-2 ^{G31R}
$w^{IR}; \text{dicer-2}^{L811f;X}/Df; \text{GFP::Dicer-2}^{F225G}$	GFP::Dicer-2 ^{F225G}

In addition, flies rescued with a wild-type version of endogenous *dicer-2* recombined with the *dicer-2* deficiency were used as controls in some experiments (*dicer-2*^{rescue}).

All the complemented flies were obtained by the same crossing strategy presented in Chapter I - **Figure 13A**. As a general remark, this study was conducted at the same time as the flies' small RNA HTS described in Chapter I. That is why some of the following figures are actually the same as the ones shown in the previous chapter with the GFP::Dicer-2^{F225G} flies results being added. In order to be able to compare the results obtained from the different complemented flies, we needed to make sure of the proper expression of the different versions of GFP-tagged Dicer-2. First, an RT-qPCR on *dicer-2* mRNA (**Figure 35A**) shows:

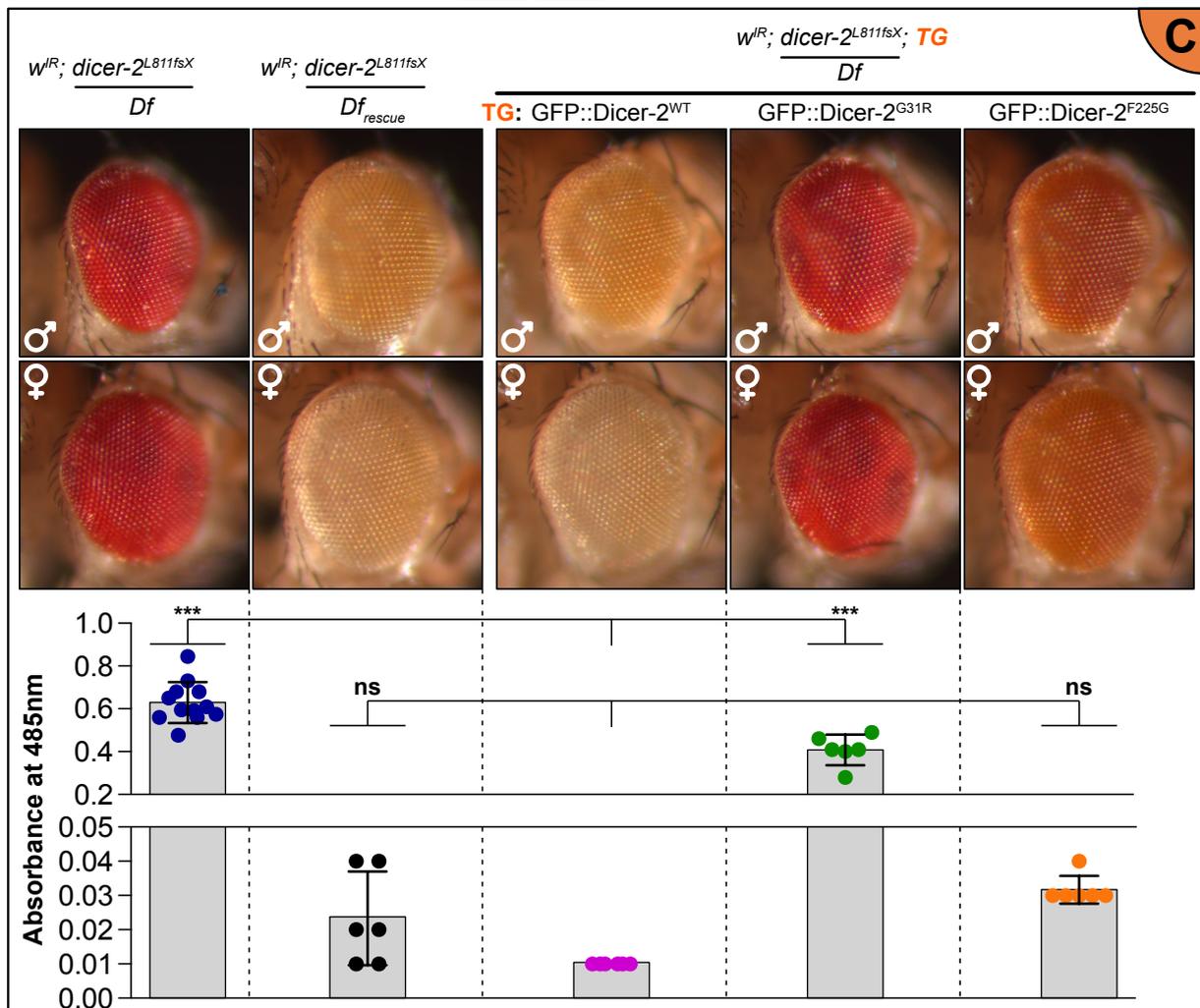
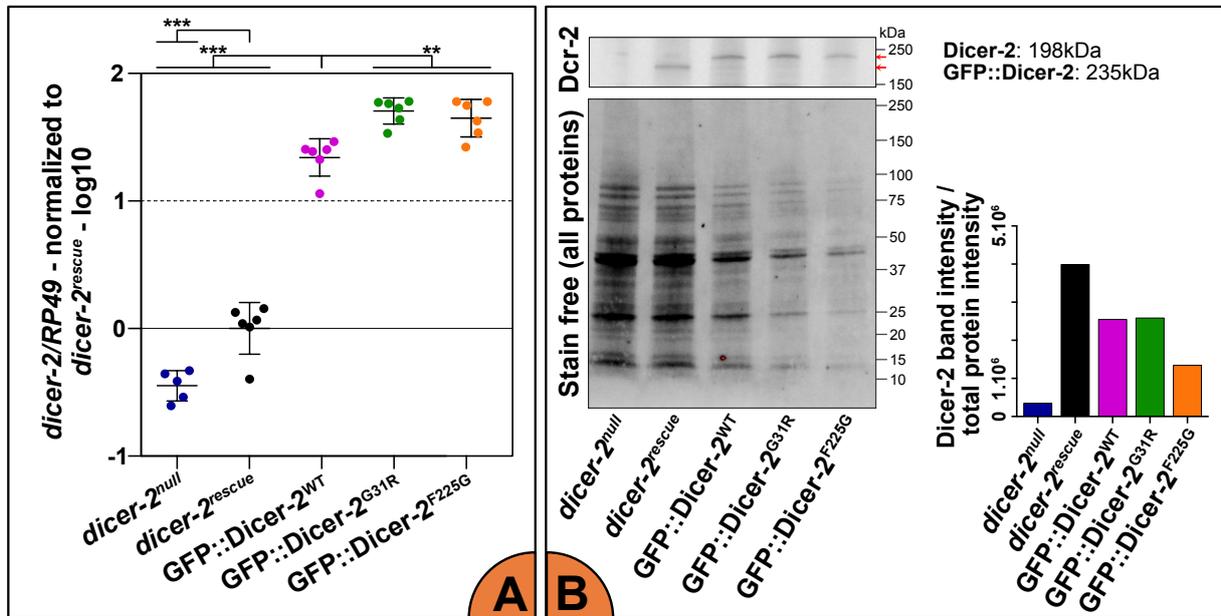


Figure 35: Characterization of flies used for small RNA HTS.

A) Relative mRNA level of *dicer-2* in comparison to the house-keeping gene *RP49* measured by RT-qPCR (n=2, biological triplicates, 3 males and 3 females per sample, error bars represent standard deviation between data points, all data were also normalized to *dicer-2^{rescue}* control, statistical test: one-way ANOVA). **B)** Western blot showing Dicer-2 protein level in the flies. Stain free exposition of the blot is used for Dicer-2 protein quantity normalization (n=1). **C)** Eye color of the different flies according to their gender. **Top:** pictures; **Bottom:** absorbance at 485nm (n=3, biological duplicates, 10 females per sample, error bars represent standard deviation between data points, statistical test: one-way ANOVA).

- A small but significant decrease of *dicer-2* mRNA in *dicer-2^{null}* flies in comparison to *dicer-2^{rescue}* flies.
- A significantly higher quantity of *dicer-2* mRNA in all GFP::*Dicer-2* flies in comparison to *dicer-2^{rescue}* flies. This difference can be explained by the poly-ubiquitin promoter used to express *Dicer-2* tagged versions. Of note, we previously tried to use a 2kb sequence upstream of *dicer-2* as a promoter but this construct failed to restore and complement a *dicer-2* null mutant.
- A small but significant increase of *dicer-2* mRNA in GFP::*Dicer-2^{G31R}* and GFP::*Dicer-2^{F225G}* flies in comparison to GFP::*Dicer-2^{WT}* flies. This observation could suggest that these point mutations affect the stability of *dicer-2* mRNA.

However, western blot analysis shows a similar amount of *Dicer-2* proteins regardless of the complemented version of *Dicer-2* expressed (*Figure 35B*). Such a result was already observed in RFP::*Dicer-2* rescued flies (Girardi et al., 2015). In conclusion, results obtained from GFP::*Dicer-2* complemented flies can be compared between each other and should not reflect a difference in the quantity of protein expressed.

As a side note, this western blot done with the described genotypes altogether was only done once, hence the lack of statistics. However, several western blots with the different flies' genotypes were performed individually and yielded similar results.

Small RNA HTS of libraries constructed with RNA samples from TRIS, DCV and VSV injected adult flies from all 4 genotypes was conducted (*Table 3* – Chapitre I). In this first part of the Chapter III, we will focus on the endo-siRNA pathway while virus specific reads will be investigated in the second part. As a general remark, no injection-dependent (TRIS, DCV or VSV) differences could be observed when looking at the endo-siRNA pathway. That is why, for the sake of visualization, all libraries will be represented on the following graphics regardless of the injection that was made but grouping flies from the same genotype under the same color. All libraries were treated the same way regarding trimming, second demultiplexing and normalization.

b. Study of the Dicer-2 dependent RNAi pathways

GFP::Dicer-2^{F225G} flies have an intermediary phenotype regarding the *w^{IR}* transgene

Like all the flies used for this study, GFP::Dicer-2^{F225G} flies have the *GMR-w^{IR}* transgene inserted in their genome on the X chromosome. It is composed of the third exon of the gene *white* which is repeated in an inverted orientation and separated by a *white* intron. The eye specific transcription and splicing of the intron will result in a perfect hairpin dsRNA of 629bp long (Lee et al., 2004). This dsRNA is diced by Dicer-2 and induces the siRNA pathway in a R2D2 and Loqs-PD-dependent manner (Marques et al., 2010b). The exact mechanism by which the hairpin is processed by Dicer-2 remains unknown but *w^{IR}* mapping siRNAs are abundantly present and can be readily detected by northern blot. These siRNAs downregulate *white* expression, which results in a white eye phenotype.

As described in Chapter I, *dicer-2^{null}* flies have red eyes while *dicer-2^{rescue}* and GFP::Dicer-2^{WT} flies have white eyes (**Figure 35C - top**). A point mutation in the helicase domain of Dicer-2 in GFP::Dicer-2^{G31R} flies results in a red color of the eyes. This result suggests the requirement for a functional ATPase activity of the helicase domain to efficiently process the *w^{IR}* hairpin. Surprisingly, flies expressing Dicer-2 with another helicase point mutation, GFP::Dicer-2^{F225G}, presented an intermediary orange color of the eyes. Even if statistical difference was not proven, confirmation of these visual phenotypes was obtained by performing an eye color quantification on heads of females of the different genotypes (**Figure 35C - bottom**). This observation by itself suggests that GFP::Dicer-2^{F225G} is able to activate the RNAi pathway, although at a lower efficiency than the wild-type protein.

Because of this observation, two hypotheses were raised: GFP::Dicer-2^{F225G} is not able to efficiently activate the RNAi pathway because (1) the produced *w^{IR}* siRNAs are not efficiently loaded on Ago2 or (2) the level of *w^{IR}* siRNAs produced is too low. This second hypothesis was tested by looking at the small RNA HTS data. We plotted the size distribution of *w^{IR}* mapping reads for all the libraries and could make the following observations (**Figure 36A**):

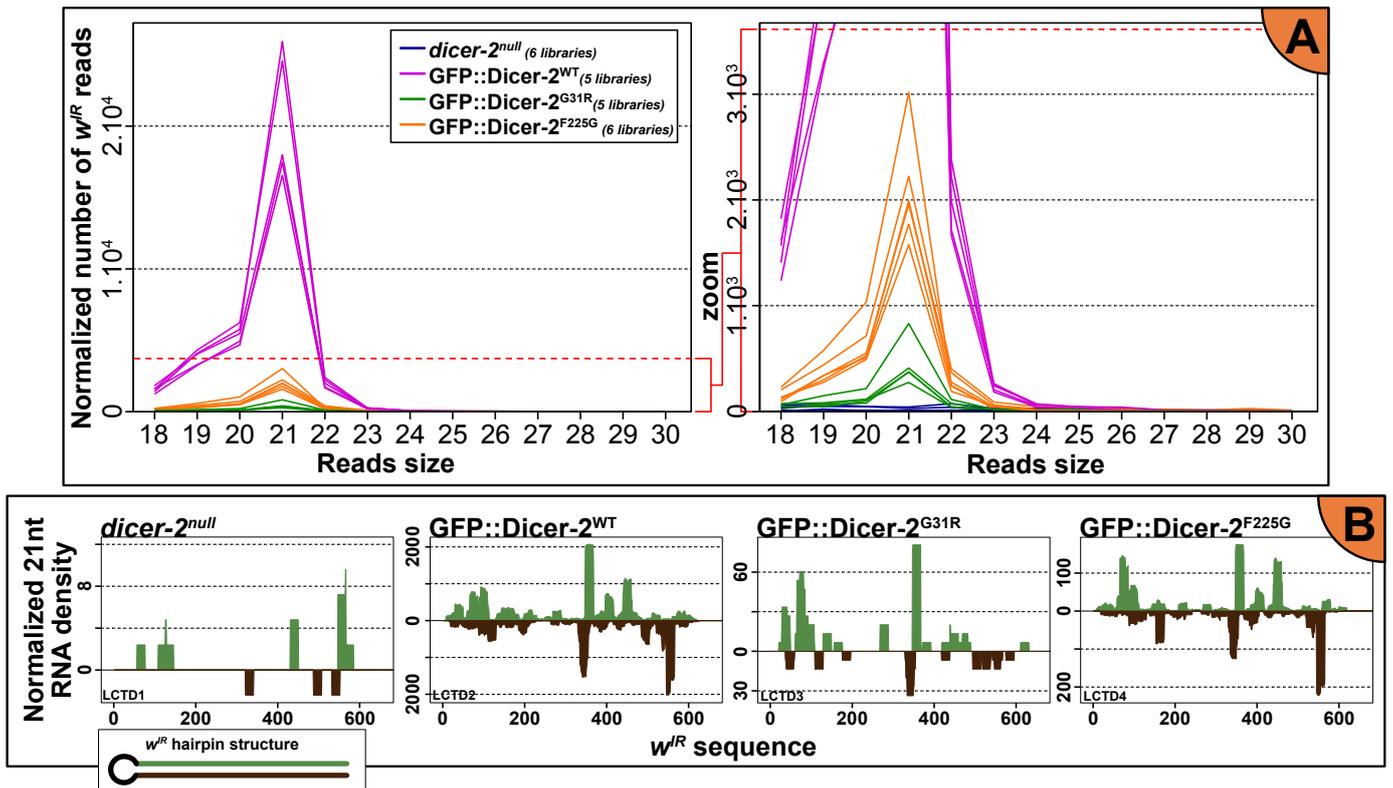


Figure 36: $GFP::Dicer-2^{F225G}$ flies have an intermediary phenotype regarding the w^R transgene.

A) Normalized size distribution of w^R mapping reads. All results from flies with similar genotype are gathered under the same color. Each curve represents an individual library. Right plot is a zoom of the left plot. **B)** Normalized w^R coverage by 21nt long reads. Shown plots are representative of all the plots obtained when looking at the data coming from flies with a similar genotype regardless of the injection performed.

- No w^{IR} mapping reads are observed in *dicer-2^{null}* flies.
- A few w^{IR} mapping reads are identified in GFP::*Dicer-2^{G31R}* flies and importantly, a peak at 21nt corresponding to Dicer-2 products is present.
- More of these reads with the same size distribution are observed in GFP::*Dicer-2^{F225G}* flies.
- Finally, a huge amount of w^{IR} mapping reads are present in GFP::*Dicer-2^{WT}* flies. Again, the peak at 21nt indicates Dicer-2-dependent products.

These sequencing data validate the hypothesis that GFP::*Dicer-2^{F225G}* is able to produce w^{IR} derived siRNAs but in a much lower quantity than GFP::*Dicer-2^{WT}* flies. This number of siRNAs is probably not sufficient to fully activate the RNAi pathway, hence an orange color of the eyes. Likewise, GFP::*Dicer-2^{G31R}* flies produce barely detectable amounts of w^{IR} siRNAs which explains their red eyes similar to *dicer-2^{null}* flies. In conclusion, these results validate the w^{IR} RNAi efficiency readout and directly correlate the amount of w^{IR} mapping 21nt long reads to the color of the eyes. However, the difference between the two helicase point mutations impact on Dicer-2 processing of w^{IR} hairpin cannot be explained yet.

To try to understand where this difference in the amount of w^{IR} derived siRNAs comes from, we plotted the distribution of 21nt long reads mapping on the w^{IR} sequence (**Figure 36B**). The idea here is to check whether this difference in number of 21nt long mapping reads reflects a global tendency across w^{IR} sequence or if it is due to a region-specific depletion of Dicer-2 products. First, when looking at GFP::*Dicer-2^{WT}* data, we can observe a very peculiar pattern of distribution of the reads with highly covered regions separated by gaps. 21nt long reads are mapping equally to both strands of the hairpin with no apparent complementarity. This unexplained coverage pattern depicts the reason why we still cannot understand the mechanism of action of Dicer-2 on w^{IR} hairpin. Phasing and offset analyses showed no siRNA signature regardless of the region analyzed. However, absence of w^{IR} coverage in *dicer-2^{null}* flies again indicates Dicer-2 dependency of these reads. When looking at the results of other genotypes, we can see that the only difference with the GFP::*Dicer-2^{WT}* pattern lies in the amount of reads mapping at each position. Indeed, no obvious depletion of sequence coverage can be detected in GFP::*Dicer-2^{F225G}* or GFP::*Dicer-2^{G31R}* flies. This result suggests that the mechanism of

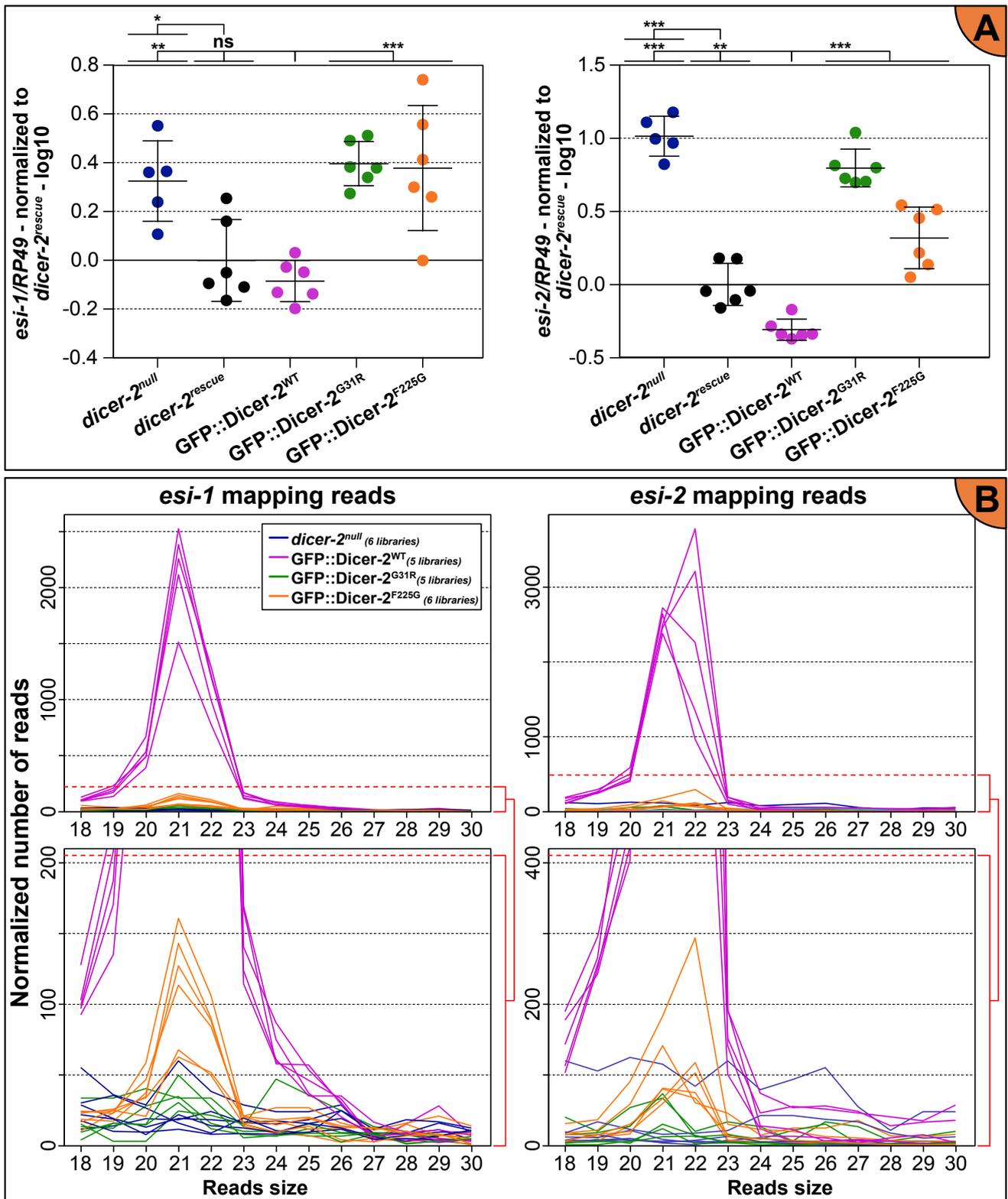


Figure 37: GFP::*Dicer-2^{F225G}* flies have an intermediary phenotype regarding *esi-1* and *esi-2* loci.

A) Levels of *esi-1* (left) and *esi-2* (right) RNA precursors normalized to the house-keeping gene *RP49* (n=2, biological triplicates, error bars represent standard deviation, all data were also normalized to *dicer-2^{rescue}* control, statistical test: one-way ANOVA). **B)** Normalized size distribution of *esi-1* (left) and *esi-2* (right) mapping reads. All results from flies with similar genotype are gathered under the same color. Each curve represents an individual library. Bottom plots are zooms of the upper plots.

action of Dicer-2 on the w^{IR} hairpin might not be directly impacted by the mutations we study but rather reflect a problem of substrate accessibility that could be due to its sensing.

GFP::Dicer-2^{F225G} flies have an intermediary phenotype regarding the endo-siRNA pathway

The w^{IR} hairpin RNA is similar to the endogenous structured loci that are precursors of natural endo-siRNAs. That is why we decided to look at the production by all these flies of endogenous endo-siRNAs. The two most represented locus in term of endo-siRNAs generation are the *esi-2* (hp-CG4068) and *esi-1* (hp-CG18854) loci (Czech et al., 2008b; Okamura et al., 2008b). Contrary to the w^{IR} transgene, no visual phenotype is associated to the generation of siRNAs from these loci, but their presence can be directly detected by northern blot or small RNA HTS. Of note, silencing of *mu508* coding gene by hp-CG4068-derived endo-siRNAs was reported. Thus, *mu508* transcripts level could have been used as indirect measurement of the endo-siRNA pathway activity. However, as previously experimented, we first tested for the processing of these endo-siRNA precursors by simple RT-qPCR in the testes of flies (Marques et al., 2010b).

Importantly, with this experiment, we are detecting the level of non-cleaved RNA precursors and not the level of Dicer-2 products (**Figure 37A**). Therefore, a high quantity of *esi-1* and *esi-2* RNAs can be detected in *dicer-2^{null}* flies in comparison to flies expressing a wild-type version of Dicer-2. A small decrease in *esi-2* precursor RNA can be observed in GFP::Dicer-2^{WT} flies in comparison to *dicer-2^{rescue}* flies. On the contrary, GFP::Dicer-2^{G31R} flies present a level of *esi-2* and *esi-1* RNAs similar to the *dicer-2^{null}* flies, suggesting the inability of this helicase mutant to process these *bona fide* substrates. GFP::Dicer-2^{F225G} flies also have a *dicer-2^{null}* mutant level of *esi-1* RNA. However, they present an intermediate level of *esi-2* RNAs. In conclusion, GFP::Dicer-2^{F225G} is less efficient at processing *esi-2* and *esi-1* RNA precursors than GFP::Dicer-2^{WT} but more efficient than GFP::Dicer-2^{G31R} at processing *esi-2* RNA. The difference of GFP::Dicer-2^{F225G} flies phenotype between *esi-1* and *esi-2* RNAs might reflect a different mode of action of Dicer-2 on these apparently similar substrates.

As a complementary approach, we checked for the quantity of endo-siRNAs from these two loci produced by each fly genotype. This information was given by the small RNA HTS data and the size distribution of *esi-2* and *esi-1* mapping reads (**Figure 37B**). The following observations could be made:

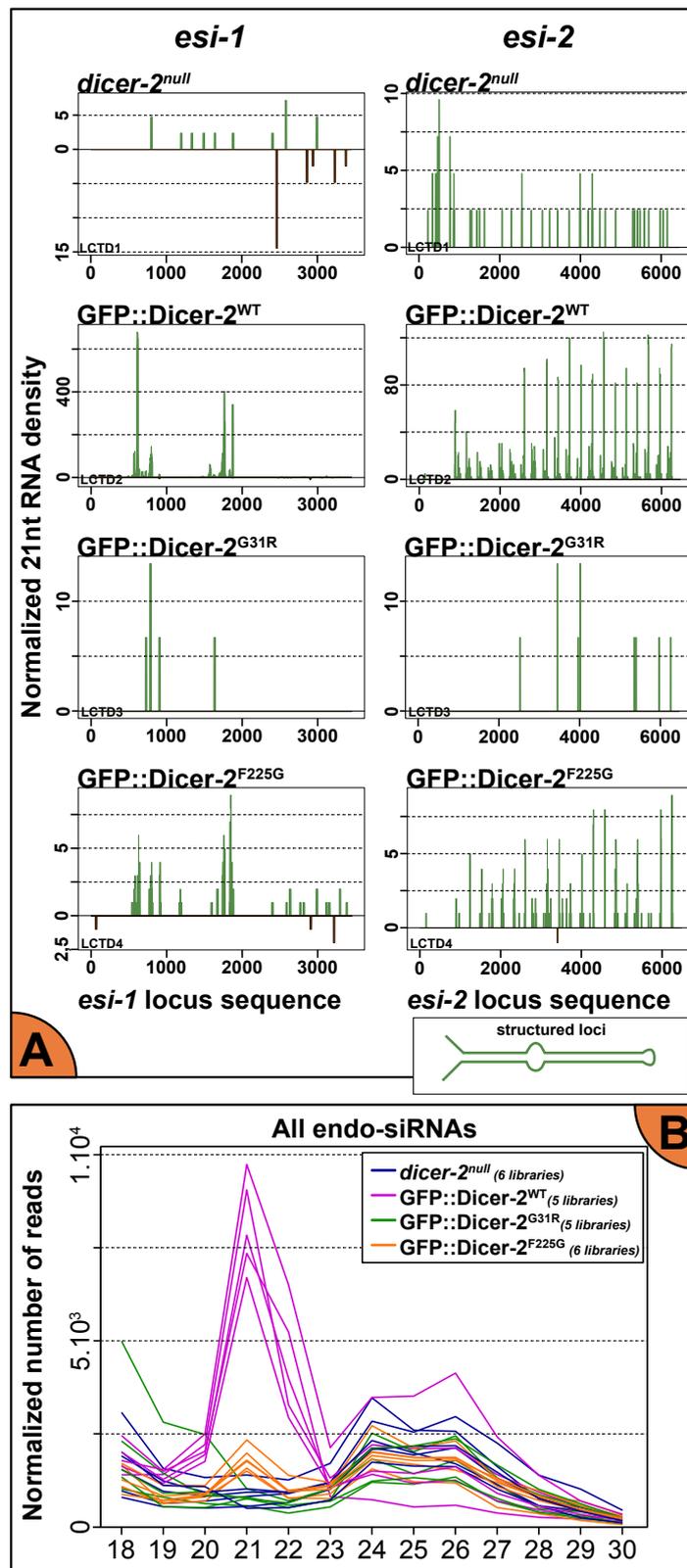


Figure 38: GFP::*Dicer-2^{F225G}* flies have an intermediary phenotype regarding the endo-siRNA pathway.

A) Normalized *esi-1* (left) or *esi-2* (right) sequence coverage by 21nt long reads. Shown plots are representative of all the plots obtained when looking at the data coming from flies with a similar genotype regardless of the injection performed. **B)** Normalized size distribution of endo-siRNA loci mapping reads. All results from flies with similar genotype are gathered under the same color. Each curve represents an individual library.

- An important peak at 21nt is observed in GFP::Dicer-2^{WT} flies for both *esi-1* and *esi-2* mapping reads. An additional 22nt peak of *esi-2* mapping reads can also be observed in these flies.
- Nearly no *esi-2* or *esi-1* mapping reads could be observed in *dicer-2^{null}* and GFP::Dicer-2^{G31R} flies.
- A small but detectable amount of these reads could be observed in GFP::Dicer-2^{F225G} with a peak at 21-22nt.

Because they are absent in *dicer-2^{null}* flies, we can conclude that these endo-siRNAs are Dicer-2 specific products as expected. We previously saw that GFP::Dicer-2^{WT} flies present a low level of *esi-2* and *esi-1* RNA precursors (**Figure 37A**). This result directly correlates with the high levels of corresponding endo-siRNAs. As expected, GFP::Dicer-2^{F225G} is much less efficient than GFP::Dicer-2^{WT} at producing endo-siRNAs from these 2 loci. Finally, as foreseen with the previous experiment, GFP::Dicer-2^{G31R} flies presents the same profile of reads distribution as *dicer-2^{null}* mutant flies suggesting its incapacity to process these endo-siRNA precursors.

Still little is known about the exact Dicer-2 processing mechanism of these endo-siRNA precursors. These long RNAs contain repeats that are predicted to fold in hairpin structures with undetermined extremities and generate phased small RNA duplexes (Czech et al., 2008b; Okamura et al., 2008b). Interestingly, our results indicate a different impact of the F225G and G31R helicase mutations on Dicer-2 processing of these RNAs. Indeed, while GFP::Dicer-2^{G31R} is barely able to produce *esi-1* or *esi-2* derived siRNAs, GFP::Dicer-2^{F225G} is still able to process these substrates but to a lower extent than GFP::Dicer-2^{WT}. Thus, in an attempt to understand where this difference in quantity of reads could come from, we plotted the distribution of 21nt long reads on *esi-1* and *esi-2* sequences (**Figure 38A**). The same observation that was made for w^{JR} mapping reads can be made here: a peculiar pattern of reads distribution is conserved between GFP::Dicer-2^{F225G} and GFP::Dicer-2^{WT} flies. The only difference lies in the number of reads mapping at each position. GFP::Dicer-2^{G31R} flies, however, present a very low number of reads which makes impossible the recognition of the distribution pattern. Thus, Dicer-2 mode of action on *esi-1* and *esi-2* precursors does not seem to be impacted by the F225G mutation. Again, the difference in amount of *esi-1* and *esi-2* mapping reads could come from a problem of accessibility of Dicer-2 to its substrate.

We wanted to check whether this observation was only true for *esi-2* and *esi-1* derived endo-siRNAs only or if it reflects a global trend of the endo-siRNA pathway. Thus, we looked at the size distribution of reads mapping on any endo-siRNA source, encompassing structured loci, cis-NATS and transposable elements (Ghildiyal et al., 2008; Okamura et al., 2008b) (**Figure 38B**). The same observations as the ones made for *esi-2* and *esi-1* loci can be made: (1) a huge peak at 21-22nt in GFP::*Dicer-2*^{WT} flies, (2) a smaller peak at 21-22nt in GFP::*Dicer-2*^{F225G} flies and (3) no peak at 21-22nt in *dicer-2*^{null} and GFP::*Dicer-2*^{G31R} flies. In conclusion, generation of the 21-22nt endo-siRNAs is *Dicer-2*-dependent, is strongly impacted (but still present) by the F225G mutation and is abolished by the G31R mutation. Of note, it could be interesting to individually study the different sources of endogenous siRNAs as they are not predicted to present the same characteristics, notably the same extremities.

It has to be noted that some of these endo-siRNA loci, especially transposable elements, are also targeted and controlled by the piRNA pathway in the gonads (germline and somatic support cells) of drosophila (Handler et al., 2013; Senti and Brennecke, 2010). This explains the broad peak at 24-30nt when looking at the size distribution of all endo-siRNA loci mapping reads (**Figure 38B**). This peak is quite homogenous between all libraries and is *Dicer-2*-independent.

***Dicer-2*^{F225G} is less efficient at binding its cofactor Loqs-PD**

All these results taken together led us to wonder about the implication of Loqs-PD in the GFP::*Dicer-2*^{F225G} phenotype. Indeed, GFP::*Dicer-2*^{F225G} flies present the same orange eye phenotype and the same accumulation of *esi-2* and *esi-1* precursors as *loqs* null mutant flies (Marques et al., 2010b). Moreover, Loqs-PD is a known cofactor of *Dicer-2*, which enhances its production of siRNAs from synthetic dsRNAs with blunt or suboptimal extremities (i.e 5' overhang) and from endogenously encoded dsRNAs (Marques et al., 2010b; Miyoshi et al., 2010; Sinha et al., 2015; Trettin et al., 2017; Zhou et al., 2009). Finally, interaction of Loqs-PD C-terminal domain to *Dicer-2* was previously mapped to its helicase domain (Hartig and Förstemann, 2011; Miyoshi et al., 2010). Thus, mutation of this phenylalanine 225 in the helicase domain of *Dicer-2* could interfere with its interaction with Loqs-PD.

We first took advantage of an in-house interactome of *Dicer-2* in adult flies to see whether we could detect this interaction *in vivo*. This interactome is based on immuno-

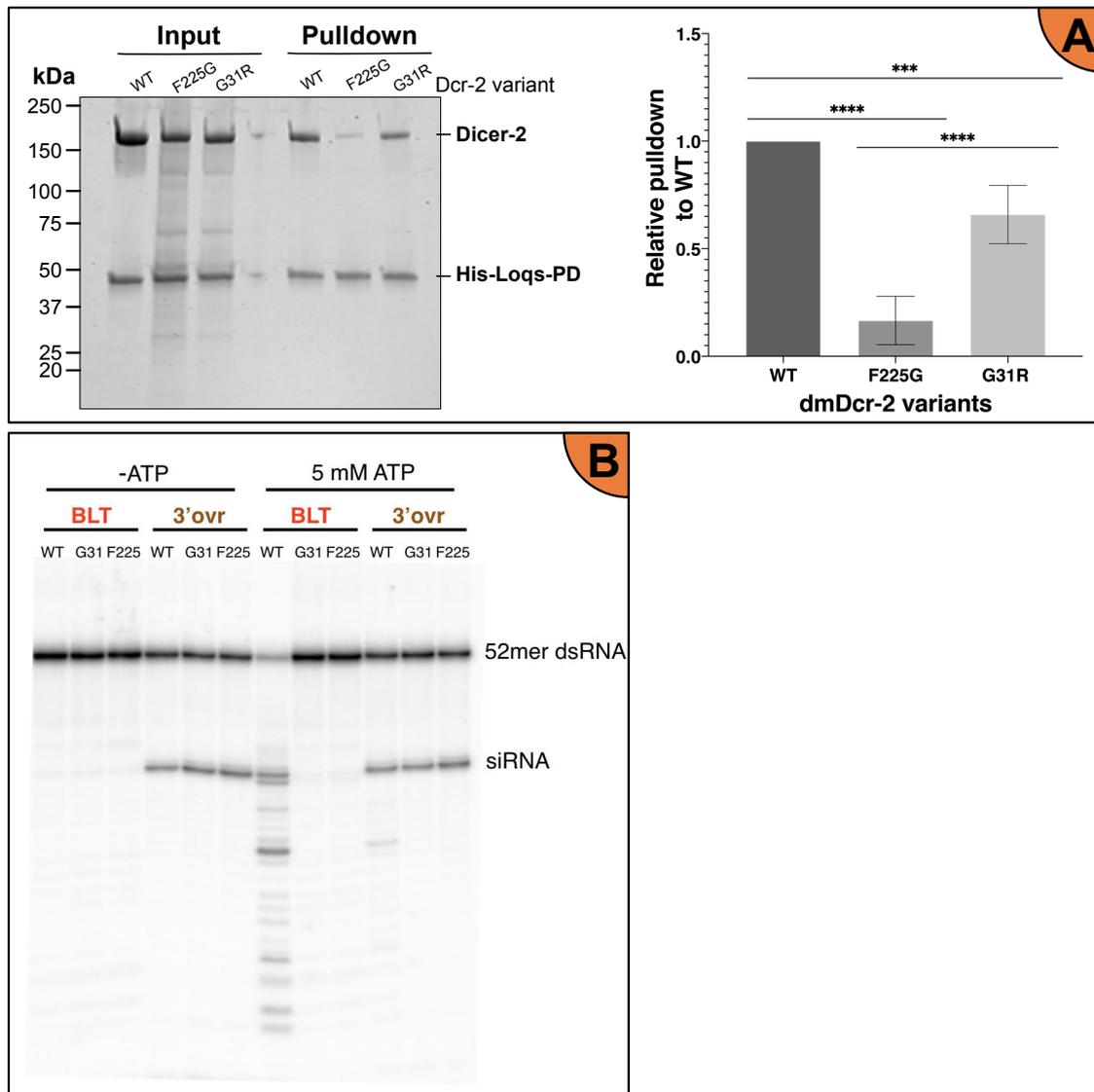


Figure 39: In vitro characterization of Dicer-2^{WT}, Dicer-2^{G31R} and Dicer-2^{F225G}.

A) His-Loqs-PD pulldown experiment. Left part of the figure shows a representative SDS PAGE Coomassie-stained gel with input and pulldown fractions. Bands intensities in the pulldown fractions were measured using ImageJ and plotted on the right part of the figure (n=3, error bars represent standard deviation between experimental replicates, statistical test used: unpaired T test). **B)** Acrylamid gel of a single-turnover cleavage experiment. **BLT**: blunt dsRNA; **3'ovr**: 2nt 3' overhang dsRNA; **WT**, **G31** and **F225**: Dicer-2^{WT}, Dicer-2^{G31R} and Dicer-2^{F225G} respectively.

precipitations of GFP-tagged Dicer-2 followed by mass spectrometry. While R2D2 always appears to be the #1 interactant of Dicer-2, Loqs-PD could never be found as an interactant (18 independent samples). This result is intriguing but could be explained by the fact that interaction between Loqs-PD and Dicer-2 might be very transient, occurring at specific stages of development or in specific tissues. Direct proofs of Loqs-PD interaction come from pull-downs of overexpressed and tagged Loqs-PD (Hartig and Förstemann, 2011). In such an experimental setup, a hypothetically transient interaction between Dicer-2 and Loqs-PD would be easier to catch. That is why we decided to test the interaction between Dicer-2^{F225G} and Loqs-PD *in vitro*. This experiment was done by the team of Pr. Brenda Bass already presented in Chapter II.

Briefly, different variants of Dicer-2 (WT, G31R or F225G) were produced, purified and incubated with purified His-Loqs-PD. Interaction between the two proteins was checked by performing His pulldown and SDS PAGE gel stained with Coomassie (**Figure 39A**). From this experiment it appears that (1) the F225G mutation dramatically decreases the interaction between Dicer-2 and Loqs-PD and (2) the G31R mutation only has a weak but significant effect on this interaction. These results have to be tempered by the fact that Dicer-2^{F225G} and Dicer-2^{G31R} purifications were not as clear as the Dicer-2^{WT} one. Indeed, products of intermediate size can be seen in the input of both these conditions and could interfere with Loqs-PD interaction. In conclusion, even though we only have indirect proof of it *in vivo*, these last *in vitro* data support our hypothesis stating that Dicer-2^{F225G} is not or less able to bind Loqs-PD. This lack of interaction may cause a less efficient processing of endogenous dsRNAs and explain the endo-siRNA related phenotypes observed in flies.

Dicer-2^{F225G} and Dicer-2^{G31R} process blunt dsRNA differently from GFP::Dicer-2^{WT}

Before going in a further *in vivo* characterization of the flies, we took advantage of having the three variants of Dicer-2 purified to perform a single turnover cleavage assay. With this experiment, we wanted to see if, beside their different interaction strength with Loqs-PD, Dicer-2^{G31R} and Dicer-2^{F225G} on their own would process different dsRNA substrates similarly. Basically, 52 nucleotides long dsRNAs bearing blunt or 2nt 3' overhang extremities were incubated with the three Dicer-2 variants in the presence or absence of ATP (**Figure 39B**). This experiment was done by the team of Pr. Brenda Bass according to the protocol of Sinha et al., 2018.

To begin with, 52nt blunt dsRNA cannot be processed into siRNAs by any of the three variants of Dicer-2 when no ATP is added into the reaction. This result is in accordance with previously published data using Dicer-2^{WT} (Sinha et al., 2015; Welker et al., 2011). Then, 3' overhang dsRNA is identically processed in a single siRNA band by the three variants of Dicer-2 in an ATP-independent manner. Dicer-2^{WT}, in presence of ATP, processes blunt and 3' overhang dsRNA but produces different cleavage patterns. Indeed, while a single siRNA band is observed with the 3' overhang dsRNA substrate, a multitude of shorter and longer products can be observed in addition to the siRNA band when Dicer-2^{WT} is incubated with the blunt dsRNA. This result, once again, fits with the previously proposed model of differential Dicer-2 mode of action depending on the substrate extremity (Sinha et al., 2018). In this model, blunt dsRNA molecules are threaded through the helicase domain in an ATP-dependent manner which can result in different sizes of cleavage products. On the contrary, 3' overhang dsRNAs are bound by the PAZ-platform domain and get precisely cut when approaching the RNaseIII domains of Dicer-2. This could explain why both helicase mutations have no impact on the processing of 3' overhang dsRNA substrate. Finally, both Dicer-2^{G31R} and Dicer-2^{F225G} are unable to process blunt dsRNA. Therefore, G31R and F225G mutations of the helicase domain have no impact on the processing of 3' overhang dsRNA but inhibit cleavage of blunt dsRNA. Further *in vitro* characterization of Dicer-2 variants interaction with their substrates should be done and will be discussed later.

In conclusion, study of the F225G and G31R mutations of Dicer-2 helicase revealed that they have a different impact on Dicer-2 substrate processivity *in vivo*. Indeed, while both variants of Dicer-2 are impaired for the production of endo-siRNAs from diverse source, it appears that GFP::Dicer-2^{F225G} still retains the ability to dice those substrates but in a much less efficient way than GFP::Dicer-2^{WT}. In the case of Dicer-2^{F225G}, this decrease in quantity of endo-siRNAs may be linked to a less efficient binding of its cofactor Loqs-PD. However, interaction of Dicer-2^{G31R} with the same cofactor was only mildly affected. In addition, purified Dicer-2^{F225G} and Dicer-2^{G31R} behave similarly on synthetic dsRNAs *in vitro*. Thus, differences in phenotypes between GFP::Dicer-2^{F225G} and GFP::Dicer-2^{G31R} flies remains unexplained but hypotheses will be discussed later.

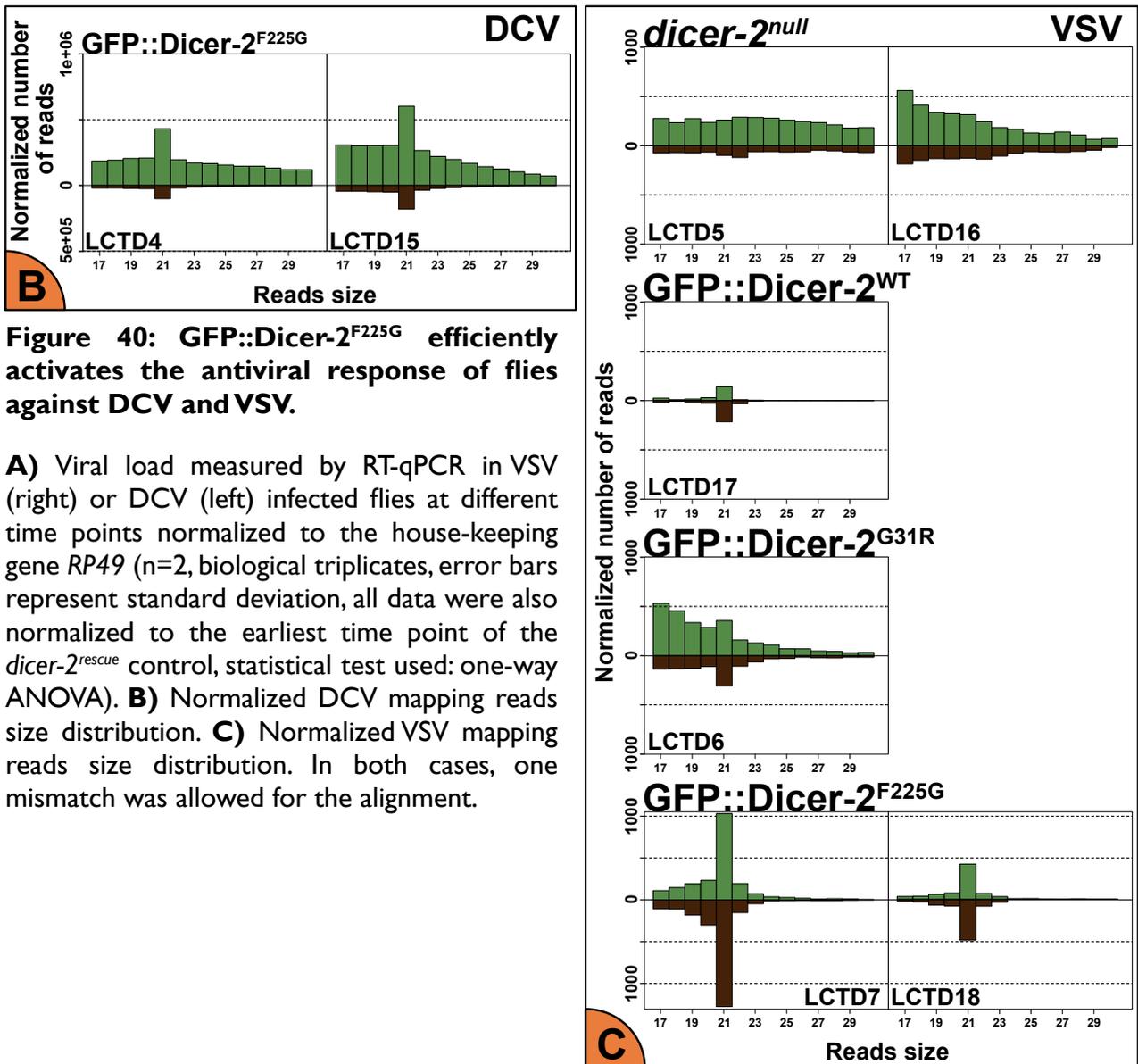
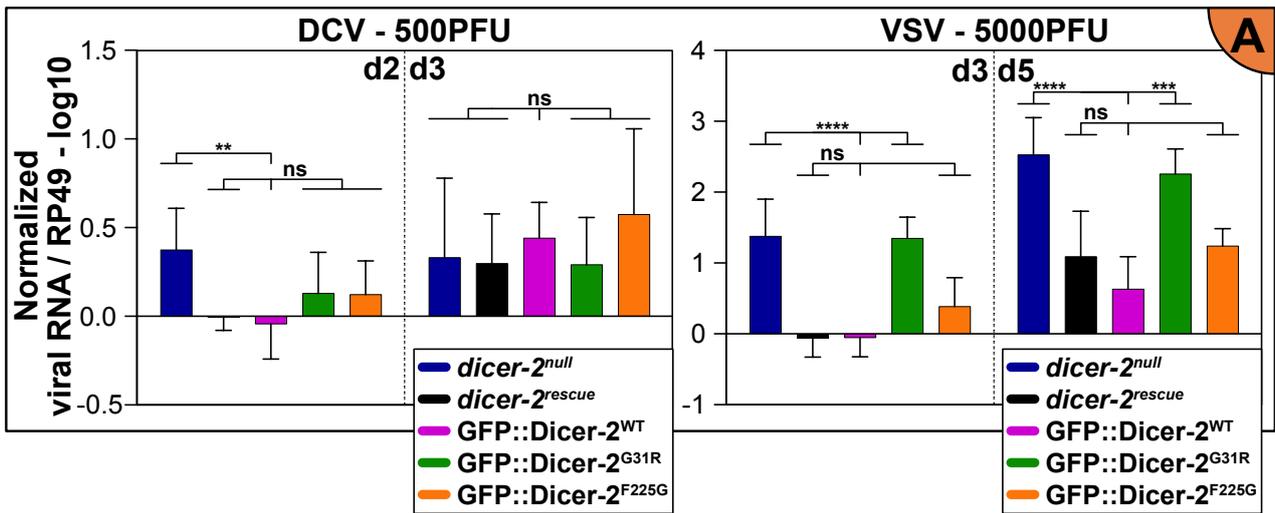


Figure 40: GFP::*Dicer-2*^{F225G} efficiently activates the antiviral response of flies against DCV and VSV.

A) Viral load measured by RT-qPCR in VSV (right) or DCV (left) infected flies at different time points normalized to the house-keeping gene *RP49* (n=2, biological triplicates, error bars represent standard deviation, all data were also normalized to the earliest time point of the *dicer-2*^{rescue} control, statistical test used: one-way ANOVA). **B)** Normalized DCV mapping reads size distribution. **C)** Normalized VSV mapping reads size distribution. In both cases, one mismatch was allowed for the alignment.

II. Impact of two Dicer-2 helicase mutations on the antiviral siRNA pathway

Interestingly and in a way that we still do not understand, the antiviral siRNA pathway is independent of Loqs-PD (Marques et al., 2013). Thus, a lack of interaction between Dicer-2 and Loqs-PD should not impact the flies' defense against viruses. However, the F225G mutation was identified by comparison with the RIG-I CTD, which is responsible for the sensing of dsRNA extremities (Luo et al., 2011; Sinha et al., 2018). One could imagine that such a mutation would impair the Dicer-2 helicase-mediated sensing of viral dsRNA. Therefore, antiviral activity of GFP::Dicer-2^{F225G} flies was investigated.

GFP::Dicer-2^{F225G} efficiently activates the antiviral response of infected flies against DCV and VSV

Impact of the F225G mutation of Dicer-2 on the antiviral RNAi pathway was first investigated by performing injections of viruses in adult flies and monitoring of viral loads. We used flies from the previously described genotypes injected with either DCV or VSV. As explained in Chapter I, DCV is a dicistrovirus with a positive ssRNA genome while VSV belongs to the *Rhabdoviridae* family and has a negative ssRNA genome. It has been decided to use those two viruses because of their difference in term of transcription and replication strategies. Indeed, DCV produces a single long RNA encompassing both of its ORFs and the required structures for its translation (IRES and IGR). The viral dsRNA intermediate of replication is thought to have blunt extremities with a 5'-linked VPg at both extremities. On the contrary, VSV transcribes its genome into several mRNAs that are all independently capped and polyadenylated. Replication is primed by 5' triphosphate leader and trailer RNAs and probably give rise to dsRNA intermediates with 5' triphosphate blunt dsRNA. Thus, studying these two very different viruses might highlight common and unique features of the corresponding antiviral response and Dicer-2 sensing.

To begin with, the viral load of DCV or VSV injected flies was checked at different time point of the infection by RT-qPCR. For DCV, 500PFU were injected in each fly and their viral load was checked after 2 or 3 days of infection (*Figure 40A – left*). For VSV, 5000PFU were injected in each fly and their viral load was checked after 3 or 5 days of infection (*Figure 40A – right*). This difference in parameters used is due to the fact that VSV is not a natural drosophila pathogen and thus, is less efficient at replicating in flies

than the drosophila specific virus DCV. Moreover, VSV does not encode for a viral suppressor of RNAi (VSR) which makes it more sensitive to the antiviral RNAi.

First, no significant difference could be identified between viral loads of DCV or VSV infected GFP::Dicer-2^{WT} and *dicer-2^{rescue}* flies. This explains why all comparisons will be made with GFP::Dicer-2^{WT} flies as a point of reference. Then, one can notice an increase in viral load between VSV d3 and d5 pi regardless of the flies' genotype, which suggests an ongoing productive viral infection. This increase is not observed in DCV infected flies and shows that some flies already reached a plateau of infection 2d pi.

In VSV infected flies and for both time points, a significantly higher viral load is detected in *dicer-2^{null}* and GFP::Dicer-2^{G31R} flies in comparison to GFP::Dicer-2^{WT}. This result emphasizes the already known importance of Dicer-2 and its helicase domain in the antiviral response of drosophila (Deddouche et al., 2008; Marques et al., 2013; Mueller et al., 2010; Sabin et al., 2013). Interestingly, the F225G mutation, like the G31R one, is located in the helicase domain of Dicer-2 but does not seem to have a significant impact on VSV or DCV viral load. As explained in the introduction, this mutation has been predicted to inhibit binding of Dicer-2 to blunt dsRNA and to have no effect on 3'overhang dsRNA substrate processing (Sinha et al., 2018). The same study proposes that substrates can either be bound through the helicase domain or through the PAZ-platform domain depending on the dsRNA extremity. Thus, the difference between GFP::Dicer-2^{F225G} and GFP::Dicer-2^{G31R} phenotypes could come from a divergent mode of action on the viral dsRNA intermediate of replication that would ultimately result in a difference in amount of virus derived siRNAs.

In order to check for the production of viral siRNAs, we once again looked at the small RNA HTS data obtained from DCV (3d pi) or VSV (5d pi) infected flies. In the case of DCV infected flies, we can see that GFP::Dicer-2^{F225G} produces a wild-type number of 21nt long mapping reads (*Figure 40B* & Chapter I – *Figure 15A*). On the contrary, as explained in Chapter I, GFP::Dicer-2^{G31R} flies present a number of 21nt long DCV mapping reads really low when considering degradation products curve. Thus, the viral load data directly correlate with the amount of virus derived 21nt long reads generated by each Dicer-2 variant.

In the case of VSV infected flies, size distribution of virus mapping reads (*Figure 40C*) revealed that:

- *dicer-2^{null}* and GFP::*Dicer-2^{G51R}* flies present a significant number of VSV reads longer or shorter than 21nt. These products present a similar signature to the degradation products identified in DCV infected flies and are *Dicer-2*-independent. They once again highlight the fact that in addition to the very efficient *Dicer-2*-mediated antiviral RNAi pathway, viral RNAs can be targeted by host degradation pathways.
- GFP::*Dicer-2^{WT}* flies have a low number of VSV mapping reads but only present a peak at 21nt and no apparent degradation products. This low number of reads was already previously observed and could be explained by the fact that *Dicer-2^{WT}* expressing flies are able to mount an efficient antiviral response, decreasing the number of *Dicer-2* targets (Mueller et al., 2010).
- In addition to the degradation product signature, GFP::*Dicer-2^{G51R}* flies present a peak at 21nt. Thus, the total number of 21nt long reads is composed of GFP::*Dicer-2^{G51R}* products but also of the abundant degradation products. That is why it is hard to compare the *Dicer-2* specific products between GFP::*Dicer-2^{WT}* and GFP::*Dicer-2^{G51R}* flies.
- GFP::*Dicer-2^{F225G}* have mostly 21nt long VSV mapping reads and a few shorter or longer reads. From this plots, it even appears that these flies produce more VSV derived siRNAs than GFP::*Dicer-2^{WT}* flies.

In conclusion, the observed viral load differences for both viruses are directly correlated with the amount of virus derived siRNAs that can be identified by small RNA HTS. GFP::*Dicer-2^{G51R}* flies produce less virus derived siRNAs than GFP::*Dicer-2^{WT}* flies, hence their higher susceptibility to VSV. The same conclusion can be drawn from DCV data even though viral load results are less clear (reached plateau of infection). On the contrary, GFP::*Dicer-2^{F225G}* flies behave as GFP::*Dicer-2^{WT}* flies regarding viral load and amount of virus derived siRNAs. Therefore, these two mutations located in the helicase domain of *Dicer-2* do not impact its function in a similar manner. Hypotheses regarding the mechanisms at play behind these mutations will be discussed later.

No peculiar distribution of the virus derived reads could be observed when looking at the libraries constructed with VSV infected samples (data not shown). This is why these

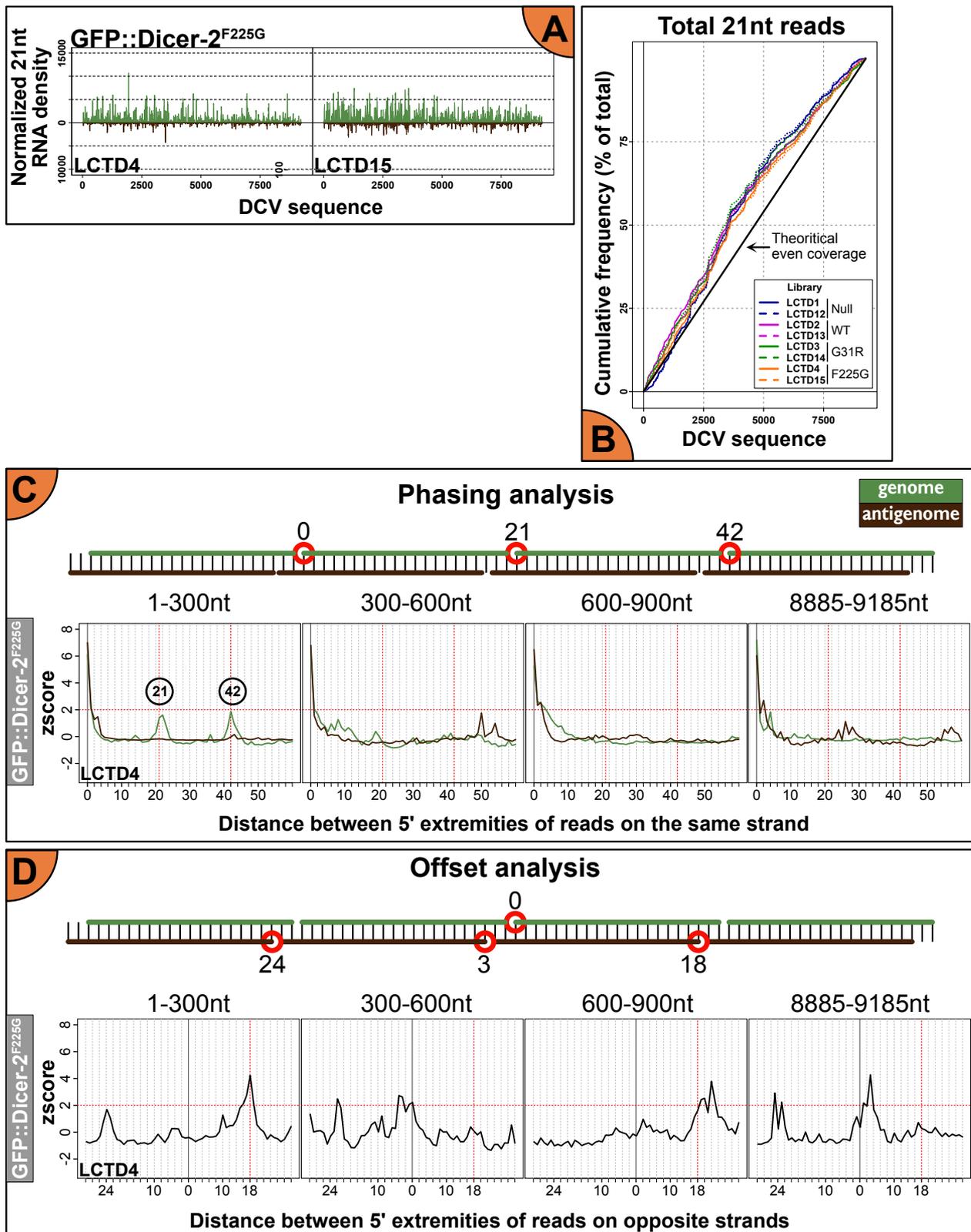


Figure 4I: A Dicer-2 signature can be identified in the 5' region of DCV in GFP::Dicer-2^{F225G} flies.

- A)** Normalized DCV sequence coverage by 21nt long reads. **B)** Cumulative frequency of DCV coverage by 21nt long reads. The black line represents what would be a perfectly even coverage of the sequence. **C)** Phasing analysis of I library representative of the corresponding genotype. **D)** Offset analysis of I library representative of the corresponding genotype.

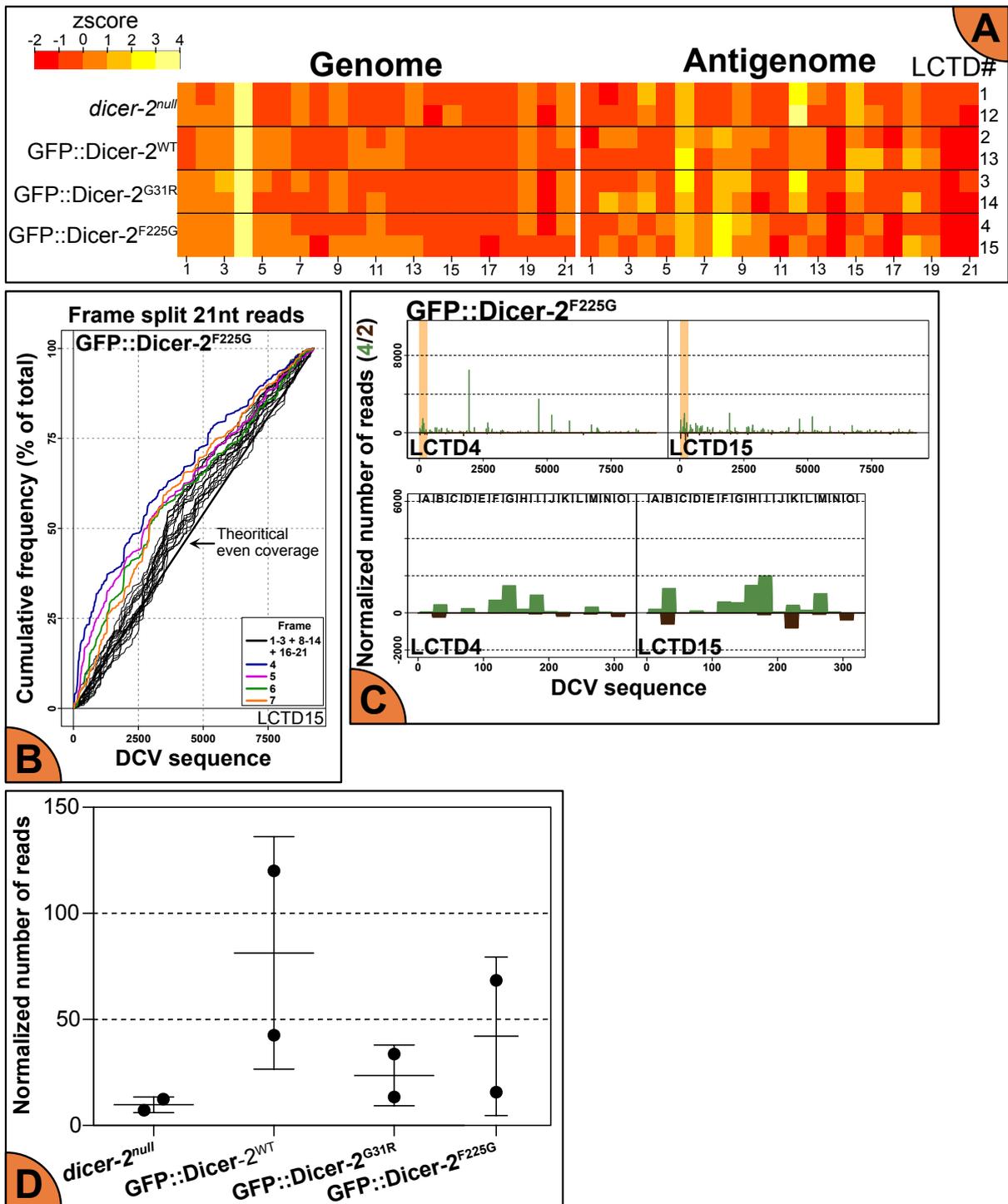


Figure 42: GFP::*Dicer-2^{F225G}* has a precise entry point on viral dsRNA in the 5' region of DCV corresponding to the domain I.

A) Frame enrichment calculated individually for DCV genome and antigenome mapping 21nt long reads. **B)** Cumulative frequency of DCV coverage by independent frames. **C)** Normalized DCV coverage by 21nt long reads in frame 4/2. **D)** Normalized number of the specific 19nt long read mapping at position 25-43 on the antigenome of DCV. Error bars represent the standard deviation between the two libraries.

results will mostly be kept for the discussion. However, because of the apparent similarity of the antiviral response of GFP::Dicer-2^{WT} and GFP::Dicer-2^{F225G} flies, we decided to investigate whether the same entry point of Dicer-2 on DCV dsRNA could be identified in GFP::Dicer-2^{F225G} flies.

GFP::Dicer-2^{F225G} has a precise entry point on viral dsRNA in the 5' region of DCV

The same procedure as the one explained in Chapter I was followed to analyze the small RNA HTS data of DCV infected GFP::Dicer-2^{F225G} flies. Plots can directly be compared with the ones presented in Chapter I as the same scale was always used.

Once again, distribution of all the 21nt long reads on DCV sequence shows no specific region enrichment in GFP::Dicer-2^{F225G} flies (*Figure 41A e B*). Phasing and offset analyses reveal a strong siRNA signature in the very 5' of DCV sequence that quickly disappears the further we go from the genomic 5' extremity (*Figure 41C e D*). Then, study of the frame enrichment of 21nt long reads revealed that frame 4/2 and subsequent frames 5/3 and 6/4 are enriched in the 5' region of DCV roughly corresponding to the domain I (*Figure 42A e B*). Finally, an identical distribution of frame 4/2 reads in this region as well as the presence of the very peculiar 19nt long read at position 25-45 makes the GFP::Dicer-2^{F225G} phenotype identical to the GFP::Dicer-2^{WT} one (*Figure 42C e D*). In conclusion, the same specific entry point on DCV domain I dsRNA can be identified in GFP::Dicer-2^{WT}, GFP::Dicer-2^{G51R} and GFP::Dicer-2^{F225G} flies. Thus, the difference in subsequent antiviral RNAi pathway activation probably lies in the number of following siRNAs generated and loaded on Ago2.

Conclusions – Discussions – Perspectives

The helicase domain of drosophila Dicer-2 was previously proposed to have several central roles in its mechanism both *in vitro* and *in vivo*:

- Interaction with Loqs-PD and R2D2 cofactors (Hartig and Förstemann, 2011; Miyoshi et al., 2010; Nishida et al., 2013; Trettin et al., 2017). It is not clear yet whether these proteins compete for binding to Dicer-2 but interaction between Loqs-PD and Dicer-2 was only detected by performing immunoprecipitation of overexpressed and tagged Loqs-PD while R2D2 is always found as #1 interactant in mass spectrometry analyzes (in-house Dicer-2 interactome).
- Binding and hydrolysis of ATP that could fuel unwinding activity of dsRNA substrate and/or Dicer-2 conformational change (Sinha et al., 2015, 2018; Welker et al., 2011).
- Sensing and threading of blunt dsRNA substrate. These roles of Dicer-2 helicase were suggested since a while but direct evidences of it were just recently obtained by using cryo-electron microscopy (Sinha et al., 2018).

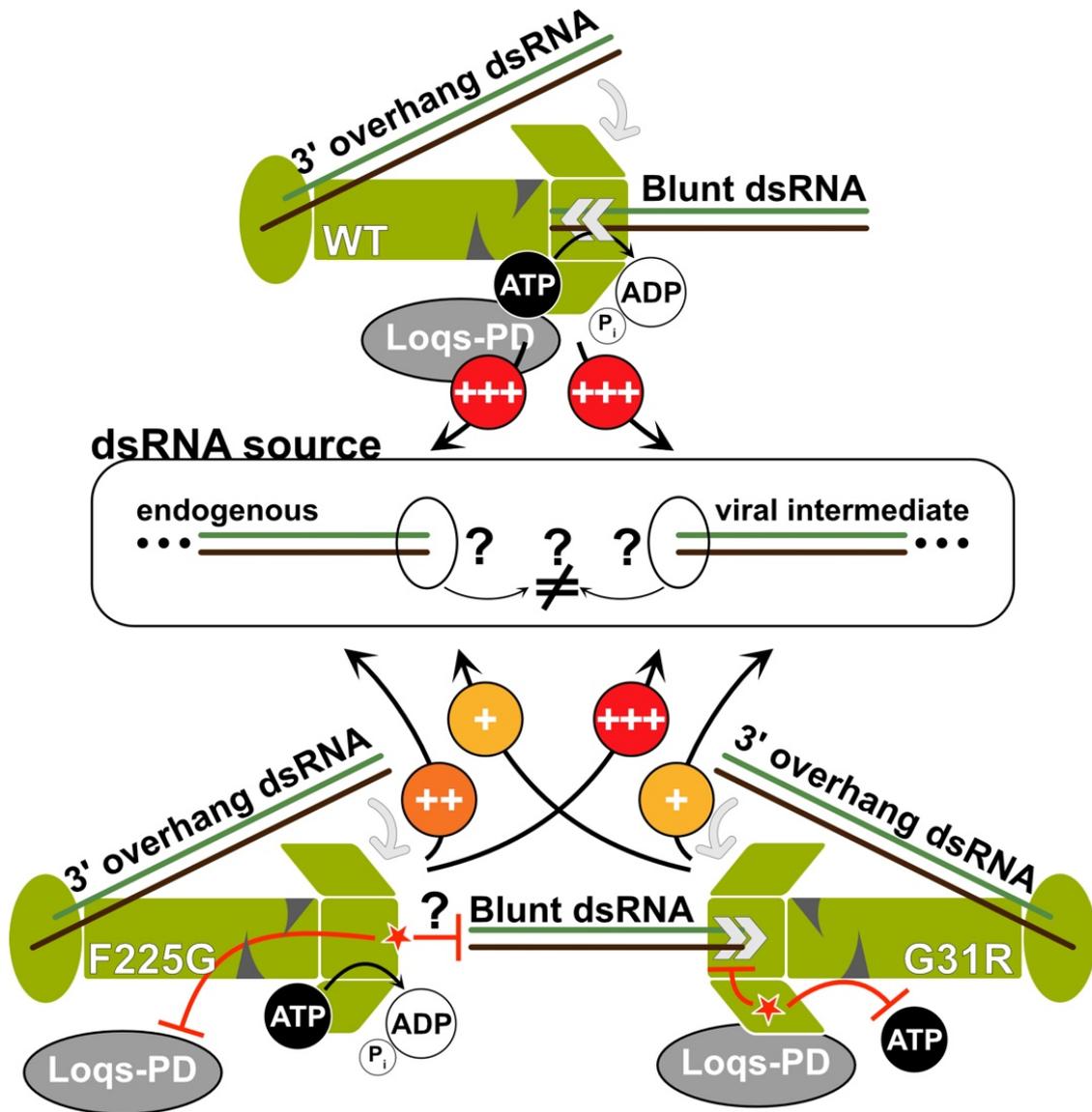
Of note, human Dicer was previously shown to be more efficient at substrate processing *in vitro* when depleted from its helicase domain suggesting an autoinhibition role of this domain on Dicer activity (Ma et al., 2008). Thus, in this case, Dicer helicase is not responsible for the sensing of its substrate.

The work conducted in this chapter is the fruit of a close collaboration with the team of Pr. Brenda Bass and aimed at better characterizing Dicer-2 helicase role in its endogenous and antiviral functions. To do so, we combined the *in vivo* study of transgenic flies expressing different Dicer-2 variants with *in vitro* experiments conducted with recombinant Dicer-2 proteins. Two point mutations of the helicase domain were used for this study:

- Dicer-2^{G31R} was shown to be unable to bind ATP, to strongly decrease siRNA production but to retain its ability to transfer siRNAs to Ago2 (Cenik et al., 2011; Förstemann et al., 2007; Fukunaga et al., 2014; Lee et al., 2004).
- Dicer-2^{F225G} is still able to process 2nt 3' overhang dsRNA but cannot dice blunt dsRNA either because of an impaired sensing activity or because of a strong conformational change in the helicase domain (Sinha et al., 2018).

The results obtained in this chapter as well as the hypotheses I will discuss can be summed up in the following table and figure (with *in vivo*: Null = *dicer-2^{null}*; WT = GFP::*Dicer-2^{WT}*; G31R = GFP::*Dicer-2^{G31R}*; F225G = GFP::*Dicer-2^{F225G}* and *in vitro*: WT = *Dicer-2^{WT}*; G31R = *Dicer-2^{G31R}*; F225G = *Dicer-2^{F225G}*):

	Null	WT	G31R	F225G	Dicer-2 variant
Flies charac.	-	++	+++	+++	<i>dicer-2</i> mRNA
	-	+++	+++	+++	Dicer-2 protein
	RED	WHITE	RED	ORANGE	Color of the eye
endo siRNA pathway	-	+++	+	++	wIR derived siRNAs
	+++	-	+++	+++	esi-1 RNA precursor
	+++	-	+++	+	esi-2 RNA precursor
	-	+++	-	++	esi-1 derived siRNAs
	-	+++	-	+	esi-2 derived siRNAs
	-	+++	-	+	All endo siRNAs
Antiviral pathway	+++	+	+++	+	DCV viral load
	+++	+	+++	+	VSV viral load
	-	+++	+	+++	viral derived siRNAs
	N/A	YES	YES	YES	Precise entry of Dicer-2 in DCV domain I
<i>In vitro</i> charac.	N/A	+++	++	+	Interaction with Loqs-PD
	N/A	-	-	-	BLT dsRNA (-) ATP
	N/A	+++ (21)	+++ (21)	+++ (21)	3' overhang dsRNA (-) ATP
	N/A	+++ (~size)	-	-	BLT dsRNA (+) ATP
	N/A	+++ (21)	+++ (21)	+++ (21)	3' overhang dsRNA (+) ATP



- a. How to explain the lack of differences between Dicer-2^{G31R} and Dicer-2^{F225G} processing *in vitro*?

By performing *in vitro* cleavage assays, Dicer-2^{G31R} and Dicer-2^{F225G} revealed themselves to behave identically regardless of the substrate provided. However, as previously stated, both mutations are not supposed to have similar impact on Dicer-2 functions. Using the current model of Dicer-2 action depending on dsRNA substrate extremities (Sinha et al., 2018), the following hypotheses could be made:

- Processing of 2nt 3' overhang is not impaired and remains ATP-independent in Dicer-2^{G31R} and Dicer-2^{F225G} as Dicer-2 helicase domain is not implicated in the processing of such a substrate.
- It was shown that the F225G mutation in Dicer-2 helicase inhibits binding of blunt dsRNA substrate. This would explain why such a substrate is not processed by Dicer-2^{F225G}.
- On the other hand, Dicer-2^{G31R} should still be able to bind blunt dsRNA. The explanation for the lack of processivity of this substrate could be due to the impaired ATP binding and hydrolysis that should fuel the dsRNA threading through the helicase domain. Thus, a blunt dsRNA would remain stuck in the helicase domain of Dicer-2^{G31R} and could not be processed. *In vitro* binding assays of Dicer-2^{G31R} to blunt dsRNA should be done to confirm this hypothesis.

b. What is the role of Loqs-PD?

One puzzling observation lies in the differences of endo siRNA and antiviral RNAi pathways activities detected between Dicer-2^{G31R} and Dicer-2^{F225G} *in vivo*. Dicer-2^{F225G} behaves as Dicer-2^{WT} regarding the antiviral RNAi pathway but shows a diminished endo siRNA pathway activity. On the other hand, both roles of Dicer-2 are severely impaired upon G31R mutation. One crucial determinant of these differences could be the interaction between Dicer-2 and Loqs-PD. Indeed, *in vitro*, Dicer-2^{G31R} is still able to bind Loqs-PD while this interaction is strongly diminished with Dicer-2^{F225G} variant. In addition, data obtained in GFP::*Dicer-2^{F225}* flies recapitulate the phenotypes previously observed in *loqs^{null}* flies (Marques et al., 2010b).

Taking into account the literature and the recently obtained data, a role for Loqs-PD in the sorting of dsRNA substrates of Dicer-2 could be proposed. A similar role for the TRBP cofactor of human Dicer was already shown. In this case, TRBP acts as a gatekeeper for hDicer, inhibiting the targeting of cellular RNAs other than pre-miRNAs (Fareh et al., 2016). Furthermore, drosophila R2D2 was shown to inhibit Dicer-2 processing of Dicer-1 pre-miRNA targets (Cenik et al., 2011). Thus, Loqs-PD could be able to discriminate endogenous sources of siRNAs from other cellular RNAs and viral intermediates of replication and present them to Dicer-2 helicase domain. One suggestion that could be made to explain the very transient interaction between Loqs-PD and Dicer-2 would be that Loqs-PD primary binding is not with Dicer-2 but with the endogenous

dsRNA. Once bound to the dsRNA target, Loqs-PD would traffic up to Dicer-2 and quickly interact with its helicase domain to transfer its load. In this model, Loqs-PD is not stably associated with Dicer-2, which could explain why their interaction was not detected *in vivo*. However, we still cannot understand how Loqs-PD could recognize its targets: through their extremities, secondary structures or maybe post-transcriptional modifications? Of note, *in vitro* Loqs-PD was shown to enhance Dicer-2 processivity of blunt dsRNA substrates suggesting that endogenous targets may be recognized as blunt substrates and threaded through the helicase (Trettin et al., 2017). In order to check this model, it could be interesting to investigate Loqs-PD RNA interaction by performing Loqs-PD immunoprecipitations followed by northern blot on endo siRNA precursors such as *esi-1* and *esi-2*. This experiment, if performed in *dicer-2^{null}* or GFP::*Dicer-2^{F225G}* flies should yield a majority of full length precursors in comparison to GFP::*Dicer-2^{WT}* flies. Alternatively, the recent advances in single-molecule fluorescence could be used to check for the binding constants of Dicer-2 – Loqs-PD and endo siRNA precursors – Loqs-PD.

c. How can I reconcile *Dicer-2^{G31R}* and *Dicer-2^{F225G}* phenotypes with the model?

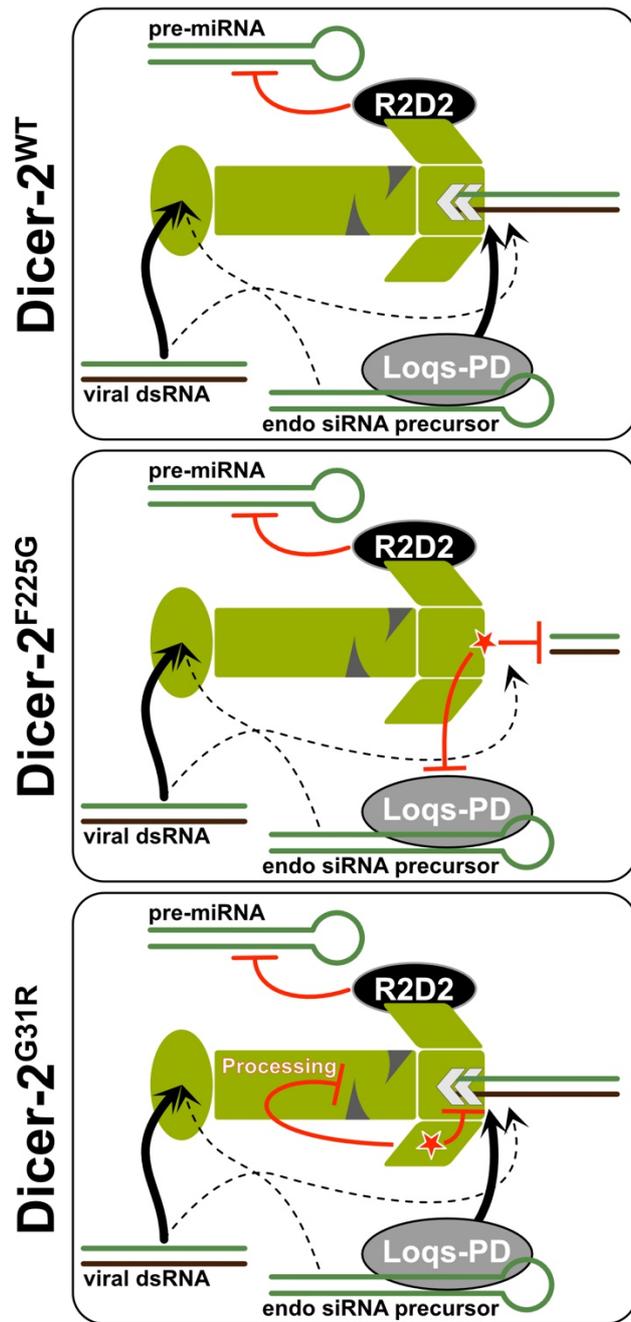
Dicer-2^{F225G}

In the case of *Dicer-2^{F225G}*, the answer is quite straight-forward: the processing of viral dsRNA intermediates does not require the action of Loqs-PD (Marques et al., 2013), which explains why the *Dicer-2* – Loqs-PD disrupting mutation *F225G* has no effect on viral load or virus derived siRNAs amount. Interestingly, this model coupled to the similar sequencing data obtained in DCV infected *Dicer-2^{WT}* and *Dicer-2^{F225G}* flies suggests that recognition of viral dsRNA could be done not by the helicase domain but by the PAZ-platform domain. This hypothesis goes against the inherent model stating that a processive action of *Dicer-2* is responsible for the generation of a high number of virus-derived siRNAs and favors a distributive mode of action of *Dicer-2* on its viral dsRNA substrate. In order to investigate this, we are currently generating flies expressing a *Dicer-2* variant mutated in its phosphate binding pocket of the PAZ domain (*Dicer-2^{PP}*, Kandasamy and Fukunaga, 2016). This mutant was shown to have an impaired processivity of 2nt 3' overhang dsRNA but retains a wild-type processivity of blunt dsRNA (Sinha et al., 2018). If our hypotheses are true, *Dicer-2^{PP}* expressing flies should present a higher viral load, a reduced survival rate and a reduced number of virus derived siRNAs in comparison to *Dicer-2^{WT}* or *Dicer-2^{F225G}* flies.

When looking at the endo siRNA pathway, Dicer-2 was shown to process its targets with the help of Loqs-PD (Marques et al., 2013; Miyoshi et al., 2010). Thus, if Loqs-PD works as a “target delivery cofactor” for Dicer-2, the absence of interaction between these two proteins would result (as observed in Dicer-2^{F225G} flies) in a diminished number of endo siRNAs. The sequencing of Loqs-PD bound RNAs previously proposed could show whether it is able to bind endo siRNA precursors independently of its interaction with Dicer-2. Though, it is not clear how Dicer-2^{F225G} would still be able to produce a detectable amount of endo siRNAs. We cannot exclude the fact that it could still be able to encounter an endogenous dsRNA target by chance and process it using its PAZ-platform domain as sensor. Alternatively, *in vitro* binding assays revealed a strongly impaired but not abolished binding of Loqs-PD to Dicer-2^{F225G}. Thus, the remaining interacting proteins could still deliver a reduced number of endogenous targets to Dicer-2.

Dicer-2^{G31R}

Dicer-2^{G31R} is still able to bind to Loqs-PD, which should enable its processing of endogenous targets. However, in our model, Loqs-PD will bring the endogenous targets of Dicer-2 to a helicase domain unable to hydrolyze ATP and to thread the dsRNA substrate. This will result in a stalled Dicer-2^{G31R} and probably in a conformational change of the helicase domain. Such a protein could still be able to bind viral or endogenous dsRNA using its PAZ-platform domain, but a steric obstruction effect of the helicase domain would inhibit the subsequent dicing mediated by the RNase III domains. One way to test this hypothesis would be to perform binding and dicing assay where the different Dicer-2 variants would have first been incubated with blunt dsRNA followed by a second incubation with 2nt 3' overhang dsRNA. As previously mentioned, a Dicer-2^{PP} should also be used in these experiments to discriminate helicase mediated binding from PAZ-platform one.



d. What did we learn about viral RNA sensing?

This study did not allow the deciphering of the exact mechanism by which Dicer-2 senses viral RNAs. The predominant model, based on RIG-I mechanism and *in vitro* data, proposes that Dicer-2 senses and discriminate the extremities of dsRNA molecules. Thus, it was suggested that a similar mechanism must be at play *in vivo*. The major limitation I encountered during my Ph.D was the lack of information regarding the extremities of the viruses I studied. By using two different viruses (DCV and VSV) with expected different extremities (VPg and 5' PPP respectively), we expected to see genotype-related

differences in the sequencing data that could reflect distinct recognition mode of Dicer-2. However, virus derived siRNAs were efficiently produced upon DCV and VSV infection in Dicer-2^{WT} and Dicer-2^{F225G} conditions. This result suggests that the viral extremities might not be the primary targets of Dicer-2 *in vivo* and that the helicase domain might not be the virus sensing domain of Dicer-2.

In order to investigate Dicer-2 sensing domain and entry point on its *bona fide* targets, a CLIP-seq approach should be developed. The perspective of this experiment is the reason why all complemented flies are expressing a GFP N-terminally fused to Dicer-2. Indeed, available Dicer-2 antibodies are not specific enough to be used for this method. In addition, flies expressing RNAseIII mutants' variants of Dicer-2 were already generated in the lab in order to stall Dicer-2 on its entry point and make its identification easier.

As a personal side note, not having the time to start this project during my Ph.D might be my biggest regret as I am sure that it will bring a lot of answers but also, maybe, many more questions.

Concluding remarks

All the conclusions regarding the results obtained during my Ph.D were already presented in the corresponding chapters. Still, the following points can be considered as a take home message from this manuscript:

- Small RNA HTS of DCV and CrPV infected cells and flies highlighted an early, precise and ATP-independent entry of Dicer-2 on the viral dsRNA corresponding to the domain I of dicistroviruses.
- *In vitro* study of DCV domain I allowed the modeling of its 2D structure and the characterization of its sensitivity to host endonucleolytic cleavage.
- A RNAi-based screen in S2 cells identified 20 candidate genes presenting an impact on DCV viral load upon knock-down that are currently being validated.
- The *in vivo* study of two different helicase point mutations of Dicer-2 allowed me to refine the previously proposed model of Dicer-2 mechanism.

The small RNA sequencing data I obtained still need to be further investigated and coupled with other methods mentioned in the discussions in order to check for validity of our proposed hypotheses.

At the time I am writing this, important questions remain unanswered: (1) What are all the drosophila Dicer-2 targets *in vivo* and are they all processed in siRNAs? (2) What are the exact characteristics of these targets and how are they discriminated between endogenous siRNA precursors, viral intermediates of replication and other cellular RNAs? (3) What is the role of Loqs-PD in this sorting mechanism? (4) Where is the entry point of Dicer-2 on its substrates *in vivo*?

Finally, the work presented in this manuscript was highly multidisciplinary and required the individual expertise of many different teams and people. Thus, I am more than ever convinced that the answers to all these questions can only arise from strong collaborations between *in vitro* and *in vivo* people, molecular biologists, virologists, bioinformaticians...

Materials & Methods

Small RNA High Throughput Sequencing

a. Preparation of RNA samples

For S2 cells' small RNA HTS

Schneider 2 (S2) cells were grown in Schneider medium (Biowest) complemented with 10% heat inactivated fetal bovine serum, 2mM glutamax, 100U/mL penicillin and 100µg/mL streptomycin (Life technologies). The cells are regularly passed twice a week at a ¼ dilution.

First, 250µl of $2 \cdot 10^6$ cells/ml ($5 \cdot 10^5$ cells) are seeded in 24 flat wells plates. After 30min, plates are put on ice and in the cold room for 30min. Medium is then removed from the wells and 250µl of cold infective solution (DCV or CrPV virus stock diluted to the appropriate MOI in complemented medium) are dispensed. The MOIs used for this screen are of 10 for DCV and 0,05 for CrPV. Cells are kept on ice for 1h with gentle shaking every 5min before the infective solution is removed. Cells are then washed with 500µl of cold PBS and 500µl of complemented medium are dispensed in the wells. Incubation is carried on in the incubator for 0, 3, 6 or 12h. Cells are harvested at the appropriate time points and spun down in RNase-free Eppendorf tubes. The complemented medium is then removed and replaced by 300µl of RNase-free TRIZOL (Ambion). The tubes are then vortexed and frozen at -20°C overnight. Tubes are defrozen on ice and vortexed after addition of 60µl of RNase-free chloroform. After a 5min rest at room temperature (RT), tubes are spun 15min at 12.500rpm (4°C). Upper aqueous phase is transferred to another tube containing 300µl of RNase-free isopropanol (99%). Tubes are vortexed, let to rest 10min at RT and spun for 10min at 12.500rpm (4°C). The RNA pellet is washed two times with 500µl of RNase-free ethanol (70%), let dry at RT and resuspended in 10µl of nuclease free water. RNA concentration was determined by nanodrop. This experimental was done in biological duplicates.

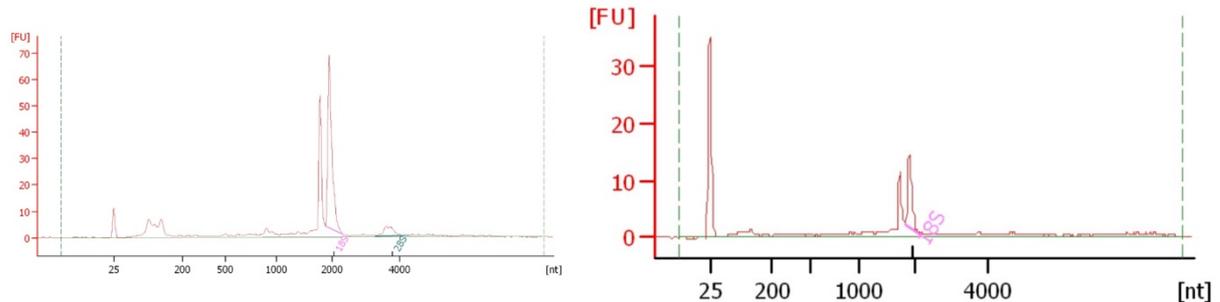
For flies' small RNA HTS

Flies of the different genotypes were obtained by crossing and collected as F1 at 0-3d old. After 3 additional days on fresh food (flies are then 3-6d old), 4,6nl of either TRIS (10mM, pH7,5), DCV (500PFU) or VSV (5000PFU) solutions are injected by intrathoracic injection (Nanoject II apparatus; Drummond Scientific). Flies are then kept for 3 (TRIS and DCV condition) to 5 (VSV condition) days at 25°C. For each RNA sample, 6 flies (3 males and 3 females) are collected and frozen at -80°C overnight in a Precellys tube with ceramic beads. 600µl of RNase-free TRIZOL (Ambion) are added to the tube and flies are crushed using a Precellys shaking apparatus. Then, 120µl of RNase-free chloroform are added to the tubes followed by extensive vortexing. The same steps of RNA precipitation, washing and resuspension as for the cells was then followed.

DCV, CrPV and VSV virus stocks were produced as described in (Kemp et al., 2013).

Checking for quality of extracted RNAs

Samples were run in a Bioanalyzer with an RNA 6000 Nano chip (Agilent) according to manufacturer's instructions. Analysis of the profiles of all samples presented no RNA degradation and the classical dual peaks corresponding to 18S and cleaved 28S ribosomal RNAs.



Representative examples of electropherograms on cells' (left) or flies' (right) RNAs used for the preparation of small RNA libraries.

b. Checking for viral loads and copy number in RNA samples

For S2 cells' small RNA HTS

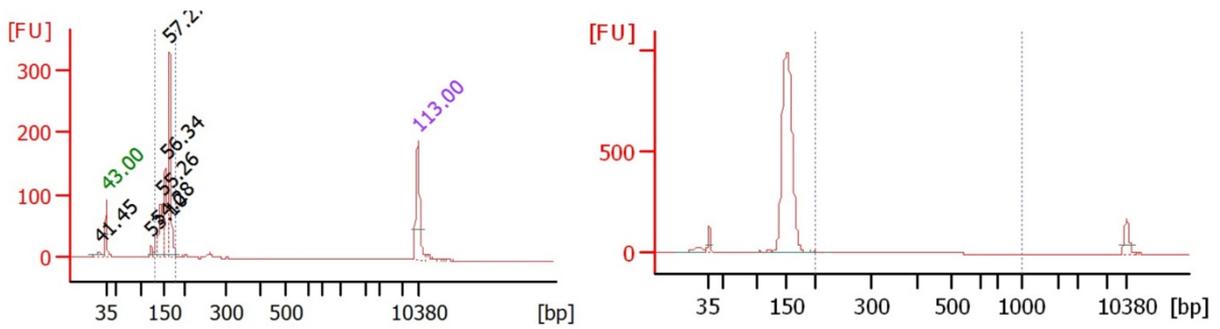
Strand-specific reverse transcription was done on 1 μ g of RNA using iScript Reverse Transcription Supermix (Bio-Rad) with custom Tag containing primers. A (-) reverse transcriptase condition was added to check for aspecific signal amplification. Quantitative PCR was done using the SYBR Green master mix (Bio-Rad). The qPCR cycling conditions were as followed: 98°C – 15s // (95°C – 2s / 60°C – 30s) X 35.

For flies' small RNA HTS

Reverse transcription of 1 μ g of RNA was done using the iScript gDNA Clear cDNA synthesis kit (Bio-Rad) according to manufacturer's instructions. Quantitative PCR was done identically to the one done for S2 cells' RNA samples.

c. Preparation of small RNA libraries

The NEBNext Multiplex Small RNA Library Prep Set for Illumina (NEB) was used to prepare libraries with modifications brought to manufacturer's instructions: (1) 500ng of extracted RNA was used as starting point, (2) all reactions volume were divided by 1.5, (3) custom RNA adaptors were used for individual libraries instead of the normal 5'SR Adaptor, (4) for the small RNA sequencing performed in flies, an additional 2S blocking primer was added to the step of hybridization of RT primers. Size selection after final PCR amplification was done on a 6% non-denaturing acrylamide gel. Elution of the products from the gel was done in 500 μ l of 300mM NaCl solution overnight at 4°C. Filtration on Costar tubes (Merck) was followed by precipitation in 1,5ml of ethanol (100%) with 10 μ g of glycogen (ThermoFisher) per sample for 4h at -20°C. Tubes are spun for 20min – 12.500rpm (4°C) and pellets are washed with 500 μ l ethanol (70%). The pellets are resuspended in nuclease free water and concentrations are measured using a Qubit (ThermoFisher). Finally, 1 μ l are run in a Bioanalyzer with DNA 1000 chip (Agilent) to ensure quality of libraries.



Representative examples of electropherograms on cells' (left) or flies' (right) libraries. All the prepared libraries were sent for Illumina sequencing at the GenomEast sequencing platform from the IGBMC (Strasbourg).

d. Bioinformatic

Trimming and second demultiplexing

Reads were first trimmed off the 2S ribosomal RNA and adaptor sequences by using **trimmomatic-0.36** (<http://www.usadellab.org/cms/?page=trimmomatic>) with the following parameters: **ILLUMINACLIP: 2:20:7 MINLEN:15**. In order to remove the reads coming from sample bleeding, a second demultiplexing step was added. This is made possible by the usage of library-specific 5' adaptors during the reverse transcription step. Thus, each reads from a specific library starts with a distinct 4nt label followed by 2 random nucleotides (6nt internal label). The second demultiplexing was performed by using the 4nt sequence as a **grep** pattern and the internal labels were trimmed using the same **trimmomatic-0.36** tool (**HEADCROP:6**).

Alignments

The alignment of the reads was done using **bowtie 1.2.2** (<http://bowtie-bio.sourceforge.net/manual.shtml>) with the following parameters: **-v 1 --best --no-unal**. The reference genomes used for the alignments are the following ones:

DCV	NC_001834.1
CrPV	NC_003924.1
VSV	EU849003.1
esi1	FBgn0285991
esi2	FBgn0285992
w ^{IR}	Exon 3 of FBgn0003996

Custom files for miRNA hairpins and endo siRNA sources were created by gathering all known sources of these small interfering RNAs. These files were provided by the team of Dr. João Marques.

Normalization of the libraries

We decided to use a miRNA-based normalization method inside each sequencing lane. The mean of 21-23nt long reads mapping drosophila miRNA hairpins was calculated taking into account all the libraries in the sequencing lane. Individual normalization factors were obtained by dividing the number of 21-23nt long miRNAs in each library by this value. Therefore, a “normalized” legend on the small RNA sequencing graphics refers to a number of reads that has been divided by the library specific normalization factor

Data handling

Reads were sorted according to size and frames using an in-house shell script. All plots were first created using custom R scripts and then esthetically modified using Affinity Designer. All R packages used during my Ph.D: BiocGenerics (0.26.0), Biostrings (2.48.0), car (3.0-2), datasets (3.5.1), dplyr (0.7.7), emmeans (1.3.4), fmsb (0.6.3), GenomeInfoDb (1.16.0), GenomicRanges (1.32.7), ggplot2 (3.1.0), gplots (3.0.1.1), graphics (3.5.1), grDevices (3.5.1), gsubfn (0.7), iRanges (2.14.12), lme4 (1.1-21), Matrix (1.2-14), methods (3.5.1), parallel (3.5.1), plotly (4.8.0), plyr (1.8.4), proto (1.0.0), readxl (1.1.0), reshape2 (1.4.3), Rsamtools (1.32.3), S4Vectors (0.18.3), seqinr (3.4-5), stats (3.5.1), stats4 (3.5.1), utils (3.5.1), **viRome** (0.10), XVector (0.20.0). Of note, the Amazon forest does not want me to add the scripts used in this manuscript but I would happily provide them to you upon request!

Chemical probing for determination of DCV domain I

a. Synthesis of template RNA

DCV first 1000 nucleotides were previously cloned in a plasmid (pJL662). Amplification of two different sizes of DCV domain I (first 385 (short) or 824 (long) nucleotides from 5') was done by PCR and a T7 promoter was added. Final PCR components: 10ng pJL662, 0,1mM dNTPs mix, 0,2μM both primers, 1X HF Phusion Buffer, 1U Phusion polymerase (ThermoFisher), 3% DMSO. The PCR cycling conditions were as followed: 95°C – 60s // (95°C – 15s / 57°C – 15s / 72°C – 30s) X 33 // 72°C – 1min. PCR products were checked for good size by migration on 1% agarose gel, purified (NaCl – ethanol precipitation) and resuspended in 20μl nuclease free water.

In vitro transcription reaction: 1X TMSDT buffer (40mM Tris HCl pH 8,1, 22mM MgCl₂, 1mM spermidine, 5mM DTT, 0,01% Triton X-100), 8mM of each NTP, 40U RNasin (Promega), 5μl of in-house T7 RNA polymerase (provided by Dr. Franck Martin team), 20μl purified PCR product. Transcription is first carried on for 1h at 37°C in a water bath. Then, 2μl of pyrophosphatase (1mg/ml - Roche) are added to the reaction and let 30min at 37°C. Finally, 2μl of DNaseI (2U/μl - ThermoFisher) are added to the reaction and let 60min at 37°C. RNA products were checked for good size by migration on a 4% denaturing acrylamide gel, purified by a phenol:chloroform step and resuspended in 100μl nuclease free water. RNA products were kept at -80°C and prepared fresh for each experimental replicate.

b. Chemical modification

Modifications by DMS or CMCT were performed on 2 pmoles of each RNA (long and short). For DMS modification, each RNA is incubated for 15min in DMS buffer (50mM Na Cacodylate, pH7.5, 5mM MgCl₂ and 100mM KCl) and 1µg of yeast total tRNA and then modified with 1.25% DMS reagent (diluted with ethanol 100%) for 10min at 20°C. For CMCT modification, each RNA is incubated for 20min in CMCT buffer (50mM Na borate 50mM; 5mM MgCl₂; 100mM KCl) and 1µg of yeast total tRNA. Then modifications were performed with 10.5g/l CMCT reagent for 20min at 20°C. Both DMS and CMCT modification reactions are stopped on ice. Modified RNAs are precipitated with ethanol 100%, 0.250mM NaCl and 0,2µg glycogen. Pellets were dried and resuspended in 7µl autoclaved milli-Q water.

c. Modified nucleotide detection by primer extension

Reverse transcription was carried out in 20µl reaction volume with 2pmoles of RNA and 0,9pmoles of 5' fluorescently labelled primers. We used 5'Vic and 5'Ned primers of same sequence for all reverse transcription reactions.

First, the modified RNAs are denaturated at 95°C for 2min. Then, fluorescent primers are annealed for 2min at 65°C followed by incubation on ice for 2min. Primer extension is performed in a buffer containing 83mM KCl, 56mM Tris-HCl (pH 8.3), 0,56mM of each of the four deoxynucleotides (dNTP), 5,6mM DTT and 3mM MgCl₂. Reverse transcriptions were performed with 1 unit of Avian Myoblastosis virus (AMV) reverse transcriptase (Promega) at 42°C for 2min, then 50°C for 30min and finally 65°C for 5min. In parallel, sequencing reactions were performed in similar conditions, but containing 0.5mM dideoxythymidine or dideoxycytidine triphosphate (ddTTP or ddCTP). Then, the synthesized cDNAs were phenol:chloroform extracted, precipitated, after centrifugation the pellets were washed, dried and resuspended in 10µl deionized Hi-Di formamide (highly deionized formamide). Samples were loaded on a 96-well plate for sequencing on an Applied Biosystems 3130xl genetic analyzer. The resulting electropherograms were analyzed using QuSHAPE software (<https://weeks.chem.unc.edu/qushape/>), which aligns signal within and across capillaries, as well as to the dideoxy references of nucleotide at specific position and corrects for signal decay.

d. Drawing of DCV domain I secondary structure

Nucleotides were classified in 5 different categories based on their corresponding reactivity value (determined by the mean of the experimental triplicates). The secondary structure model prediction was initiated using mfold (<http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form>) and then edited and refined using Affinity Designer according to our reactivity values.

***In vitro* cleavage assays (Chapter II)**

a. Using recombinant proteins (performed by the team of Pr. Brenda Bass)

Dicer-2 proteins were produced and purified as explained in Sinha and Bass, 2017. Blunt dsRNA substrate was produced as explained in (Welker et al., 2011). *In vitro* transcribed DCV domain I RNA (first 313 nucleotides) was resuspended in 50mM Tris pH8 and 20mM KCl and then refolded (heated at 95°C for 3min and slowly cooled-down to room temperature) or not. Cleavage assay was carried on with 30nM Dicer-2 (+/- Loqs-PD) and 1nM RNA substrate at 25°C and in the following reaction buffer: 25mM Tris pH8, 100mM KCl, 10mM MgCl₂, 1mM TCEP and 5mM ATP. Reaction was stopped by addition of 2 volumes of 2x formamide loading buffer (95% formamide, 18 mM EDTA, 0.025% SDS, xylene cyanol, bromophenol blue) and products were loaded on a 10% denaturing acrylamide gel. A Decade Marker (Ambion) was used for size determination.

b. Using embryonic extracts

Preparation of embryonic extracts

Flies from different genotypes were collected to lay eggs on apple juice-agar plates for 5h. Embryos are then collected and bleached (50% bleach) for 2 min under constant agitation to remove the chorion. After extensive wash with water, embryos are quickly dried, collected in a Precellys tube with 2 ceramic beads and flash-frozen using liquid nitrogen. Embryos are crushed two times using the Precellys shaking apparatus: one time without and one time with 50µl buffer (30mM HEPES pH 7,5, 100mM KOAc, 2mM MgOAc, 10% glycerol and 1mM DTT). Tubes are spun two times for 20min at 12.500rpm (4°C) and the supernatant collected. After determination of protein concentration by Bradford, protein extracts are diluted to a concentration of 1,5mg/ml, aliquoted and frozen at -80°C.

Preparation of RNA substrate

DCV first 1000 nucleotides were previously cloned in a plasmid (pJL662). Amplification of three different sizes of DCV domain I (117, 189 and 264 nucleotides from 5') was done by PCR and a T7 promoter was added. Final PCR components: 20ng pJL662, 0,25mM dNTPs mix, 2µM both primers, 1X HF Phusion Buffer, 1U Phusion polymerase (ThermoFisher), 3% DMSO. The PCR cycling conditions were as followed: 95°C – 1min // (95°C – 15s / 55°C – 15s / 72°C – 1min30s) X 33 // 72°C – 1min. PCR products were checked for good size by migration on agarose gel, purified (phenol:chloroform and NaCl – ethanol precipitation) and resuspended in 20µl nuclease free water. *In vitro* transcription was carried on identically to the “chemical probing” part. For body-labeled radioactive RNA, 1/10th of the UTP nucleotide added to the transcription mix is α -³²P UTP. For capped RNA, a ScriptCap m⁷G Capping System kit (CELLSCRIPT) was used according to manufacturer's instructions and with addition of α -³²P GTP. Radioactive RNAs were gel purified and electroeluted to ensure that no smaller RNA sizes were present in

the mix. Radioactivity of each RNA product was measured (in counts per min or cpm) using a Scintillation System from Beckman Coulter (LS 6500).

Preparation of T1 ladder

T1 enzyme is a ribonuclease catalyzing cleavage of ssRNA 3' of guanosine residues. Radioactive (50.000 cpm) and capped RNA of 264nt long is incubated at 95°C for 1min in a 9µl solution containing 0,1mg/ml of yeast RNA and 1X RNA sequencing buffer (Ambion T1 RNase kit). The solution is cooled-down on ice prior to addition of 1µl of desired T1 enzyme dilution. A final dilution of 1/10th of the stock (1U/ml) was kept after optimization. After 15min of incubation at room temperature, 6µl of STOP mix (0,6M NaAc pH6, 3mM EDTA, 0,1µg/µl tRNA) and 5µl formamide dye 2X (Thermo Fisher) are added. 10µl are loaded on 8% denaturing acrylamide gels alongside products of cleavage experiments for determination of their size.

Cleavage assay

A cleavage reaction is composed of 10µl of radioactive capped or radioactive body-labeled RNA (50.000cpm), 10µl buffer (30mM HEPES pH7,4, 100mM KOAc, 2mM MgOAc, 10% glycerol, 1mM DTT) and 30µl of embryonic extracts (40µg protein). This solution is first let to equilibrate for 15min at room temperature before addition of 10µl of MgOAC-ATP 5mM (t=0). Incubation is carried on for the designated amount of time and stopped by addition of 140µl water and 200µl phenol:chloroform solution. Tubes are vortexed and spun 10min at 12.500rpm before NaCl – ethanol precipitation of the upper phase. Radioactive RNA pellet is resuspended in 20µl of formamide dye 2X and 10µl are loaded on a 8% denaturing acrylamide gel to check for size.

5' Rapid Amplification of cDNA Ends

For the sake of simplicity, only the latest optimized RACE protocol will be detailed here (*Figure 10*).

RNA extraction

Extraction of RNAs was done differently according to their origin: viral stock's RNAs were extracted using the QIAamp Viral RNA Mini Kit (Qiagen), flies' (3 males, 3 females, 3d pi with DCV 500PFU) and cells' (5.10⁶ infected cells at MOI 0,01, 20h pi) RNAs were extracted using a TRIZOL:chloroform method described in the preparation of RNA samples for small RNA HTS.

Reverse transcription

Strand specific reverse transcription was performed using SuperScript III Reverse Transcriptase enzyme (ThermoFisher) according to manufacturer's instructions with a custom RT primer. The resulting cDNAs are precipitated twice using ammonium acetate (1/4 V NaAc 10M + 2,5 V EtOH 100%), washed twice with EtOH 70% and resuspended in 20µl water.

3' tailing

Addition of 3' G nucleotides was performed using a terminal transferase enzyme (NEB) according to manufacturer's instructions and only providing dGTP to the reaction. The resulting products are precipitated and washed with the same methods as previously described (final volume 20µl).

PCR1

Final PCR components: 5µl of poly G-tailed cDNA, 0,2mM dNTPs mix, 0,4µM both primers, 1X HF Phusion Buffer, 0,5U Phusion polymerase (ThermoFisher). The PCR cycling conditions were as followed: (94°C – 5min / 62°C – 5min / 75°C – 5min) // (94°C – 30s / 62°C – 30s / 72°C 30s) X 24 // 72°C – 1min. PCR products were checked for good size by migration on agarose gel, purified (phenol:chloroform and NaCl – ethanol precipitation) and resuspended in 20µl nuclease free water.

BstII digestion

Enzymatic digestion of PCR products was done using 10µl of purified PCR1 product and BstII enzyme (NEB) according to manufacturer's instructions in a total volume of 50µl. Incubation at 37°C was done for the indicated amount of time. Restriction profiles are visualized on an agarose gel and products are phenol:chloroform purified.

PCR2

Final PCR components: 5µl of poly G-tailed cDNA, 0,2mM dNTPs mix, 0,4µM both primers, 1X HF Phusion Buffer, 0,5U Phusion polymerase (ThermoFisher). The PCR cycling conditions were as followed: 94°C – 2min // (94°C – 30s / 62°C – 30s / 72°C 30s) X 11-14 // 72°C – 1min. PCR products were checked for good size by migration on agarose gel, purified (PCR cleanup kit from Qiagen) and resuspended in 15µl nuclease free water.

Cloning, transformation and sequencing of PCR2 products

Addition of a 3'-dA overhang is done by incubating for 20min at 70°C 500ng of purified PCR2 products with Taq polymerase (5U, Invitrogen), MgCl₂ (2,5mM), dATP (0,1mM) and Taq buffer (1X) in a final volume of 10µl. 2µl of this solution are then taken to perform a ligation in a pJET plasmid using the CloneJET PCR Cloning Kit (Thermo) according to manufacturer's instructions. 2µl of ligation products are transformed in chemo-competent DH5α bacteria and plated on ampicillin (100µg/ml) plates. Growing colonies are picked for on-colony PCR and amplified for plasmid extraction by miniprep (illustra plasmidPrep Mini Spin Kit – GE Healthcare) and sequencing of the insert.

RNAi-based screen in S2 cells

a. Knock-down of candidate genes for the RNAi screen

S2 cells were regularly passed as previously described. 20.10⁶ cells are collected, spun-down and resuspended in 13,4ml complemented medium at a concentration of 1,5.10⁶ cells/ml. 30µl of this medium (4,5.10⁴ cells) are dispensed in the internal wells (not at the borders of the plates) of six 96 U-shaped well

plates. Cells are let to attach for 1h in the incubator. Medium is removed from the wells and 40µl of dsRNA solution is added to the cells (20µl serum free medium + 20µl of 0,1µg/µl dsRNA). dsRNAs were provided ready-to-use by the Drosophila RNAi Screening Center at Harvard Medical School. After 5h of incubation in the incubator, 160µl of complemented medium are added to each well. Wells located at the periphery of the plate are filled with 200µl of PBS to prevent evaporation over the course of the experiment. Plates are then sealed with parafilm and put in the incubator for three full days before doing the viral infection.

b. Synchronized viral infection

Cell concentration is measured from extra wells identically seeded. All the plates are put on ice and in the cold room for 30min. Medium is removed from the wells and 50µl of cold infective solution (DCV virus stock diluted to the appropriate MOI in complemented medium) are dispensed. The MOI used for this screen is of 0,01. Cells are kept on ice for 1h with gentle shaking every 5min before the infective solution is removed. Cells are then washed with 150µl of cold PBS and 200µl of complemented medium are dispensed in the wells. Incubation is carried on in the incubator for 20h.

c. RNA extraction and quantitative RT-PCR

Cells lysis, reverse transcription and quantitative PCR were performed using the Cell-To-Ct kit (Ambion). In comparison to manufacturer's instructions, SYBR RT buffer and RT enzyme mix volumes were divided by two. Moreover, the SYBR used was not taken from the kit but from another manufacturer (Biorad). The qPCR cycling conditions were as followed: 95°C – 10s // (95°C – 15s / 60°C – 1min) X 40.

d. Analysis of the data

The threshold cycle (Ct) of each sample is automatically calculated by linear regression. Then, ratio between RP49 and DCV values was calculated as follow:

$$R = \frac{(E_{target})^{\Delta Ct_{target} (mean_{control-sample})}}{(E_{ref})^{\Delta Ct_{ref} (mean_{control-sample})}}$$

With:

E_{target} : efficiency of DCV primers calculated for the specific qPCR plate using dilutions of a standard plasmid
 E_{ref} : efficiency of RP49 primers calculated for the specific qPCR plate using dilutions of a standard plasmid
 ΔCt_{target} : PCR cycle at which fluorescence exceeded determined threshold (Ct) for DCV reaction
 ΔCt_{ref} : PCR cycle at which fluorescence exceeded determined threshold (Ct) for RP49 reaction
 $mean_{control}$: dsLacZ was used as experimental control. Because it is present several times on experimental and qPCR plates, the mean of all the ΔCt values for this specific plate is used for normalization.
 $sample$: ΔCt value corresponding to the tested dsRNA treated sample

Because this experiment was a large-scale screen done in triplicate and involving massive handling of the experimental plates, we statistically searched for bias in the data obtained after qPCR. To do this, a mixed effect model taking into account variations in plates, rows and columns was calculated (R – **lmer** function). The statistical significance of plates, rows and columns effects were tested and proven by ANOVA. Thus,

the corrected ratios (*estimates*) extracted from the mixed effect model were taken into account for further analysis. Statistical significance of a dsRNA effect was calculated by comparing the obtained *estimates* to the average of all data (R – `emmeans` function).

e. dsRNA cell toxicity assay

The exact same protocol of cells seeding and dsRNA soaking as for the RNAi-based screen was followed using the selected dsRNA candidates. 3d post soaking, toxicity of the dsRNAs used is determined by quantifying the mitochondrial activity of treated cells in comparison to untreated ones using an MTS assay kit (Promega CellTiter 96) following manufacturer's instructions. A one-way ANOVA test is used to determine which dsRNA had a significant effect on mitochondrial activity (comparison to dsLacZ treated cells).

Generation of transgenic flies expressing GFP::Dicer-2

a. *Drosophila* genetics, fly maintenance and injections.

dicer-2 mutant flies (*dcr-2^{L811f,x}*) (Lee et al, 2004) were crossed with the deficiency Df(2R) BSC45 (Bloomington stock #7441) or the Df(2R)BSC45–Dcr-2 rescue (previously described in Kemp et al., 2013) lines. All flies contained the GMR-*w^{IR}* transgene (Lee and Carthew, 2003) located on the X-chromosome. Flies were fed on standard cornmeal–agar medium at 25 °C. All flies used were Wolbachia-free.

b. GFP::Dicer-2 complemented flies establishment

A PCR fragment corresponding to the whole *dicer-2* cDNA was cloned into the pENTR/D-TOPO Gateway entry vector using the pENTR directional TOPO cloning kit (Invitrogen) for N-terminal fusion (Majzoub et al., 2014). The *dicer-2* cDNA was then transferred to in-house *Drosophila* transgenic expression vectors by LR recombination using the Gateway LR clonase II enzyme mix (Invitrogen). We established a transgenic vector allowing the expression of GFP::Dicer-2 under the control of the *poly-ubiquitin* promoter with a specific site insertion in the *Drosophila* genome. Basically, the segment containing *poly-ubiquitin* promoter was amplified by PCR from the destination vector pURW (kindly provided by Jean-René Huynh's laboratory ([DGRC#1282](#))) using the primers (OJL2914-NheI-ubi-F- 5'-cgctaGCAAACAGCGCTGACTTTGAG -3' and OJL2915-BamHI-ubi-R-5'-GGGGATCCGCCCTTGATTATTCTGC -3'). The *poly-ubiquitin* promoter was then cloned in a *Drosophila* transgenic vector (called UAS_t-attB PmeI V) provided by Jean-Marc Reichhart containing the *mini-white* gene as a transgenic marker, 5xUAS_t followed by a polylinker, a terminator site SV40 and an attB cassette allowing site specific insertion in flies containing attP site. The UAS_t-attB PmeI V was digested by NheI-BamHI to remove the UAS_t sequences and replace it by the *poly-ubiquitin* promoter (1949bp), this clone was called U-attB (pJL629) allowing the cloning of cDNA under the control of *poly-ubiquitin* promoter. The segment containing the GFP coding sequence and the N-terminal gateway cassette from the destination vector pURW ([DGRC#1282](#)) was amplified by PCR using the primers (OJL2916-

KpnI-eGFP-F- 5'- cggtaCCGGCATGGTGAGCAAGGGCG -3' and OJL2917-BglII-pW-R-5'- ccagatCTGCAGGTTCGACAAAGGTTAAGC -3'). The cassette was the subcloned in a pJET vector (Thermo Fisher). The final *Drosophila* transgenic vector called UGW_attB (*poly-ubiquitin* promoter-GFP-N-terminal gateway cassette – pJL633) was obtained by the insertion of the purified GFP-N-terminal gateway cassette (KpnI-BglIII of 2959bp) in the vector U-attB (pJL629) at the restriction sites KpnI-BamHI to remove the SV40 terminator. Of note, the cloning of gateway cassette containing a *ccdB* gene required the used of specific competent cells (One Shot™ *ccdB* Survival™ competent cells - Thermo Fisher). The vector UGW_attB (pJL633) was fully sequence to validate the cloning.

As we previously validate that a N-terminal GFP fusion of Dicer-2 under the *poly-ubiquitin* promoter inserted in independent sites on the *Drosophila* genome was able to rescue the RNAi pathway (Girardi et al., 2015), we generated several GFP-Dicer-2 variants by PCR mutagenesis and inserted them in the same genomic site to be able to compare easily transgenic lines to each other.

The GFP::Dicer-2 constructs were all inserted on the 3rd chromosome at the position 89E11 (BL#9744). Transgenic lines were generated by BestGene (<https://thebestgene.com/>). As *dicer-2* gene is located on the 2nd chromosome, we could complement *dicer-2* null mutant with the GFP::Dicer-2 constructs. All transgenic constructs were put in the following genetic background: [$w^{IR};dicer-2^{L811f;X}/CyO$; GFP::Dicer-2]. According to the experiments, flies were crossed with a deficiency covering *dicer-2* gene to obtain hemizygote flies for *dicer-2* gene [$w^{IR}; Df(2R)BSC45/CyO$], *dicer-2* null mutation [$w^{IR};dicer-2^{L811f;X}/CyO$] or the genomic rescue of the *dicer-2* gene [$w^{IR}; Df(2R)BSC45, Dcr-2$ -Rescue/ CyO] described in Kemp et al., 2013.

Characterization of GFP::Dicer-2 complemented flies

a. *dicer-2* mRNA quantification by RT-qPCR

Extraction of flies' RNAs and *dicer-2* mRNA determination was done identically to what has been done in the small RNA HTS part.

b. Dicer-2 protein quantification by western blot

60 flies (30 males, 30 females) of each genotype were collected and frozen overnight at -80°C in Precellys tubes with ceramic beads. Flies are then shredded a first time without any additional liquid and a second time with 600µl of lysis buffer (30mM HEPES KOH pH7,5, 50mM NaCl, 2mM Mg(OAc)₂, 1% NP40, 2X cOmplete Protease Inhibitor Cocktail (Roche)). Tubes are spun two times at 12.500rpm for 10min (4°C) to remove fly debris. Resulting protein extract concentration is measured by Bradford technic at 595nm. 40µg of proteins are run on a 4-12% acrylamide gel (Bio-Rad) after 5min heating at 95°C. Semi-dry transfer to nitrocellulose membrane was performed with Biorad TransBlot Turbo machine. Membranes were blocked with 5% non-fat dry milk in TBS-Tween 0,05% one hour at room temperature and incubated overnight at 4°C with primary antibody in non-fat dry milk 2% TBS-Tween 0,05%. After washing, the secondary antibody fused to horse-radish peroxidase (HRP) was added to the membrane in non-fat dry

milk 2% TBS-Tween 0,05% for one hour at room temperature. Membranes were then washed and revealed with the enhanced chemiluminescence reagent (GE Healthcare) in a ChemiDoc (Bio-Rad) apparatus.

Antibodies used and dilutions:

Target	Reference	Origin	Dilution
Dicer-2	ab4732 - Abcam	Rabbit	1/1.000
Anti-rabbit	3918-8816-31 - Millipore	Goat	1/10.000

c. Eye color quantification

For each sample, 10 heads of female flies aged from 3 to 5 days were frozen in liquid nitrogen and homogenized twice at 25Hz during 1min30s. 1mL of AEA (solution with 30% EtOH, 1% HCl) was then added and the coloration was analyzed with a spectrophotometer at 485nm. At least 6 independent samples were used for each genotype.

d. esi-1 and esi-2 mRNA precursors quantification by RT-qPCR

Expression of *esi-1* and *esi-2* precursors was checked in flies' testes. To do so, dissection of the testes was performed in PBS and the tissues were frozen in 300µl of TRIZOL (Ambion) at -80°C overnight in a Precellys tube with ceramic beads. Tubes were thawed on ice and homogenized in a Precellys apparatus. 60µl of chloroform are added to the tube before vortexing. The same steps of RNA precipitation, washing and resuspension as for the cells small RNA HTS extraction method was then followed. Reverse transcription of 1µg of RNA was done using the iScript gDNA Clear cDNA synthesis kit (Bio-Rad) according to manufacturer's instructions. Quantitative PCR was done identically to the one done for S2 cells' small RNA HTS samples.

Pulldown assay

dmDcr-2 WT or mutants (2µM) was incubated with His-Loqs PD WT (4µM) in pulldown assay buffer (25mM TRIS pH8, 175mM KCl, 10mM MgCl₂, 10mM imidazole, 1mM TCEP, 5% glycerol, 0.1% NP-40) at 4°C for 1h. His-select resin was added to the reactions and rotated at 4°C for 2h. Centrifugation and removal of the supernatant removed most unbound proteins. Resin and bound proteins were washed two times with pulldown assay to remove additional unbound proteins. Bound proteins were eluted with elution buffer (pulldown assay buffer with 300mM imidazole). Proteins were resolved by SDS PAGE and stained with Coomassie. Bands were quantified using ImageJ.

Primers (5'-3')

Strand specific RT primers with Tag (**bold**).

DCV	(-) strand	TTGGCATTAAAGGCTATCGG TCATCGGTATGCACATTGCT
	(+) strand	TTGGCATTAAAGGCTATCGG CGCATAACCATGCTCTTCTG
CrPV	(-) strand	TTGGCATTAAAGGCTATCGG GCTGAAACGTTCAACGCATA
	(+) strand	TTGGCATTAAAGGCTATCGG CCACTTGCTCCATTTGGTTT

qPCR primers.

Tag	Strand spec. qPCR	
		TTGGCATTAAAGGCTATCGG
DCV	Fw	TCATCGGTATGCACATTGCT
	Rv	CGCATAACCATGCTCTTCTG
RP49	Fw	GCCGCTTCAAGGACAGTATCT
	Rv	AAACGCGGTTCTGCATGAG
CrPV	Fw	GCTGAAACGTTCAACGCATA
	Rv	CCACTTGCTCCATTTGGTTT
Dicer-2	Fw	AAACTGATGCCAGGTGGAAG
	Rv	ATTCCAAAACGCTCAACAC
esi-1 precursor	Fw	GGTGCTGCGCATACCTTT
	Rv	CAAGGCTAGGGCTCGTCA
esi-2 precursor	Fw	CAAACACCCACACACATACACA
	Rv	CCAGGGCGCTACATTCAATA

Primers used for PCR mutagenesis. Nucleotides different from wild-type sequence are in red.

G31R mutation	Fw	TTGTCTACCTGCCCA CA AGATCTGGGAAAACGTT C
	Rv	GAACGTTTTCCAGAT C TTGTGGGCAGGTAGACAA
F225G mutation	Fw	CACAGAGGTCATGGTGTCC GG TCCACATCAAGAGCAAGTG
	Rv	CACTTGCTCTTGATGTGGAC CC GGACACCATGACCTCTGTG

Primers used for synthesis of cleavage assays RNA substrates.

T7-DCV5'	Fw	ATATTAATACGACTCACTATAGGG tttatatcgtgtgtacatataaatatgtacacacggc
PCR (117)	Rv	ccatcaacaatgaaaaccgtaaagcataataacc
PCR (189)	Rv	ccttaccatcaacaatgaaaacaacgtatcag
PCR (264)	Rv	ggccatcaacattgaaaacaataaggcataattc

5' RNA adaptors for small RNA sequencing library building. Manufacturer's adaptor is annotated as "original" and internal labels of custom primers are in **bold**. N stands for any nucleotide and r letters between each nucleotide represents the fact that primers are RNA primers.

RA5mod01	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCr ArCrGrUrNrN
RA5mod02	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCr CrGrUrArNrN
RA5mod03	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCr GrUrCrArNrN
RA5mod04	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCr CrCrArGrNrN
RA5mod05	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCr ArGrCrCrNrN
RA5mod06	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCr UrUrArGrNrN
RA5mod07	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCr GrArUrCrNrN
RA5mod08	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCr UrArGrUrNrN
RA5mod09	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCr ArGrGrCrNrN
RA5mod10	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCr CrArArUrNrN
RA5mod11	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCr GrCrUrArNrN
RA5mod12	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCr UrUrCrGrNrN
RA5 original	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC

Primers used for determination of DCV domain I chemical probing. Sequence in **bold** corresponds to the T7 promoter sequence.

T7-DCV5'	Fw	ATATTAATACGACTCACTATAGG TTAATATCGTGTGTACAT
PCR short	Rv	CCAGATATCAAATCAGTAAAGAG
PCR long	Rv	CAGGCCATACTTTTATCAGATTC
RT primer	VIC/NED 5' labeled	GGACTAACTCAGTATACCCTACT

RACE related primers (**bold** sequences are adapters, D stands for all nucleotides except C, N stands for all nucleotides).

RT primer	Rv	CATTACAAATCCAAAAGACGCATC
PCR1	Rv	CCTTGGCACCCGAGAATTCCACT TAAGTTGAGATGTAATCTTTG
	Fw	GTTTACAGAGTTCTACAGTCCGACGATC ACGTCCCCCCCCDN
PCR2	Rv	AATGATACGGCGACCACCGAGATCTACACGTTTACAGTTCTACAGTCCGA
	Fw	CAAGCAGAAGACGGCATAACGAGATTGTTGGGTGACTGGAGTTCCCTTGGCACCCGAGAATTCCA
On-colony PCR and seq.	Fw	GTAAAACGACGGCCAG
	Rv	CAGGAAACAGCTATGAC

Bibliography

Andino, R., Rieckhof, G.E., and Baltimore, D. (1990). A functional ribonucleoprotein complex forms around the 5' end of poliovirus RNA. *Cell* 63, 369–380.

Barton, D.J., O'Donnell, B.J., and Flanagan, J.B. (2001). 5' cloverleaf in poliovirus RNA is a cis-acting replication element required for negative-strand synthesis. *EMBO J.* 20, 1439–1448.

Bevilacqua, P.C., Ritchey, L.E., Su, Z., and Assmann, S.M. (2016). Genome-Wide Analysis of RNA Secondary Structure. *Annu. Rev. Genet.* 50, 235–266.

Bleichert, F., and Baserga, S.J. (2007). The Long Unwinding Road of RNA Helicases. *Mol. Cell* 27, 339–352.

Cenik, E.S., Fukunaga, R., Lu, G., Dutcher, R., Wang, Y., Tanaka Hall, T.M., and Zamore, P.D. (2011). Phosphate and R2D2 Restrict the Substrate Specificity of Dicer-2, an ATP-Driven Ribonuclease. *Mol. Cell* 42, 172–184.

Cherry, S., and Perrimon, N. (2004). Entry is a rate-limiting step for viral infection in a *Drosophila melanogaster* model of pathogenesis. *Nat. Immunol.* 5, 81–87.

Cherry, S., Kunte, A., Wang, H., Coyne, C., Rawson, R.B., and Perrimon, N. (2006). COPI activity coupled with fatty acid biosynthesis is required for viral replication. *PLoS Pathog.* 2, e102.

Civril, F., Bennett, M., Moldt, M., Deimling, T., Witte, G., Schiesser, S., Carell, T., and Hopfner, K.-P. (2011). The RIG-I ATPase domain structure reveals insights into ATP-dependent antiviral signalling. *EMBO Rep.* 12, 1127–1134.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., et al. (2008a). An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 453, 798–802.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., et al. (2008b). An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 453, 798–802.

Deddouche, S., Matt, N., Budd, A., Mueller, S., Kemp, C., Galiana-Arnoux, D., Dostert, C., Antoniewski, C., Hoffmann, J.A., and Imler, J.-L. (2008). The DExD/H-box helicase Dicer-2 mediates the induction of antiviral activity in *drosophila*. *Nat. Immunol.* 9, 1425–1432.

Ewart, G.D., and Howells, A.J. (1998). [15] ABC transporters involved in transport of eye pigment precursors in *Drosophila melanogaster*. In *Methods in Enzymology*, (Elsevier), pp. 213–224.

Fairman-Williams, M.E., Guenther, U.-P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* 20, 313–324.

Fareh, M., Yeom, K.-H., Haagsma, A.C., Chauhan, S., Heo, I., and Joo, C. (2016). TRBP ensures efficient Dicer processing of precursor microRNA in RNA-crowded environments. *Nat. Commun.* 7.

Feng, Q., Hato, S.V., Langereis, M.A., Zoll, J., Virgen-Slane, R., Peisley, A., Hur, S., Semler, B.L., van Rij, R.P., and van Kuppeveld, F.J.M. (2012). MDA5 Detects the Double-Stranded RNA Replicative Form in Picornavirus-Infected Cells. *Cell Rep.* 2, 1187–1196.

Förstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D. (2007). *Drosophila* microRNAs Are Sorted into Functionally Distinct Argonaute Complexes after Production by Dicer-1. *Cell* 130, 287–297.

Fukunaga, R., Colpan, C., Han, B.W., and Zamore, P.D. (2014). Inorganic phosphate blocks binding of pre-miRNA to Dicer-2 via its PAZ domain. *EMBO J.* 33, 371–384.

Galiana-Arnoux, D., Dostert, C., Schneemann, A., Hoffmann, J.A., and Imler, J.-L. (2006). Essential function in vivo for Dicer-2 in host defense against RNA viruses in *drosophila*. *Nat. Immunol.* 7, 590–597.

Gamarnik, A.V., and Andino, R. (1997). Two functional complexes formed by KH domain containing proteins with the 5' noncoding region of poliovirus RNA. *RNA N. Y.* N 3, 882–892.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L.W., Zapp, M.L., Weng, Z., et al. (2008). Endogenous siRNAs Derived from Transposons and mRNAs in *Drosophila* Somatic Cells. *Science* 320, 1077–1081.

Girardi, E., Lefèvre, M., Chane-Woon-Ming, B., Paro, S., Claydon, B., Imler, J.-L., Meignin, C., and Pfeffer, S. (2015). Cross-species comparative analysis of Dicer proteins during Sindbis virus infection. *Sci. Rep.* 5, 10693.

Goubau, D., Schlee, M., Deddouche, S., Pruijssers, A.J., Zillinger, T., Goldeck, M., Schuberth, C., Van der Veen, A.G., Fujimura, T., Rehwinkel, J., et al. (2014). Antiviral immunity via RIG-I-mediated recognition of RNA bearing 5'-diphosphates. *Nature* 514, 372–375.

Griffiths, J.A., Richard, A.C., Bach, K., Lun, A.T.L., and Marioni, J.C. (2018). Detection and removal of barcode swapping in single-cell RNA-seq data. *Nat. Commun.* 9.

Gross, L., Vicens, Q., Einhorn, E., Noireterre, A., Schaeffer, L., Kuhn, L., Imler, J.-L., Eriani, G., Meignin, C., and Martin, F. (2017). The IRES5'UTR of the dicistrovirus cricket paralysis virus is a type III IRES containing an essential pseudoknot structure. *Nucleic Acids Res.* 45, 8993–9004.

Han, Y.-H., Luo, Y.-J., Wu, Q., Jovel, J., Wang, X.-H., Aliyari, R., Han, C., Li, W.-X., and Ding, S.-W. (2011). RNA-Based Immunity Terminates Viral Infection in Adult *Drosophila* in the Absence of Viral Suppression of RNA Interference: Characterization of Viral Small Interfering RNA Populations in Wild-Type and Mutant Flies. *J. Virol.* 85, 13153–13163.

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. *Mol. Cell* 50, 762–777.

Hartig, J.V., and Förstemann, K. (2011). Loqs-PD and R2D2 define independent pathways for RISC generation in *Drosophila*. *Nucleic Acids Res.* 39, 3836–3851.

Herold, J., and Andino, R. (2001). Poliovirus RNA Replication Requires Genome Circularization through a Protein–Protein Bridge. *Mol. Cell* 7, 581–591.

Jan, E., and Sarnow, P. (2002). Factorless Ribosome Assembly on the Internal Ribosome Entry Site of Cricket Paralysis Virus. *J. Mol. Biol.* 324, 889–902.

Jaworski, E., and Routh, A. (2017). Parallel ClickSeq and Nanopore sequencing elucidates the rapid evolution of defective-interfering RNAs in Flock House virus. *PLOS Pathog.* 13, e1006365.

Jiang, F., Ramanathan, A., Miller, M.T., Tang, G.-Q., Gale, M., Patel, S.S., and Marcotrigiano, J. (2011). Structural basis of RNA recognition and activation by innate immune receptor RIG-I. *Nature* 479, 423–427.

Kandasamy, S.K., and Fukunaga, R. (2016). Phosphate-binding pocket in Dicer-2 PAZ domain for high-fidelity siRNA production. *Proc. Natl. Acad. Sci.* 201612393.

Kemp, C., Mueller, S., Goto, A., Barbier, V., Paro, S., Bonnay, F., Dostert, C., Troxler, L., Hetru, C., Meignin, C., et al. (2013). Broad RNA interference-mediated antiviral immunity and virus-specific inducible responses in *Drosophila*. *J. Immunol. Baltim. Md* 1950 190, 650–658.

Khong, A., Bonderoff, J., Spriggs, R., Tammpere, E., Kerr, C., Jackson, T., Willis, A., and Jan, E. (2016). Temporal Regulation of Distinct Internal Ribosome Entry Sites of the Dicistroviridae Cricket Paralysis Virus. *Viruses* 8, 25.

King, L.A., and Moore, N.F. (1988). Evidence for the presence of a genome-linked protein in two insect picornaviruses, cricket paralysis and *Drosophila C* viruses. *FEMS Microbiol. Lett.* 50, 41–44.

Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 40, e3–e3.

Kloc, A., Rai, D.K., and Rieder, E. (2018). The Roles of Picornavirus Untranslated Regions in Infection and Innate Immunity. *Front. Microbiol.* 9.

Kolakofsky, D., Kowalinski, E., and Cusack, S. (2012). A structure-based model of RIG-I activation. *RNA* 18, 2118–2127.

Kunzelmann, S., Böttcher, R., Schmidts, I., and Förstemann, K. (2016). A Comprehensive Toolbox for Genome Editing in Cultured *Drosophila melanogaster* Cells. *G3 Bethesda Md* 6, 1777–1785.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct Roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA Silencing Pathways. *Cell* 117, 69–81.

Li, H., Li, W.X., and Ding, S.W. (2002). Induction and suppression of RNA silencing by an animal virus. *Science* 296, 1319–1321.

Linder, P., and Jankowsky, E. (2011). From unwinding to clamping — the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* 12, 505–516.

Lu, C., Xu, H., Ranjith-Kumar, C.T., Brooks, M.T., Hou, T.Y., Hu, F., Herr, A.B., Strong, R.K., Kao, C.C., and Li, P. (2010). The Structural Basis of 5' Triphosphate Double-Stranded RNA Recognition by RIG-I C-Terminal Domain. *Structure* 18, 1032–1043.

Lu, R., Maduro, M., Li, F., Li, H.W., Broitman-Maduro, G., Li, W.X., and Ding, S.W. (2005). Animal virus replication and RNAi-mediated antiviral silencing in *Caenorhabditis elegans*. *Nature* 436, 1040–1043.

Luo, D., Ding, S.C., Vela, A., Kohlway, A., Lindenbach, B.D., and Pyle, A.M. (2011). Structural Insights into RNA Recognition by RIG-I. *Cell* 147, 409–422.

Luo, D., Kohlway, A., and Pyle, A.M. (2013). Duplex RNA activated ATPases (DRAs): Platforms for RNA sensing, signaling and processing. *RNA Biol.* 10, 111–120.

Lyons, T., Murray, K.E., Roberts, A.W., and Barton, D.J. (2001). Poliovirus 5'-Terminal Cloverleaf RNA Is Required in cis for VPg Uridylylation and the Initiation of Negative-Strand RNA Synthesis. *J. Virol.* 75, 10696–10708.

Ma, E., MacRae, I.J., Kirsch, J.F., and Doudna, J.A. (2008). Autoinhibition of Human Dicer by Its Internal Helicase Domain. *J. Mol. Biol.* 380, 237–243.

Majzoub, K., Hafirassou, M.L., Meignin, C., Goto, A., Marzi, S., Fedorova, A., Verdier, Y., Vinh, J., Hoffmann, J.A., Martin, F., et al. (2014). RACK1 controls IRES-mediated translation of viruses. *Cell* 159, 1086–1095.

Marques, J.T., Kim, K., Wu, P.-H., Alleyne, T.M., Jafari, N., and Carthew, R.W. (2010a). Loqs and R2D2 act sequentially in the siRNA pathway in *Drosophila*. *Nat. Struct. Mol. Biol.* 17, 24–30.

Marques, J.T., Kim, K., Wu, P.-H., Alleyne, T.M., Jafari, N., and Carthew, R.W. (2010b). Loqs and R2D2 act sequentially in the siRNA pathway in *Drosophila*. *Nat. Struct. Mol. Biol.* 17, 24–30.

Marques, J.T., Wang, J.-P., Wang, X., de Oliveira, K.P.V., Gao, C., Aguiar, E.R.G.R., Jafari, N., and Carthew, R.W. (2013). Functional specialization of the small interfering RNA pathway in response to virus infection. *PLoS Pathog.* 9, e1003579.

Martins, N., Lemoine, A., Santiago, E., Paro, S., Imler, J.-L., and Meignin, C. (2019). A Transgenic Flock House Virus Replicon Reveals an RNAi Independent Antiviral Mechanism Acting in *Drosophila* Follicular Somatic Cells. *G3 Bethesda Md* 9, 403–412.

Miyoshi, K., Miyoshi, T., Hartig, J.V., Siomi, H., and Siomi, M.C. (2010). Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in *Drosophila*. *RNA N. Y. N* 16, 506–515.

Mohr, S.E. (2014). RNAi screening in *Drosophila* cells and in vivo. *Methods* 68, 82–88.

Moore, N.F., Kearns, A., and Pullin, J.S. (1980). Characterization of cricket paralysis virus-induced polypeptides in *Drosophila* cells. *J. Virol.* 33, 1–9.

Moore, N.F., Reavy, B., Pullin, J.S.K., and Plus, N. (1981). The polypeptides induced in *Drosophila* cells by *Drosophila C* virus (strain Ouarzazate). *Virology* 112, 411–416.

Mueller, S., Gausson, V., Vodovar, N., Deddouche, S., Troxler, L., Perot, J., Pfeffer, S., Hoffmann, J.A., Saleh, M.-C., and Imler, J.-L. (2010). RNAi-mediated immunity provides strong protection against the negative-strand RNA vesicular stomatitis virus in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 107, 19390–19395.

Nakashima, N., and Ishibashi, J. (2010). Identification of the 3C-protease-mediated 2A/2B and 2B/2C cleavage sites in the nonstructural polyprotein precursor of a dicistrovirus lacking the NPGP motif. *Arch. Virol.* 155, 1477–1482.

Nakashima, N., and Nakamura, Y. (2008). Cleavage sites of the “P3 region” in the nonstructural polyprotein precursor of a dicistrovirus. *Arch. Virol.* 153, 1955–1960.

Nakashima, N., and Shibuya, N. (2006). Multiple coding sequences for the genome-linked virus protein (VPg) in dicistroviruses. *J. Invertebr. Pathol.* 92, 100–104.

Nayak, A., Berry, B., Tassetto, M., Kunitomi, M., Acevedo, A., Deng, C., Krutchinsky, A., Gross, J., Antoniewski, C., and Andino, R. (2010). Cricket paralysis virus antagonizes Argonaute 2 to modulate antiviral defense in *Drosophila*. *Nat. Struct. Mol. Biol.* 17, 547–554.

Nayak, A., Kim, D.Y., Trnka, M.J., Kerr, C.H., Lidsky, P.V., Stanley, D.J., Rivera, B.M., Li, K.H., Burlingame, A.L., Jan, E., et al. (2018). A Viral Protein Restricts

Drosophila RNAi Immunity by Regulating Argonaute Activity and Stability. *Cell Host Microbe* 24, 542-557.e9.

Nishida, K.M., Miyoshi, K., Ogino, A., Miyoshi, T., Siomi, H., and Siomi, M.C. (2013). Roles of R2D2, a cytoplasmic D2 body component, in the endogenous siRNA pathway in *Drosophila*. *Mol. Cell* 49, 680–691.

Ookamura, K., Chung, W.-J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008a). The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* 453, 803–806.

Okamura, K., Chung, W.-J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008b). The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* 453, 803–806.

Paro, S., Imler, J.-L., and Meignin, C. (2015). Sensing viral RNAs by Dicer/RIG-I like ATPases across species. *Curr. Opin. Immunol.* 32, 106–113.

Pathak, K., and Nagy, P. (2009). Defective Interfering RNAs: Foes of Viruses and Friends of Virologists. *Viruses* 1, 895–919.

Paul, A.V., and Wimmer, E. (2015). Initiation of protein-primed picornavirus RNA synthesis. *Virus Res.* 206, 12–26.

Paul, A.V., Yin, J., Mugavero, J., Rieder, E., Liu, Y., and Wimmer, E. (2003). A “Slide-back” Mechanism for the Initiation of Protein-primed RNA Synthesis by the RNA Polymerase of Poliovirus. *J. Biol. Chem.* 278, 43951–43960.

Peisley, A., Lin, C., Wu, B., Orme-Johnson, M., Liu, M., Walz, T., and Hur, S. (2011). Cooperative assembly and dynamic disassembly of MDA5 filaments for viral dsRNA recognition. *Proc. Natl. Acad. Sci.* 108, 21010–21015.

Peisley, A., Jo, M.H., Lin, C., Wu, B., Orme-Johnson, M., Walz, T., Hohng, S., and Hur, S. (2012). Kinetic mechanism for viral dsRNA length discrimination by MDA5 filaments. *Proc. Natl. Acad. Sci.* 109, E3340–E3349.

Perrimon, N., and Mathey-Prevot, B. (2007). Applications of High-Throughput RNA Interference Screens to Problems in Cell and Developmental Biology. *Genetics* 175, 7–16.

Persson, H., Søkilde, R., Pirona, A.C., and Rovira, C. (2017). Preparation of highly multiplexed small RNA sequencing libraries. *BioTechniques* 63.

Pestova, T.V. (2003). Translation elongation after assembly of ribosomes on the Cricket paralysis virus internal ribosomal entry site without initiation factors or initiator tRNA. *Genes Dev.* 17, 181–186.

Plaskon, N.E., Adelman, Z.N., and Myles, K.M. (2009). Accurate Strand-Specific Quantification of Viral RNA. *PLoS ONE* 4, e7468.

Putnam, A.A., and Jankowsky, E. (2013). DEAD-box helicases as integrators of RNA, nucleotide and protein binding. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* 1829, 884–893.

Rieder, E., Paul, A.V., Kim, D.W., van Boom, J.H., and Wimmer, E. (2000). Genetic and biochemical studies of poliovirus cis-acting replication element cre in relation to VPg uridylylation. *J. Virol.* 74, 10371–10380.

van Rij, R.P., Saleh, M.-C., Berry, B., Foo, C., Houk, A., Antoniewski, C., and Andino, R. (2006). The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in *Drosophila melanogaster*. *Genes Dev.* 20, 2985–2995.

Routh, A., Ordoukhanian, P., and Johnson, J.E. (2012). Nucleotide-resolution profiling of RNA recombination in the encapsidated genome of a eukaryotic RNA virus by next-generation sequencing. *J. Mol. Biol.* 424, 257–269.

Rybak-Wolf, A., Jens, M., Murakawa, Y., Herzog, M., Landthaler, M., and Rajewsky, N. (2014). A Variety of Dicer Substrates in Human and *C. elegans*. *Cell* 159, 1153–1167.

Sabin, L.R., Zheng, Q., Thekkat, P., Yang, J., Hannon, G.J., Gregory, B.D., Tudor, M., and Cherry, S. (2013). Dicer-2 processes diverse viral RNA species. *PloS One* 8, e55458.

Sanjuán, R. (2012). From Molecular Genetics to Phylodynamics: Evolutionary Relevance of Mutation Rates Across Viruses. *PLoS Pathog.* 8, e1002685.

Sanjuán, R., Nebot, M.R., Chirico, N., Mansky, L.M., and Belshaw, R. (2010). Viral Mutation Rates. *J. Virol.* 84, 9733–9748.

Sato, S., Li, K., Kameyama, T., Hayashi, T., Ishida, Y., Murakami, S., Watanabe, T., Iijima, S., Sakurai, Y., Watashi, K., et al. (2015). The RNA Sensor RIG-I Dually Functions as an Innate Sensor and Direct Antiviral Factor for Hepatitis B Virus. *Immunity* 42, 123–132.

Schlee, M., Roth, A., Hornung, V., Hagmann, C.A., Wimmenauer, V., Barchet, W., Coch, C., Janke, M., Mihailovic, A., Wardle, G., et al. (2009). Recognition of 5' Triphosphate by RIG-I Helicase Requires Short Blunt Double-Stranded RNA as Contained in Panhandle of Negative-Strand Virus. *Immunity* 31, 25–34.

Senti, K.-A., and Brennecke, J. (2010). The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet.* 26, 499–509.

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and Mechanism of Helicases and Nucleic Acid Translocases. *Annu. Rev. Biochem.* 76, 23–50.

Sinha, N.K., Trettin, K.D., Aruscavage, P.J., and Bass, B.L. (2015). *Drosophila* Dicer-2 Cleavage Is Mediated by Helicase- and dsRNA Termini-Dependent States that Are Modulated by Loquacious-PD. *Mol. Cell* 58, 406–417.

Sinha, N.K., Iwasa, J., Shen, P.S., and Bass, B.L. (2018). Dicer uses distinct modules for recognizing dsRNA termini. *Science* 359, 329–334.

Sinha, R., Stanley, G., Gulati, G.S., Ezran, C., Travaglini, K.J., Wei, E., Chan, C.K.F., Nabhan, A.N., Su, T., Morganti, R.M., et al. (2017). Index switching causes “spreading-of-signal” among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. *BioRxiv*.

Steil, B.P., and Barton, D.J. (2009). Cis-active RNA elements (CREs) and picornavirus RNA replication. *Virus Res.* 139, 240–252.

Steimer, L., and Klostermeier, D. (2012). RNA helicases in infection and disease. *RNA Biol.* 9, 751–771.

Suhasini, A.N., and Brosh, R.M. (2013). DNA Helicases Associated with Genetic Instability, Cancer, and Aging. In *DNA Helicases and DNA Motor Proteins*, M. Spies, ed. (New York, NY: Springer New York), pp. 123–144.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P.D. (2004). A protein sensor for siRNA asymmetry. *Science* 306, 1377–1380.

Trettin, K.D., Sinha, N.K., Eckert, D.M., Apple, S.E., and Bass, B.L. (2017). Loquacious-PD facilitates *Drosophila* Dicer-2 cleavage through interactions with the helicase domain and dsRNA. *Proc. Natl. Acad. Sci. U. S. A.* 114, E7939–E7948.

Tuiskunen, A., Leparc-Goffart, I., Boubis, L., Monteil, V., Klingstrom, J., Tolou, H.J., Lundkvist, A., and Plumet, S. (2010). Self-priming of reverse transcriptase impairs strand-specific detection of dengue virus RNA. *J. Gen. Virol.* 91, 1019–1027.

Turner, D.H., and Mathews, D.H. (2010). NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.* 38, D280–D282.

van der **V**alk, T., Vezzi, F., Ormestad, M., Dalén, L., and Guschanski, K. (2018). Estimating the rate of index hopping on the Illumina HiSeq X platform. *BioRxiv*.

Veen, A.G., Maillard, P.V., Schmidt, J.M., Lee, S.A., Deddouche-Grass, S., Borg, A., Kjær, S., Snijders, A.P., and Reis e Sousa, C. (2018). The RIG-I-like receptor LGP2 inhibits Dicer-dependent processing of long double-stranded RNA and blocks RNA interference in mammalian cells. *EMBO J.* 37.

Vogt, D.A., and Andino, R. (2010). An RNA Element at the 5'-End of the Poliovirus Genome Functions as a General Promoter for RNA Synthesis. *PLoS Pathog.* 6, e1000936.

Wang, Y., Ludwig, J., Schuberth, C., Goldeck, M., Schlee, M., Li, H., Juranek, S., Sheng, G., Micura, R., Tuschl, T., et al. (2010). Structural and functional insights into 5'-ppp RNA pattern recognition by the innate immune receptor RIG-I. *Nat. Struct. Mol. Biol.* 17, 781–787.

Warsaba, R., Sadasivan, J., and Jan, E. (2019). Dicistrovirus-Host Molecular Interactions. *Curr. Issues Mol. Biol.* 83–112.

Watanabe, M., Iwakawa, H., Tadakuma, H., and Tomari, Y. (2017). Biochemical and single-molecule analyses of the RNA silencing suppressing activity of CrPV-1A. *Nucleic Acids Res.* 45, 10837–10844.

Weber, M., Sediri, H., Felgenhauer, U., Binzen, I., Bänfer, S., Jacob, R., Brunotte, L., García-Sastre, A., Schmid-Burgk, J.L., Schmidt, T., et al. (2015). Influenza Virus Adaptation PB2-627K Modulates Nucleocapsid Inhibition by the Pathogen Sensor RIG-I. *Cell Host Microbe* 17, 309–319.

Welker, N.C., Maity, T.S., Ye, X., Aruscavage, P.J., Krauchuk, A.A., Liu, Q., and Bass, B.L. (2011). Dicer's Helicase Domain Discriminates dsRNA Termini to Promote an Altered Reaction Mode. *Mol. Cell* 41, 589–599.

Wickersheim, M.L., and Blumenstiel, J.P. (2013). Terminator oligo blocking efficiently eliminates rRNA from *Drosophila* small RNA sequencing libraries. *BioTechniques* 55.

Wu, B., Peisley, A., Richards, C., Yao, H., Zeng, X., Lin, C., Chu, F., Walz, T., and Hur, S. (2013). Structural Basis for dsRNA Recognition, Filament Formation, and Antiviral Signal Activation by MDA5. *Cell* 152, 276–289.

Yang, S., Yu, J., Fan, Z., Gong, S., Tang, H., and Pan, L. (2018). Bub1 Facilitates Virus Entry through Endocytosis in a Model of *Drosophila* Pathogenesis. *J. Virol.* 92.

Yao, H., Dittmann, M., Peisley, A., Hoffmann, H.-H., Gilmore, R.H., Schmidt, T., Schmid-Burgk, J.L., Hornung, V., Rice, C.M., and Hur, S. (2015). ATP-Dependent Effector-like Functions of RIG-I-like Receptors. *Mol. Cell* 58, 541–548.

Yoneyama, M., Onomoto, K., Jogi, M., Akaboshi, T., and Fujita, T. (2015). Viral RNA detection by RIG-I-like receptors. *Curr. Opin. Immunol.* 32, 48–53.

Zhao, C., Liu, F., and Pyle, A.M. (2018). An ultraprocessive, accurate reverse transcriptase encoded by a metazoan group II intron. *RNA* 24, 183–195.

Zhou, R., Czech, B., Brennecke, J., Sachidanandam, R., Wohlschlegel, J.A., Perrimon, N., and Hannon, G.J. (2009). Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform. *RNA* 15, 1886–1895.

Annexe : 10% de la thèse traduit en français

Préambule

Pendant la période de mon doctorat, j'ai eu la chance d'écrire un chapitre d'un livre intitulé " Insect Molecular Virology: Advances and Emerging Trends". Ce livre a été édité par l'éminent chercheur et professeur Bryony C. Bonning et publié en juin 2019. Le résultat de ce travail d'écriture est une revue globale des mécanismes de défense antivirale des insectes avec un accent particulier sur les récepteurs de l'immunité innée qui détectent les infections virales. Comme l'indique le titre de ce manuscrit, mon travail de doctorat portait principalement sur la détection des ARN viraux par Dicer-2 dans *Drosophila melanogaster*.

Ainsi, en raison de la pertinence de mon chapitre de livre et de sa bibliographie liée dans le cadre de mon travail de doctorat, j'ai décidé de l'utiliser comme une introduction globale de mon manuscrit. Ensuite, une petite transition mettra en évidence les principales questions qui ont guidé la rédaction des trois chapitres suivants. Enfin, une introduction plus approfondie de certains aspects spécifiques du domaine sera faite dans les chapitres correspondants.

Outro (terme couramment utilisé en musique pour désigner le contraire d'une introduction)

Tandis que d'importants travaux sont en cours pour tenter de découvrir de nouveaux senseurs de virus dans la drosophile, mes travaux de doctorat ont porté sur la caractérisation du seul capteur connu d'acides nucléiques viraux qui alimente la voie de l'ARNi antiviral : Dicer-2. Des études *in vitro* ont déjà proposé des modèles de mécanismes d'action de Dicer-2 sur des substrats d'ARNdb en fonction de leur nature, de leur longueur et de leur extrémité, mais aussi de leur interaction avec plusieurs cofacteurs. Tous ces paramètres qui peuvent être étroitement contrôlés dans des conditions *in vitro* sont autant d'inconnues lorsqu'il s'agit d'études *in vivo*. En effet, bien que des cibles endogènes, exogènes et virales de Dicer-2 aient été identifiées, on sait peu de choses sur les caractéristiques exactes de ces ARNdb. Ainsi, deux questions principales ont guidé la rédaction de ce manuscrit :

Comment Dicer-2 accède à ses substrats et quelles sont leurs caractéristiques ?

En raison de son double rôle dans les voies endo-siRNA et ARNi antivirales, Dicer-2 doit être capable de détecter et de discriminer une grande diversité potentielle de molécules d'ARNdb. La détection des ARN viraux est rendue encore plus difficile par leur fort potentiel d'adaptabilité lié au taux de mutation de leur enzyme répliquatrice. Ainsi, de nombreux contre-mécanismes ont évolué et compliquent leur détection. Néanmoins, Dicer-2 est capable d'accéder à de nombreux virus et de traiter les molécules d'ARNdb formées pendant leur cycle d'infection. Comment et où Dicer-2 parvient-il à contourner les défenses virales (Chapitre I) ? Quelles sont les caractéristiques de ces points faibles potentiels (chapitre II) ? J'ai essayé de répondre à ces questions en utilisant les virus dicistrovirus *Drosophila C Virus* et *Cricket Paralysis Virus* comme modèles d'infection.

Quel est le rôle de l'hélicase Dicer-2 et de ses cofacteurs associés ?

Il est frappant de constater que les protéines impliquées dans l'immunité antivirale de divers organismes possèdent toutes la même organisation spécifique du domaine hélicase. Dans le cas des récepteurs de type RIG chez les mammifères, par exemple, ce domaine s'est avéré nécessaire pour la reconnaissance de leur cible et la régulation de leur activité. Ainsi, l'implication du domaine hélicase de Dicer-2 dans la détection de ses cibles virales et endogènes a été étudiée. En outre, deux mutations distinctes de ce domaine ont été étudiées dans le but de dissocier son rôle de détection possible de son activité ATPase (chapitre III).

Introduction – Chapitre I

Les dicistrovirus sont des virus non enveloppés dont le génome est relativement petit (~8-10kb) avec un ARN messager monopartite positif et une petite capsidie icosaédrique (~30nm). Ils ont un large tropisme dans le phylum des Arthropodes, sont extrêmement divers et représentent des menaces agricoles et économiques majeures dans le monde entier. En raison de leurs caractéristiques et de la similitude en termes de symptômes de maladies (paralysie, par exemple), ils ont été initialement classés dans la famille des Picornaviridae. Cependant, leur organisation génomique ne se compare pas ce qui justifie la classification de ces virus dans des familles distinctes. En effet, les picornavirus ont un génome monocistronique codant d'abord pour les protéines structurales, puis pour les protéines non structurales, tandis que les dicistrovirus ont un génome bicistronique codant pour les protéines non structurales dans l'ORF1 et les protéines structurales dans l'ORF2. Malgré l'étude approfondie des interactions des dicistrovirus avec leur hôte, de nombreuses lacunes subsistent dans notre compréhension des mécanismes des différentes étapes de l'infection.

La plupart des détails mécanistes du cycle d'infection par les dicistrovirus proviennent d'études menées sur la drosophile et deux virus du genre Cripavirus : Drosophila C Virus (DCV) et le Cricket Paralysis Virus (CrPV). Une description détaillée de l'entrée, de la réplication, de la traduction et du conditionnement des dicistrovirus a été revue dans Warsaba et al, 2019. En bref, les virus s'attachent très probablement d'abord à un récepteur de surface cellulaire non encore identifié et entrent dans la cellule par la voie de l'endocytose clathrinodépendante (Cherry et Perrimon, 2004 ; Yang et al., 2018). Ensuite, le génome viral est libéré dans le cytoplasme et ciblé pour la traduction et la réplication. La traduction des deux ORF, indépendante de la coiffe, est médiée par le ribosome hôte et produit deux polyprotéines (Moore et al., 1980, 1981). La digestion ultérieure de ces polyprotéines par la protéase de type 3C codée viralement donne naissance à des protéines matures non structurales (ORF1) et structurales (ORF2) (Nakashima et Ishibashi, 2010 ; Nakashima et Nakamura, 2008). Parallèlement, la réplication est assurée par l'ARN polymérase ARN dépendante qui interagit avec des protéines hôtes liées à la queue poly A et à la VPg liée au génome viral en 5'. Le mécanisme par lequel les brins sens et antisens sont produits est encore obscure et a surtout été déduit des études sur les picornavirus (expliqué en détail au chapitre II et étudié dans Paul et Wimmer, 2015). Enfin, la famille des dicistrovirus comprend des

virus lytiques et non lytiques, ce qui complique notre compréhension de leurs stratégies de transmission (Figure 1).

Tel que présenté dans l'introduction générale de ce manuscrit, les insectes et plus particulièrement les drosophiles sont capables de monter des réponses antivirales contre une grande diversité de virus. Plusieurs mécanismes immunitaires innés conservés au cours de l'évolution peuvent être identifiés comme antiviraux, à savoir les voies JAK-STAT, Toll, IMD et Heat-Shock. En outre, des facteurs de restriction exprimés de façon constitutive peuvent également participer à la lutte contre les virus chez les insectes. Bien que tous ces mécanismes de défense soient la plupart du temps spécifiques du virus, la voie des ARN interférents demeure la principale réponse antivirale générale chez les insectes. Comme expliqué précédemment, Dicer-2 est le seul senseur connu d'acides nucléiques viraux et activateur de la voie siRNA antivirale chez la drosophile. Cette enzyme est décrite comme une endonucléase d'ARNdb entrant par l'extrémité de son substrat et générant des siARNs dérivés du virus. Une expression altérée de cette enzyme catalytique entraîne une charge virale plus élevée ainsi qu'un taux de survie réduit des mouches injectées avec le dicistrovirus DCV (Figure 2 et Galiana-Arnoux et al., 2006 ; van Rij et al., 2006). Ainsi, Dicer-2 doit être capable de détecter et de cliver l'intermédiaire de réplication ARNdb produit par DCV (Figure 1). Cependant, les dicistrovirus sont hautement protégés et cachés du système immunitaire et nous ne comprenons toujours pas comment Dicer-2 est capable de les détecter.

Premièrement, les ARN génomiques des dicistrovirus sont protégés contre le mécanisme de dégradation de l'ARN cellulaire 5'-3' médié par Pacman par une protéine virale liée au génome (VPg) à leur extrémité 5' (King et Moore, 1988 ; Nakashima et Shibuya, 2006). L'absence d'une coiffe d'ARNm à l'extrémité 5' de leur génome les protège également des enzymes de decapping Dcp1 et Dcp2. En plus de son rôle protecteur, la VPg est proposée comme amorce protéique utilisée par la RdRp virale pour initier la synthèse des brins génomiques et antigénomiques (Paul et Wimmer, 2015). De même, on pense que la queue 3' poly A des génomes des dicistrovirus est liée par des protéines cellulaires pour favoriser la circularisation de l'ARN viral et la synthèse subséquente des brins antigénomiques par la RdRp (Herold et Andino, 2001). Cette queue poly A est identique à celle présente à l'extrémité 3' des ARNm cellulaires et protège très probablement le génome viral contre la dégradation 3'-5' à médiation

exosomique. En plus de ces caractéristiques de stabilisation génomique, la réplication des dicistrovirus se produit souvent sur les structures des membranes hôtes. Dans le cas du DCV, le remodelage du Golgi entraîne la formation de petites usines de réplication virale (~115 nm de diamètre) qui sont difficiles d'accès pour le système immunitaire (Cherry et al., 2006). Enfin, les virus ont développé leurs propres mécanismes de défense contre l'ARNi, à savoir les suppresseurs viraux de l'ARNi (VSRs). De nombreuses stratégies ont évolué pour contrer l'ARNi antiviral à de multiples niveaux et même des dicistrovirus étroitement liés comme le DCV et le CrPV codent pour des VSRs qui ont des modes d'action différents. Alors que le DCV 1A lie l'ARNdb viral intermédiaire de réplication pour bloquer le clivage par Dicer-2, le CrPV 1A interagit directement avec la protéine effectrice Ago2 pour supprimer son activation de clivage et la cibler pour la dégradation protéosomale (Nayak et al., 2010, 2018 ; van Rij et al, 2006 ; Watanabe et al., 2017). Par conséquent, les dicistrovirus semblent être extrêmement bien protégés contre les voies immunitaires antivirales (Figure 1).

Malgré toutes ces couches de protection, des preuves solides montrent que Dicer-2 joue un rôle important dans la défense contre les dicistrovirus. Ainsi, une question majeure a guidé mon doctorat pendant quatre ans : comment Dicer-2 est-il capable de détecter les ARN protégés des dicistrovirus ?

L'une des caractéristiques de l'ARNi antiviral est que les siARNs dérivés du virus et produits par Dicer-2 fournissent une empreinte de l'action du système immunitaire. Ainsi, le fait de pouvoir identifier une signature siRNA pourrait nous fournir des informations mécanistes sur la détection et l'entrée de Dicer-2 sur les virus. La méthode actuelle de choix pour étudier les petits ARNs est sans aucun doute le séquençage des petits ARN à haut débit (HTS). Au cours de mon doctorat, j'ai utilisé deux approches complémentaires basées sur cette technologie afin de chercher le point d'entrée de Dicer-2 sur deux dicistrovirus, DCV et CrPV. Tout d'abord, une étude cinétique de l'apparition des siRNAs a été réalisée en utilisant le modèle simplifié des cellules S2. Ces cellules de type macrophages peuvent être infectées de manière synchronisée, ce qui permet l'identification des siRNAs aux premiers temps d'infection. Par la suite, on a effectué le séquençage à haut débit de petits ARNs de mouches infectées par DCV exprimant différents variants de Dicer-2 pour vérifier la pertinence des données

obtenues dans les cellules et pour mieux comprendre les mécanismes moléculaires à l'œuvre.

Introduction – Chapitre II

Dans le chapitre précédent, une enquête approfondie sur l'apparition de siRNAs dans des conditions infectées (DCV et CrPV) a été menée. Le séquençage des petits ARNs à haut débit réalisé simultanément dans les mouches et dans les cellules S2 pointe vers un point d'entrée précis de Dicer-2 sur l'ARNdb viral, sans que nous puissions l'identifier précisément. Néanmoins, des preuves solides suggèrent que ce point d'entrée est situé dans la région 5' non traduite (UTR) des deux virus étudiés. Chez les dicistrovirus, cette région englobe une grande diversité de structures d'ARN de la plus haute importance pour leur réplication, traduction, virulence et interaction avec les protéines hôtes.

Pour rappel, les dicistrovirus sont des virus bicistroniques avec la traduction de l'ORF1 sous le contrôle du 5'IRES (situé dans le 5'UTR) et la traduction de l'ORF2 sous le contrôle de l'IRES interne (IGR - Chapitre I - Figure 1). Parce qu'il ne nécessite pas de facteur d'initiation de traduction, l'IGR du dicistrovirus fait partie des structures de recrutement du ribosome 40S les plus simples et est bien conservé entre ces virus (Jan et Sarnow, 2002 ; Pestova, 2003). Au contraire, les besoins en facteurs d'organisation structurelle et d'initiation de l'IRES 5' sont beaucoup plus variables d'un virus à l'autre, ce qui pourrait refléter des stratégies d'adaptation distinctes utilisées pour détourner le ribosome de l'hôte. Par exemple, le DCV et le CrPV nécessitent une protéine ribosomale appelée Rack1 pour contrôler leur traduction médiée par le 5'IRES (Majzoub et al., 2014).

Une étude récente visant à mieux comprendre la traduction médiée par le 5'IRES a mis en évidence l'interaction entre cette structure d'ARN et les sous-unités du facteur 3 d'initiation eucaryote (Gross et al., 2017). Le 5' UTR des dicistrovirus englobe à la fois cette structure 5'IRES et un domaine structuré supplémentaire situé en amont : le domaine I. Selon nos données de séquençage d'ARN, le point d'entrée de Dicer-2 serait situé dans cette région du domaine I. Dans le cadre de l'étude précitée, les structures 2D de ces deux domaines ont été déterminées pour le CrPV. Le modèle du domaine I a révélé une structure comprenant 5 tiges boucles simples séparées par de courtes régions de liaison d'ARN simple brin. Cependant, une caractérisation plus poussée de cette

région à l'aide d'une méthode de sondage chimique DMS/CMCT a révélé une structure initialement décrite dans le poliovirus comme un trèfle (données non publiées de l'équipe du Dr Franck Martin, IBMC-UPR9002, Strasbourg - Figure 23). Chez les picornavirus, il a été démontré que cette région interagit avec la RdRP viral et avec les protéines de l'hôte pour favoriser la synthèse des brins négatifs et positifs (Andino et al., 1990 ; Barton et al., 2001 ; Gamarnik et Andino, 1997 ; Herold et Andino, 2001 ; Vogt et Andino, 2010). De plus, il a été démontré que l'interaction entre la structure du trèfle, la vRdRP et la VPg est responsable du déclenchement de la synthèse de l'ARN du picornavirus à amorçage protéique (Lyons et al., 2001 ; Rieder et al., 2000). Le modèle actuellement proposé pour la synthèse des brins négatifs et positifs dans les picornavirus est présenté à la figure 19.

Dans les picornavirus, ce trèfle est situé dans la région très 5' du génome viral, qui contient le point d'entrée hypothétique de Dicer-2. Il est important de noter que (1) nous ne comprenons pas comment Dicer-2 pourrait avoir accès à une extrémité d'ARNdb non protégée différente de l'extrémité 5' virale et (2) nous ne connaissons pas la structure 2D du domaine I de DCV, qui semble également comprendre le point d'entrée de Dicer-2. Pour commencer, un modèle de la structure 2D de l'ARN du domaine I de DCV a été obtenu en effectuant un sondage chimique *in vitro*. Ensuite, dans le but de trouver le point d'entrée précis de Dicer-2, la caractérisation de la sensibilité de clivage de ce ssRNA a été déterminée *in vitro* en utilisant soit la protéine recombinante Dicer-2 soit des extraits embryonnaires de mouches. Enfin, une tentative de détermination des extrémités de 5' produites dans des conditions infectées a été faite en utilisant une méthode d'amplification rapide de l'extrémité 5' de l'ADNc (RACE). Dans l'ensemble, ces techniques ont été utilisées pour mieux comprendre cette région non caractérisée du DCV.

Introduction – Chapitre III

Les hélicases sont des protéines omniprésentes qui existent dans toutes les formes de vie cellulaire (Bleichert et Baserga, 2007 ; Linder et Jankowsky, 2011). Leur implication dans pratiquement toutes les facettes du métabolisme de l'ADN et de l'ARN est illustrée par les nombreuses maladies causées par leur dérégulation (Steimer et Klostermeier, 2012 ; Suhasini et Brosh, 2013). Les hélicases peuvent être classées en 6 superfamilles (SF) en fonction de leurs structures, fonctions et motifs de séquence partagée (Singleton et al., 2007). Les enzymes SF1 et SF2 contiennent une structure hélicase conservée composée de deux domaines de type RecA. Ces domaines englobent des motifs de séquences spécifiques nécessaires à l'hydrolyse de l'ATP et à la liaison aux acides nucléiques (Putnam et Jankowsky, 2013). L'alignement des séquences centrales des hélicases SF1 et SF2 de *S. cerevisiae*, *E. coli* et de certains virus a permis de regrouper ces deux superfamilles en 12 familles (9 en SF2 et 3 en SF1, Fairman-Williams et al., 2010).

Des acteurs importants de l'immunité antivirale des mammifères peuvent être trouvés dans la famille des récepteurs de type Rig-I (RLR) de la superfamille SF2, à savoir RIG-I, MDA5 et LGP2. En bref, la détection de l'ARN viral par RIG-I ou MDA-5 conduit à leur oligomérisation sur leur substrat d'ARNdb, à leur signalisation via leur domaine CARD N-terminal vers MAVS, à l'activation de la voie IFN et à l'expression ultérieure des ISG pour monter une réponse antivirale (examinée dans Yoneyama et al., 2015). Dans le cas du LGP2 sans CARD, en plus d'un rôle suggéré dans la modulation de l'activité des RIG-I et MDA5, il pourrait être impliqué dans la régulation négative de l'ARNi chez les mammifères par une interaction directe avec Dicer (Veen et al., 2018). Les RLR ont une organisation particulière de domaine hélicase avec un domaine Hel2i inséré entre les deux domaines de type RecA. Cette organisation particulière du domaine est conservée entre les RLR et les enzymes Dicer et joue un rôle important dans la détection des ARN (Civril et al., 2011 ; Jiang et al., 2011 ; Kolakofsky et al., 2012 ; Luo et al., 2011). Cependant, comme l'activité de déroulement de ces hélicases inférées n'a jamais été démontrée, ils ont été rebaptisés ATPases activées par l'ARN duplex ou ATPases dépendantes de l'ARN double brin (DRA, examinée dans Luo et al., 2013 ; Paro et al., 2015).

Les principales différences entre Dicer et RLR résident dans l'absence de domaines CARD N-terminaux et la présence de domaines RNaseIII supplémentaires dans Dicer (Figure 33). Cette différence dans la composition des domaines entraîne un mode d'action catalytique des enzymes Dicer sur leur substrat, ce qui explique pourquoi les enzymes Dicer sont classées comme DRA catalytiques (c) tandis que RIG-I et MDA5 sont des DRA signalétiques (s) (voir Paro et al., 2015). Toutefois, il convient de noter que la distinction stricte des deux catégories peut être discutée. En effet, les fonctions d'effecteurs antiviraux directs des RLR par déplacement des protéines virales ont déjà été suggérées (Sato et al., 2015 ; Weber et al., 2015 ; Yao et al., 2015). D'autre part, l'activité catalytique de Dicer-2 sur les ARNdb viraux n'est pas suffisante pour obtenir une réponse ARNi antivirale et nécessite l'action amplificatrice d'Ago2 (van Rij et al., 2006). De plus, il a été démontré que l'activation des gènes antiviraux par Dicer-2, comme vago, limite la réplication virale (Deddouche et al., 2008).

Il a été démontré que les enzymes DRA ont différentes spécificités de substrat et présentent des activités dépendantes et indépendantes de l'ATP. Ainsi, il a été démontré que RIG-I reconnaissait des extrémités 5' di- ou tri-phosphate d'ARNdb alors que MDA5 adopte un mode de liaison à la tige de longs ARNdb sans contact avec l'extrémité de l'ARNdb (Goubau et al, 2014 ; Lu et al, 2010 ; Schlee et al, 2009 ; Wang et al, 2010 ; Wu et al., 2013). Ensuite, l'oligomérisation de MDA5 sur son substrat est indépendante de l'ATP, mais l'hydrolyse de l'ATP favorise son désassemblage des ARNdb courts (Peisley et al., 2011, 2012). D'autre part, les monomères RIG-I peuvent lier les extrémités d'ARNdb d'une manière indépendante de l'ATP mais nécessitent une hydrolyse d'ATP pour leur oligomérisation (Goubau et al., 2014 ; Luo et al., 2011). Il est intéressant de noter que ces RLR peuvent présenter des préférences pour les substrats viraux. Dans le cas des picornavirus qui ne présentent pas de PPP génomique en 5', c'est MDA5 et non Rig-I qui est nécessaire pour monter une réponse antivirale efficace (Feng et al., 2012). Au contraire, chez la drosophile, Dicer-2 s'est avérée nécessaire pour se défendre contre tous les virus testés, quelle que soit la nature de leur génome.

Comme pour les RLR, une étude approfondie de Dicer-2 in vitro a révélé que différents modes d'action dépendants ou non de l'ATP coexistent (Cenik et al., 2011 ; Sinha et al., 2015 ; Welker et al., 2011). Ainsi, l'ARN double brin à extrémité franche déclenche une

activité processive efficace de Dicer-2 dépendante du domaine DRA et de l'ATP, de sorte qu'une seule protéine Dicer-2 clive plusieurs fois avant de se dissocier. Au contraire, une molécule d'ARNdb avec des terminaisons en porte-à-faux de 3', favorise une activité lente, indépendante de l'ATP et distributive de Dicer-2, caractérisée par la dissociation de Dicer-2 après chaque clivage. Une étude récente basée sur la cryo-microscopie électronique a proposé deux mécanismes de détection distincts de Dicer-2 en fonction des extrémités de l'ARNdb (Figure 34, Sinha et al., 2018) :

- Les ARNdb à extrémités franches sont d'abord liés par le domaine hélicase de Dicer-2. Ensuite, la molécule d'ARNdb est enfilée et déroulée grâce à l'hydrolyse de l'ATP jusqu'à ce qu'elle atteigne le domaine PAZ et qu'elle soit coupée par les domaines RNaseIII. Il faut noter que ce mécanisme peut conduire à la production d'ARN plus courts ou plus longs que les siRNAs canoniques (de 5 à ~30nt de long). Nous ne savons pas encore si cette hétérogénéité des produits Dicer-2 est pertinente *in vivo* ou si des cofacteurs de liaison empêchent le découpage aléatoire des ARNdb tirés dans l'hélicase.
- D'autre part, l'ARNdb présentant des extrémités en porte-à-faux de 3' est directement lié via son 5' P par le domaine PAZ à travers la poche de liaison au phosphate (Kandasamy et Fukunaga, 2016). La molécule d'ARNdb est finalement rapprochée des domaines RNaseIII et clivée. Ce mécanisme est indépendant de l'ATP et ne génère que des siRNAs canoniques.

Cette étude a été rendue possible par le découplage fonctionnel des deux modes d'action de Dicer-2. Ainsi, les mutations dans le domaine PAZ ont inhibé le clivage des ARNdb présentant des extrémités en porte-à-faux de 3' tout en maintenant la processivité de l'ARNdb à extrémités franches. Au contraire, une mutation dans le domaine de l'hélicase de Dicer-2 a modifié son action sur l'ARNdb à extrémité franche tout en laissant inchangée son activité distributive sur les ARNdb à extrémités en porte-à-faux de 3'. Il est intéressant de noter que cette dernière mutation a été identifiée en recherchant des similitudes entre la séquence de Dicer-2 et le domaine C-terminal de RIG-I, qui a été suggéré comme étant responsable de la reconnaissance de l'ARNdb à extrémités franches (Luo et al., 2011). Une de ces régions de la bite de Dicer-2 a été identifiée dans son domaine hélicase et une seule mutation d'une phénylalanine (F) 225 en une glycine (G) a été réalisée (Dicer-2F225G). En raison du rôle proposé du domaine C-terminal de RIG-I C dans la détection, il a été proposé que cette mutation dans Dicer-2 pourrait

avoir un impact sur le traitement des cibles endogènes ou virales de bonne foi de Dicer-2 in vivo.

Dans l'ensemble, cette étude propose un modèle d'action de Dicer-2 sur différents ARNdb synthétiques. Il est important de noter que dans les études in vitro, les séquences et les extrémités des substrats d'ARNdb utilisés sont définies arbitrairement. En effet, les caractéristiques exactes des cibles in vivo d'ARNdb de Dicer-2 de bonne foi demeurent l'une des inconnues les plus importantes dans le domaine de l'ARNi. C'est pourquoi, à la suite de cette publication marquante, il a été décidé d'étudier l'impact de la mutation Dicer-2F225G caractérisée in vitro chez la mouche. Des mouches exprimant cette version mutée de Dicer-2 ont été générées parallèlement aux mouches décrites au chapitre I. La caractérisation des voies siRNA dépendantes de Dicer-2, à savoir les voies endo-siRNA et siRNA antivirales, a été réalisée en utilisant des méthodes RT-qPCR ainsi que la puissante méthode de séquençage à haut débit des petits ARNs. Enfin, la comparaison des données obtenues à partir de différents génotypes de mouches nous a permis de mieux comprendre l'implication de l'hélicase Dicer-2 dans la détection et le traitement de ses cibles véritables.

Observations finales

Toutes les conclusions concernant les résultats obtenus lors de mon doctorat ont déjà été présentées dans les chapitres correspondants. Néanmoins, les points suivants peuvent être considérés comme le message à retenir de ce manuscrit :

- Les séquençages à haut débit des petits ARNs de cellules et de mouches infectées par DCV et CrPV ont mis en évidence une entrée précoce, précise et indépendante de l'ATP de Dicer-2 sur l'ARNdb viral correspondant au domaine I des dicistrovirus.
- L'étude in vitro du domaine I de DCV a permis la modélisation de sa structure 2D et la caractérisation de sa sensibilité au clivage endonucléolytique par des protéines présentes dans des extraits embryonnaires.
- Un criblage basé sur l'ARNi dans des cellules S2 a permis d'identifier 20 gènes candidats présentant un impact sur la charge virale du DCV. Ces gènes sont en cours de validation.
- L'étude in vivo de deux différentes mutations ponctuelles du domaine hélicase de Dicer-2 m'a permis d'affiner le modèle de mécanisme de Dicer-2 proposé précédemment.

Les données de séquençage à haut débit que j'ai obtenues doivent encore être approfondies et couplées à d'autres méthodes mentionnées dans les discussions afin de vérifier la validité des hypothèses que nous proposons.

Au moment où j'écris ces lignes, d'importantes questions demeurent sans réponse : (1) Quelles sont toutes les cibles de Dicer-2 in vivo et sont-elles toutes clivées en siRNAs ? (2) Quelles sont les caractéristiques exactes de ces cibles et comment sont-elles différenciées entre les précurseurs endogènes de siRNAs, les intermédiaires viraux de la réplication et les autres ARN cellulaires ? (3) Quel est le rôle de Loqs-PD dans ce mécanisme de tri ? (4) Où se trouve le point d'entrée de Dicer-2 sur ses substrats in vivo ?

Enfin, le travail présenté dans ce manuscrit était hautement multidisciplinaire et exigeait l'expertise individuelle de nombreuses équipes et personnes différentes. Ainsi, je suis plus que jamais convaincu que les réponses à toutes ces questions ne peuvent naître que

d'une forte collaboration entre personnes travaillant *in vitro* et *in vivo*, biologistes moléculaires, virologues, bioinformaticiens....

Loïc TALIDE
**Détection des ARNs
viraux par Dicer-2 chez la
drosophile**

Résumé

Je me suis intéressé au système de défense anti-viral majeur de *Drosophila melanogaster* qui est la voie du RNA silencing (siRNA). A ce jour, le seul senseur d'acide nucléique viral et activateur de la voie siRNA est Dicer-2. Ainsi, le travail que j'ai effectué a permis d'apporter de nouvelles informations concernant la détection des ARNs viraux par Dicer-2. L'utilisation de méthodes de séquençage à haut débit (HTS) des petits ARNs dans des cellules S2 infectées par le Drosophila C Virus (DCV) à des temps précoces m'a permis de proposer un point d'entrée précis et interne de Dicer-2 sur l'ARN double brin de ce dicistrovirus. La validation de ce point faible dans la défense du virus a été effectuée en réalisant un HTS des petits ARNs dans des mouches de différents génotypes infectées avec DCV. J'ai ensuite caractérisé plus en profondeur cette région du génome virale en déterminant tout d'abord sa structure 2D puis sa sensibilité à des clivages médiés par des extraits embryonnaires de mouches. Finalement, l'utilisation de différents variants de Dicer-2 présentant des mutations du domaine DRA m'a permis de proposer un nouveau mécanisme de fonctionnement de cette protéine.

Mots clés : Dicer-2, *D. melanogaster*, Dicistrovirus, séquençage à haut débit

Résumé en anglais

My Ph.D revolved around the study of the major antiviral defense system of *Drosophila melanogaster*: the siRNA pathway. To date, the only viral nucleic acid sensor and siRNA pathway activator in drosophila is Dicer-2. Thus, the work I have done has provided new information regarding the detection of viral RNAs by Dicer-2. The use of high throughput sequencing (HTS) methods of small RNAs in S2 cells infected with Drosophila C Virus (DCV) at early time points has allowed me to propose a precise and internal entry point for Dicer-2 on the double-stranded RNA of this dicistrovirus. The validation of this weak point in the defence of the virus was carried out by performing an HTS of small RNAs in flies of different genotypes infected with DCV. I then characterized this region of the viral genome in more depth by first determining its 2D structure and then its sensitivity to cleavages mediated by embryonic fly extracts. Finally, the use of different variants of Dicer-2 with mutations in the DRA domain allowed me to propose a new mechanism of action for this protein.

Keywords: Dicer-2, *D. melanogaster*, Dicistrovirus, high-throughput sequencing