Identification and characterization of the 21U RNA processing enzyme in *Caenorhabditis elegans*

Dissertation

Zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Nadezda Podvalnaya

Geb. am 26.09.1991 in Moskau, Russland

Mainz, 2023

Tag der mündlichen Prüfung: 15.04.2024

Table of contents

Summary.

The piRNA pathway is a widely-represented, small RNA-based mechanism that takes part in transposon silencing. It is mostly active in the germline and early embryos. Even though the piRNA pathway is present in many species, mechanistic details of the pathway vary significantly. For instance, in many animals piRNAs are produced from long precursor transcripts that are processed afterward to form multiple mature piRNAs, whereas in *C. elegans* piRNA precursors are transcribed from individual loci that each produce a single piRNA. PiRNAs in *C. elegans* are called 21U RNA, because they are 21 nucleotide long and have strong 5' U bias. 21U RNA precursors are roughly 28 nt long and contain a 5' cap. To become a 21U RNAs, these precursors undergo maturation both at the 5' and 3' end.

The mechanism of the 5' end processing of 21U RNA precursors in *C. elegans* is yet unknown. Our lab previously identified a protein complex, named PETISCO, that binds 21U RNA precursors and is required for their processing. Interestingly, PETISCO has two mutually exclusive interactors: PID-1 and TOST-1. PID-1-PETISCO only plays a role in 21U RNA maturation, while TOST-1-PETISCO has a different and still unknown maternal essential function, but does not affect 21U RNA biogenesis. However, how PETISCO can help 21U RNA production is still unclear.

I started with the purification of PETISCO proteins and obtained working antibodies against TOFU-6, IFE-3, and PID-3. In collaboration with Sebastien Falk (Vienna), the crystal structure of PETISCO was obtained and the effects of the disrupting interactions within PETISCO were tested *in vivo*.

In my second project, I identified a novel nuclease complex, which had been named PUCH (Precursor of 21U 5' end Cleavage Holoenzyme), with an essential role in piRNA precursor processing. PUCH is characterized by three subunits, TOFU-1, TOFU-2, and one of the newly identified 21U RNA pathway factors SLFL-3 or SLFL-4. The active center of PUCH is formed by Schlafen-like domains, present in all four proteins. I proved the activity of PUCH *in vivo* and *in vitro.* Analysis of the substrate requirements showed that PUCH needs m^7G-cap and U in position three to perform the cleavage.

Again, in collaboration with Sebastian Falk, the link between PETISCO and PUCH was identified. Interaction between TOFU-6 and TOFU-1 was demonstrated *in vitro* and crystal

structure of this interaction was obtained. CRISPR/CAS9 genome editing we disrupted the interaction between PUCH and PETISCO and confirmed the relevance of this interaction for 21U RNA maturation *in vivo*.

Thus, I identify PUCH as a novel 5' end processing complex that drives piRNA precursor maturation in *C. elegans*, in conjunction with the precursor-binding complex PETISCO. The catalytic center of PUCH consists of three different SLFN domains. Given the multitude of non-characterized SLFN homologs in mammalian genomes, PUCH may define a novel type of multi-subunit nuclease enzymes with many different yet unknown functions.

Zusammenfassung.

Der piRNA-Signalweg ist ein weit verbreiteter, auf kleinen RNAs basierender Mechanismus, der am Transposon-Silencing beteiligt ist. Er ist hauptsächlich in der Keimbahn und in frühen Embryonen aktiv. Obwohl der piRNA-Signalweg in vielen Spezies konserviert vorkommt, unterscheiden sich die mechanistischen Details erheblich. in vielen Tieren werden piRNAs aus langen Vorläufertranskripten gebildet, die nachfolgend zu mehreren reifen piRNAs prozessiert werden, während in C. elegans piRNA-Vorläufer von einzelnen Loci transkribiert werden, die nur jeweils eine einzelne piRNA produzieren. In C. elegans werden piRNAs auch als 21U-RNA bezeichnet, da sie 21 Nukleotide lang sind und einen starken Uracil-Bias am 5'-Ende aufweisen. Deren Vorläufer sind etwa 28 nt lang und enthalten eine 5'-Kappe. Um zu einer 21U-RNA zu werden, durchlaufen diese Vorläufer sowohl am 5'- als auch am 3'-Ende eine Reifeprozessierung.

Der Mechanismus zur Prozessierung des 5'-Endes von 21U-RNA-Vorläufern in C. elegans ist noch unbekannt. In vorausgegangenen Arbeiten konnte in unserem Labor bereits einen Proteinkomplex namens PETISCO identifiziert werden, der 21U-RNA-Vorläufer bindet und für deren Reifung erforderlich ist. Interessanterweise hat PETISCO zwei sich gegenseitig ausschließende Interaktoren: PID-1 und TOST-1. PID-1-PETISCO spielt nur bei der 21U-RNA-Reifung eine Rolle. TOST-1-PETISCO hingegen hat eine andere, noch unbekannte, jedoch mütterlicherseits essentielle Funktion, die aber die 21U-RNA-Biogenese nicht beeinflusst. Wie PETISCO zur 21U-RNA-Produktion beitragen kann, ist noch unklar.

In meiner Arbeit begann ich mit der Aufreinigung von PETISCO-Proteinen und erstellte funktionierende Antikörper gegen TOFU-6, IFE-3 und PID-3. In Zusammenarbeit mit Sebastien Falk (Wien) wurde die räumliche Struktur von PETISCO aufgeklärt und die Auswirkungen der Wechselwirkungen mit störenden Interaktoren von PETISCO in vivo getestet.

In einem weiteren Projekt beschäftigte ich mich mit der Identifikation eines neuartigen Nukleasekomplex, der den Namen PUCH (Precursor of 21U 5'-end Cleavage Holoenzyme) erhielt und eine wesentliche Rolle bei der Prozessierung von piRNA-Vorläufern spielt. PUCH besteht aus drei Untereinheiten, TOFU-1, TOFU-2 und einem der beiden neu identifizierten Fraktoren des 21U-RNA-Signalwegs, SLFL-3 oder SLFL-4. Das aktive Zentrum von PUCH wird durch "Schlafen"-ähnliche Domänen gebildet, die in allen vier Proteinen vorhanden sind. Die enzymatische Aktivität

von PUCH konnte ich in vivo und in vitro nachweisen. Eine Analyse der Substratanforderungen zeigte, dass PUCH eine m7G-cap-Struktur sowie Uracil in Position drei der piRNA benötigt, um eine Spaltung durchzuführen.

Ebenfalls in Zusammenarbeit mit Sebastian Falk wurde die Beziehung zwischen PETISCO und PUCH untersucht. Die Interaktion zwischen TOFU-6 und TOFU-1 wurde in vitro nachgewiesen und die Kristallstruktur dieser Proteine während der Interaktion konnte aufgeklärt werden. Durch CRISPR/CAS9 Genomeditierung wurde die Interaktion zwischen PUCH und PETISCO gestört und die Bedeutung dieser Interaktion für die 21U RNA-Reifung in vivo bestätigt.

Somit wurde PUCH, in Verbindung mit PETISCO, als neuartiger Prozessierungskomplex des 5'-Endes bei der Reifung von piRNA-Vorläufern in C. elegans identifiziert. Das katalytische Zentrum von PUCH setzt sich aus drei verschiedenen SLFN-Domänen zusammen. Angesichts der Vielzahl nicht charakterisierter SLFN-Homologe in Säugetieren könnte PUCH eine neue Art von Nuklease-Enzymen mit mehreren Untereinheiten und vielen verschiedenen, bisher unbekannten Funktionen darstellen.

List of abbreviations.

AGO	Argonaute
IFE-3	(Initiation Factor 4E (eIF4E) family)
TMG	2,2,7-trimethylguanosine
m^7G	7-methylguanylate
AlbA_2	acetylation lowers binding affinity
AMP	Adenosine monophosphate
Aub	Aubergine
bp	Base pairs
BRCA	BReast CAncer gene
Ce	*Caenorhabitis elegans*
CCR4-NOT	Catabolite Repression—Negative On TATA-less
$CD4^+$	Cluster of Differentiation 4
CD8+	Cluster of Differentiation 8
cGAS	Cyclic GMP-AMP synthase
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
DNMT	DNA methyltransferase
dsRNA	double-stranded RNA
dsRNA	Double-stranded RNA
Endo-siRNA	Endogenous siRNA
eGFP	Enhanced green fluorescent protein
ERH	Enhancer of Rudimentary Homologue
eAGOs	Eukaryotic AGO
eIF4E	eukaryotic translation initiation factor 4E
GST	glutathione-binding protein
GMP	guanosine monophosphate
HENN	HEN-1 of nematode
HUH	histidine-hydrophobic-histidine
H3K9me3	Histone H3 lysine 9 trimethylation
HEN1	HUA enhancer 1
hAGO2	Human AGO2
HIV	human immunodeficiency virus
IMAC	Immobilized Metal Ion Affinity Chromatography
IP	Immunoprecipitation
IP-LFQP	Immunoprecipitation followed by label free quantitative mass spectrometry
LDS	lithium dodecyl sulfate
LTR	long terminal repeats
MHC	major histocompatibility complex

MBP	Maltose-binding protein
Mel	Maternal effect lethal phenotype
mRNA	Messenger RNA
miRNA	microRNA
miRISC	microRNA RISC
MID	Middle
Mrt	Mortal germline phenotype
MALS	multiangle light scattering
NK	Natural Killer
NHEJ	non-homology end joining
NRDE	Nuclear RNAi defective
nt(s)	Nucleotide(s)
ORF	open reading frame
OPVs	Orthopoxvirus
PAMP(s)	pathogen-associated molecular pattern(s)
Piwi	P-element induced Wimpy testis
PBS	Phosphate-buffered saline
PRDE	piRNA-dependent silencing defective
PID	piRNA-induced silencing defective
PRG-1	Piwi (fruitfly) Related Gene
PAZ	Piwi Argonaute and Zwille
PRG	Piwi related gene
piRNA	Piwi-interacting RNA
PARN	Poly(A)-specific ribonuclease
PTGS	Post-transcriptional gene silencing
PTM	Post-translational modification
pAGO	Prokaryotic Ago
RPA	replication protein A
RIG-1 helicase	Retinoic acid inducible gene I
rDNA	Ribosomal DNA
rRNA	ribosomal RNA
RIP	RNA immunoprecipitation
RNAi	RNA interference
RNA Pol	RNA polymerase
RRM	RNA recognition motif
RdRP	RNA-dependent RNA Polymerase
RDE	RNAi defective
RISC	RNA-induced silencing complex
RITS	RNA-induced transcriptional silencing
Slfn	Schlafen
Slfl	Schlafen-like
SINE	short interspersed repeats
ssDNA	single-stranded DNA

ssRNA	single-stranded RNA
SEC	size-exclusion chromatography
SnRNA	small nuclear RNA
SDS PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
STING	Stimulator of interferon genes
TPRT	Target Primed Reverse Transcription
TPRT	Target Primed Reverse Transcription
TRF2	Telomeric repeat-binding factor 2
TIR	Terminal Inverted Repeats
TLR	Toll-like reseptors
TFIIA-L	Transcription factor IIA L
TREX	TRanscription-EXport
tRNA	Transfer RNA
TE(s)	Transposable elements
TGS1	Trimethylguanosine Synthase 1
TOFU	Twenty One u-rna (21U-RNA) biogenesis Fouled Up
WB	Western Blot
WAGO	Worm-specific AGO
Y2H	Yeast-2-hybrid
Zuc	Zucchini

Author contribution.

Sebastian Falk's laboratory performed the crystallization and AlphaFold analysis of PUCH and PETISCO proteins. Jannosh Hennig's laboratory performed protein NMR. Emil Karaulanov analyzed the sequencing data. Emily Nischwitz performed the MS analysis. Walter Bronkhorst obtained IP-PUCH material. Svenja Hellmann assisted me with some microinjections and genotyping. Martin Möckel purified PETISCO antibodies from the sera. Michael Musheev performed the RNA MS runs. I performed the rest of the experiments.

1. Introduction.

1.1. The problem of self and non-self

We are living in a world, where different species have to share their living spaces. Multicellular organisms evolved by being in close contact with bacteria, fungi, other microorganisms, and viruses. While the presence of some of those microorganisms may not be harmful or even beneficial for the species, contact with the other microorganisms can be detrimental and cause serious infections. Therefore, very complex immune systems have evolved, helping organisms to distinguish the presence of pathogenic elements and eliminate them from the host .

There are two kinds of immune systems – innate and adaptive. Innate immunity is present in all multicellular organisms, and it is a major immune system in plants, insects, and primitive multicellular organisms (Gonzalez et al. 2011). Innate immunity provides an immediate, but not long-lasting response for the pathogens, and this kind of immunity is shared by all organisms belonging to the same species. An adaptive immune system is present in vertebrates and provides the possibility of immune response to a wide range of changing pathogens. Each organism develops its immunological memory of the pathogens it meets during life.

1.1.1. Innate immunity.

To recognize the pathogens, organisms should be able to distinguish specific signatures, called pathogen-associated molecular patterns (PAMPs), which are present in the invaders and are absent in the host (CA. Janeway 1992). Recognition of PAMPs relies on the system of germline-encoded receptors and does not depend on previous exposure to the pathogen. One of the most known examples of this system is Toll-like receptors, which recognize the least variable elements of pathogens, and curtail their survival or virulence (Roach et. al. 2005).

The innate immune system not only recognizes “non-self” molecules but also exploits more sophisticated mechanisms, for example, it recognizes changes in “self”, caused by infection. Modifications, performed on self-molecules by pathogenic enzymes trigger the recognition or expression of certain infection-related molecules (Janeway and Medzhitov 2002; Kawai and Akira 2010). Another strategy is recognizing the “missing self, first described for the infected cells, which lost the ability to express major histocompatibility complex (MHC) class I molecules. Such cells were shown to become a target for Natural Killer cells (Karre et. al. 1986). In addition to “missing self”, the innate immune system can recognize “induced or stressed self” – an expression

of the infection-induced molecules in the cell, which helps to prevent the formation of malignant cells (Gasser and Raulet 2006; López-Larrea et al. 2008; Cerwenka and Lanier 2001).

1.1.2. Adaptive immunity.

In many years of battle between pathogens and immune systems, pathogens developed multiple mechanisms helping to avoid the recognition by the innate immune system. In vertebrates, a second kind of immune system appeared, helping to respond to the various numbers of pathogens. The adaptive response starts within 4-6 days after the infection and relies on two mechanisms, provided by two different kinds of lymphocytes – B cells and T cells.

A special class of lymphocytes, B cells is responsible for humoral immunity and targets the toxins and pathogens that are present in the blood or lymph. Each of the B cells possesses a certain kind of receptor, which is capable of recognizing one pathogen. Thus, one B cell is pretty useless, however, a repertoire of the B cells of a healthy organism covers the most pathogens. B cells stay inactive unless they meet the pathogen they can bind. After the activation, caused by binding of the pathogen to the B cell receptor, the B cell divides many times to multiply the response by producing clones. Some of those clones of the activated B cell become plasma cells, which produce antibodies against the pathogen, and some become special long-live cells that provide the memory to initiate a fast response to the pathogen in case the organism meets it again (Bonilla and Oettgen 2010).

T cells provide the response against the host cells, invaded by the pathogen. T cells are also present in an inactive state and activated by interaction with the pathogen. T cells are not capable of the recognition of the pathogen themselves and need signals from the innate immune system to start the response. There are two kinds of T cells – Helper T cell (CD4+), which is not able to kill the infected cell and affects other immune cells by producing cytokines, and killer T cell (CD8+), which is capable of killing the infected cells. T killers are especially important in the viral response (Bonilla and Oettgen 2010).

1.1.3. Immunity against nucleic acids.

Another important aspect of the immune response is a response to the invasive nucleic acids in the application of the antiviral defense. There are many mechanisms, found in both eukaryotes and prokaryotes, to distinguish non-self nucleic acids and eliminate them.

In prokaryotes, one of the most known systems, which helps to recognize foreign nucleic acids is the restriction-modification-based system. The genomes of many bacteria encode modification and restriction enzymes; modification enzymes modify host DNA so it is not cleavable by the restriction enzymes. Invader DNA would be most probably unmodified and be digested by restriction endonucleases (Koonin et al. 2017).

Another example is CRISPR/CAS9, which relies on the encoded in the genome sequences of the nucleic acids and is capable of recognizing the specific target based on the complementarity of the target to the effector RNA (Koonin et al. 2017). Prokaryotic Argonauts can detect invader DNA and cause its degradation (Hur et al. 2014; Koopal et al. 2023). I will discuss prokaryotic Argonautes in more detail in the dedicated chapter below.

In general, responses to invading nucleic acids can be divided into innate and adaptive. The innate immune response relies on the receptors that are encoded in the genome, and the adaptive immune response uses the host's experience to recognize the invaders. Thus, the CRISPR/CAS9 system would be an example of adaptive immunity: even though the effector RNA is encoded by the host, the variations of that depend on the experience of the organism (Rath et al. 2015).

There are several innate immune receptors, that recognize damage-associated nucleic acids based on some features, which are not present in the host nucleic acids. Toll-like receptors TLR3, TLR7, TLR8, and TLR9 are shown to participate in the recognition of nucleic acids (Schlee and Hartmann 2016). Reasons, why nucleic acid could be recognized as non-self are many. Bacterial DNA has been shown to give a strong immune response, and the cause is lower methylation of CpG islands compared to the host DNA (Krieg et. al 1995) and TLR9 is responsible for the detection of unmethylated CpGs (Hemmi et. al 2000). TLR3 recognizes double-stranded RNA (Alexopoulou et. al. 2001), and ssRNA can initiate the immune response provided by TLR7 and TLR8 in the immune cells (Diebold et. al. 2004).

In vitro transcribed RNA causes a stronger immune response, compared to synthetically synthesized RNA, because of the presence of 5’-triphosphate ends (Schlee et al. 2009; Hornung et al. 2006). The presence of the 5‘-triphosphate in the cytosol is a sign of a viral infection, recognition of the 5’-phosphate happens by RIG-1 helicase, which was reported to be the immune receptor (Kato et al. 2005).

DNA in the cytoplasm could be a sign of infection or tumorigenesis, therefore, recognition of its presence in the cytosol and the initiation of the immune response is a very important process. The cGAS-STING pathway is responsible for recognizing DNA and initiating the immune response (Ablasser and Chen 2019). Cyclic GMP–AMP synthase (cGAS) recognizes cytosolic DNA and binds it, the binding results in the synthesis of the message molecule, 2'3'-cGAMP (Wu et al. 2015; Zhang et al. 2013). 2'3'-cGAMP binds the simulator of Interferon Genes (STING), which initiates the start of interferon signaling (Ishikawa et al. 2009; Ishikawa and Barber 2008).

As recognition of foreign nucleic is a crucial function for host survival, there are many factors and pathways involved and I will not mention all of them here (Schlee and Hartmann 2016). At last, I want to mention RNA interference (RNAi) - one of the mechanisms, responsible for the antiviral response in plants and invertebrates (Szittya and Burgyan 2013, (Zhou and Rana 2013). In mammals, RNAi is mostly involved in the host gene regulation and transposon silencing – another kind of invasive nucleic acids, which are already located in the genome. Below, I will discuss the different types of transposons, the RNAi effect, and the piRNA pathway, as a major transposon controller. Conceptually, the piRNA pathway could be seen as the genetic "immune" system against transposons. Later in my results, I will demonstrate the connection between the innate immune system and the production of piRNAs in *C. elegans*.

1.2. Transposons.

Transposable elements (TEs, transposons) have been discovered over 80 years ago by Barbara McClintock. She observed, that phenotype changes in maize were caused by deletions and translocations in maize chromosomes. Further, transposons have been identified in almost all sequenced species from bacteria to humans (Huang et al. 2012).

Transposable elements (TEs), also known as transposons, are often described as 'selfish' DNA sequences because they can both replicate themselves and relocate within the genome. These sequences are highly prevalent and occupy a substantial portion of the genome in many organisms. For example, in maize, nearly 80% of the genome consists of transposons (SanMiguel et al. 1996), and TEs account for approximately 45% of human DNA (Mills et al. 2007). Transposable elements contribute to genome expansion (Kidwell M. 2002). While many TEs have undergone significant sequence modifications over time and are no longer active, some copies of transposons retain their mobility.

1.2.1. Types of transposons.

Transposons have two major ways to change the places in the genome – “copy and paste” and ‘cut and paste”. TEs, acting via a copy-and-paste mechanism, represent Class I of transposons and called retrotransposons, as replication of Class transposons happens via RNA-intermediate molecule. Class II of TEs, also called DNA-transposons, propagates by a “cut-and-paste” mechanism. Most of the transposons encode enzymes, which are required for the transposition and are called autonomous transposons; however in both TE Classes occur elements, that do not encode their own enzyme and their activity relies on autonomous TEs.

1.2.2. Retrotransposons.

Retrotransposons (TEs Class I) are divided into two groups, depending on the presence of the long terminal repeats (LTR): LTR-containing retrotransposons and non-LTR-retrotransposons. Autonomous non-LTR-retrotransposons LINE-1 and non-autonomous *Alu* und Sva elements were found to be mobile in the human genome. None of the LTR-retrotransposons are found active in humans (Mills et al. 2007).

LTR transposons share their transposition mechanism with retroviruses, while non-LTR transposons have a unique transposition mechanism (Eickbush 2007). The two most abundant kinds of non-LTR transposons are autonomous long interspersed repeats (LINEs) and non-autonomous short interspersed repeats (SINEs). Autonomous non-LTR retrotransposons contain one or two ORFs, which encode proteins, required for the transposition (Kolosha and Martin 1997). ORF1 encodes the protein with a nuclear acid chaperone function (Martin and Bushman 2001), which forms a trimer and binds transposon mRNA in the cytoplasm (Khazina et al. 2011). This protein is crucial for the retrotransposition process (Martin 2006). ORF2 encodes reverse transcriptase and nuclease enzymes (Mathias et al. 1991; Feng et. al 1996). The expression level of both ORFs is not equal, expression of the ORF1 is about 27 times higher than ORF2 (Taylor et al. 2013). ORF1 and ORF2 bind transposon RNA in the cytoplasm and get transferred to the nucleus, where the new insertion happens via the Target Primed Reverse Transcription (TPRT) mechanism. ORF2 protein creates a nick in the DNA using its nuclease activity and uses transposon mRNA as a template for reverse transcription (Cost et. al 2002).

SINE elements are transcribed by RNA Poll III and they do not encode their own proteins they can attract the machinery of LINEs to perform the transposition (Weiner 2002)

1.2.3. DNA transposons.

DNA transposons are another class of transposons, which does not use RNA intermediate. Even though DNA transposons are widely present in many organisms, they remain mobile only in several genomes, including Tc1 and Tc3 in *C. elegans* (Emmons et. al 1983; Colins et. al 1989). There are no autonomous DNA transposons left in mammals and the activity of DNA transposons has not been detected in the human genome for the last 50 million years (Pace and Feschotte 2007).

The majority of DNA transposons encode the enzyme, called transposase, flanked by Terminal Inverted Repeats (TIR), oriented in opposite directions to each other. Two molecules of transposase, produced by host expression machinery from the TEs sequence, recognize the TIRs and perform extraction of the transposon from the DNA, resulting in the formation of double-stranded breaks. New insertion occur in any region, containing TA-dinucleotide sequence. New transposon insertions cause duplication of the target region (Muñoz-López 2010). The choice of pathway for the reparation of the double-stranded breaks depends on the host and fully relies on the host machinery. When the breaks are repaired via the NHEJ (non-homology end joining) pathway, the formation of the transposon-specific footprint occurs (Muñoz-López 2010). If the reparation happens by homology recombination mechanism, and the sister chromosome is used as a repair template, a new copy of TE can be generated (Haber 2000).

Copies of transposable elements, which encode inactive transposase or no sequence present at all are known. These elements rely on the expression of the enzyme from another active copy of TEs. As the survival of the transposons depends on the host fitness, and increased activity of transposons reduces the survival of the host, it is hypothesized that inactive copies of transposons are helping to reduce transposition by competing with working copies for the TIR sequences.

Some DNA transposons lack TIR elements and use a rolling-circle mechanism to move across DNA. These transposons are allocated into Sub-Class II of the DNA-transposons, the most studied representative of rolling-circle transposons is the *Helitron* transposon family (Kapitonov and Jurka 2007; Thomas and Pritham 2015). Rolling-circle transposons are mostly present in plants but were also identified in eukaryotes. Instead of transposase, these transposons encode the 'HelRep' enzyme, containing DNA-helicase and replicator initiator protein functions. The

latter combines HUH-nuclease and ligase functions (Ilyna and Koopin 1992). HelRep nicks the donor DNA, peels off transposon-containing sequence, and forms a circle ssDNA. Novel insertion occurs in the TA-dinucleotide area as well, as for other DNA transposons. HelRep cleaves the acceptor strand and ligates the circle DNA to the new spot (Grabundzija et al. 2016).

Unlike standard DNA transposition, there are no double-strand breaks formed and no duplication of the target site happens during the rolling-circle transposition. The gap in the donor DNA is repaired by the host, resulting in the recovery of the initial transposon sequence (Grabundzija et al. 2016).

1.2.4. Negative consequences of the transposition.

The insertion of transposons within coding sequences often leads to disruption of gene expression, resulting in premature stop codons or the generation of incorrect mRNA. Even when transposons later move away from the invaded genes, they leave behind a 'scar' sequence due to the duplication of the transposition target site (Muñoz-López 2010).

In addition to directly affecting coding sequences, transposable elements can also impact enhancers, promoters, and other regulatory elements, resulting in global changes in the genome. For instance, primate transposons can influence DNA methylation status. LINEs and SINEs, characterized by their high GC content, tend to be targeted by the cellular methylation system (Ichiyanagi K 2013). Elevated methylation in a region can lead to the formation of heterochromatin and a reduction in the expression of nearby genes (Estécio et al. 2007).

The somatic activity of transposable elements is associated with the development of different cancer types (Iskow et al. 2010; Solyom et al. 2012). Insertion of transposons into BRCA1, BRCA2, and other genes, involved in the DNA reparation can cause cancer. Besides that, transposons are involved in multiple diseases like Fukuyama muscular dystrophy, hemophilia, and others (Kazazian 1998; Hancks and Kazazian 2012).

Not only transposons, which are already present in the genome are detrimental. Transposons can invade new species and further spread within the population, damaging DNA and reducing chances to have viable progeny (Laski et al 1986; Yu et al. 2019; McCullers and Steiniger 2017).

1.2.5. Positive consequences of transposition.

While most transposon insertions within coding regions are mostly harmful, insertion in the non-coding regions or in between the genes, as with any other mutations can result in a wide range of both positive and negative consequences. As the survival of transposons depends on the successful propagation of the host, some transposons developed mechanisms to preferably translocate to the genome areas, where their presence will be less detrimental (Sultana et al. 2017). The activity of transposable elements made a great impact on the diversity and evolution of many organisms (Sotero-Caio et al. 2017; Dubin et al. 2018) and affected the brain development of mammals (Ferrari et al. 2021).

Transposons can give a start for the novel enhancer elements. For instance, the formation of new enhancers, alternative promotors, or new exons has been shown (Ohtani and Iwasaki 2021).

Besides that, some inactive copies of transposons were domesticated by the host genome. The most known example is the syncytin-A and syncytin-B placenta-specific genes, found in the Muidae family (Dupressoir et al 2005). Another interesting example is envelope proteins with retroviral origin, which now help hosts resist new retroviral infections (Frank et al 2022).

1.2.6. Importance of the germline defense.

For transposons to remain part of the genome and pass it on to future generations, they need to integrate into the germline. However, when transposons become overly active in the germline, it often leads to sterility or a significant reduction in the number of viable offspring. Many transposition events in the germline, particularly those occurring in coding regions, can impact embryonic development and the overall health of the progeny. Even if these translocations do not hinder the development of embryos into healthy adults, the next generation may encounter the same challenges.

The inability of the host to produce offspring can also have negative consequences for transposons themselves, as they rely on host reproduction for their propagation. This has led to a lengthy process of co-evolution between hosts and transposons, during which various defense mechanisms have evolved to prevent harmful mutations (Cosby et al. 2019). This thesis will mainly focus on a TE-defence mechanism known as RNA interference (RNAi).

1.3. RNA interference.

1.3.1. Discovery of RNA interference.

The initial observations regarding the influence of exogenous RNAs on endogenous gene expression were made in plants during attempts to enhance the color of petunias through the introduction of a transgene. Surprisingly, instead of intensifying the color, the petunias became colorless (Van der Krol et al. 1990). At first, the silencing of the host genes was explained by the direct interaction of the anti-sense RNA molecules to endogenous mRNA transcripts, resulting in the inhibition of their translation. However, this explanation required a substantial amount of complementary RNA. Subsequent studies conducted by A. Fire and M. Montgomery in 1998 (Montgomery et al. 1998), demonstrated that injection of dsRNA into the gonads of *C.elegans* is causing a sequence-specific gene silencing on the post-transcriptional level. The silencing effect, caused by dsRNA was at least 10 times more efficient than in similar experiments with sense or antisense ssRNA (Fire et al. 1998). Notably, the amount of injected dsRNA was just about several molecules per cell, which is significantly lower than the level of the target mRNA. Thus, it became evident that mRNA silencing could not be solely explained by simple stoichiometric pairing (Montgomery et al. 1998).

Similar effects of dsRNA on gene expression had been observed in *Drosophila melanogaster* and other organisms, suggesting that gene silencing by dsRNA could be a general mechanism, present in many species (Kennerdell and Carthew 1998; Hammond et al. 2000). This phenomenon has been named RNAi, short for RNA interference. In 2006 Andrew Fire and Craig Mello received a Nobel Prize for the discovery of RNAi interference.

Since then RNA interference has been extensively studied. RNAi has been found to be a conserved mechanism, which is present in most eukaryotes. RNA interference is involved in the regulation of both endogenous parasitic and exogenous pathogenic elements, as well as in the regulation of the translation of protein-coding genes by initiating the mRNA degradation process, and is a promising therapeutic tool (Gregory J. Hannon 2002; Setten et al. 2019)

1.3.2. Mechanism of RNAi.

RNAi process relies on the small RNA molecule and Argonaute protein, which form a riboprotein complex together. This complex recognizes target RNA transcript based on the

sequence homology between small RNA molecules and target RNA. After the recognition, the Argonaute protein initiates the target RNA degradation process (Ketting and Cochella, 2021). Many Argonaute proteins possess nuclease activity and are capable of cleaving target mRNA themselves (PII: S0955-0674(02)00338-1; Liu et al. 2004; Jinek and Doudna 2009). Other Argonaute proteins initiate gene silencing engaging other pathways to inhibit translation or transcription of the target (Maroney et al. 2006; Meister 2007). In this introduction, I am going to mostly discuss RNA interference, performed by eukaryotic Argonaute proteins together with three main types of small RNA: miRNA, siRNA, and piRNA.

1.3.3. Argonaute proteins.

Argonaut proteins were first discovered in plants during the genetic screening of *Arabidopsis thaliana* (Bohmert et al. 1998). Mutants plants looked like small octopuses, therefore the gene had been named after *Argonauta argo – Ago1.* Argonaute proteins were found in most eukaryotes, except Saccharomyces cerevisiae (Drinnenberg et al. 2009), some bacteria and archaea (Makarova et al. 2009; Swarts et al. 2014). Four clades of eukaryotic Argonautes (eAgos) are known – AGO, PIWI, WAGO - Worm specific Asrgonautes, and *Trypanosoma* clade, which is specific for trypanosomatids. AGO clade is expressed ubiquitously in mammalian cells and participates in both viral defense and internal gene regulation in a complex with both miRNA and exogenous and endogenous siRNA (Vaucheret et al. 2004; Wang et al. 2011). PIWI Argonautes are the germline-specific clade, which acts together with piRNA to protect the germline from transposons (Aravin et al. 2006; Kalmykova et al. 2005). WAGO clade – Worm-specific Argonautes, found specifically in *C. elegans*, which bind to the secondary siRNA, produced by RNA-dependent RNA polymerases during the silencing response (Pak and Fire 2007; Yigit et al. 2006). All Eukaryotic Argonaute proteins share a common domain organization.

1.3.4. Domain organization of Argonaute proteins.

Crystal structures obtained for prokaryotic (pAgos) and eukaryotic (eAgos) Argonautes had shown that all AGO proteins share similar structural features for the main four core domains, with some differences in relative domain positions between pAgos and eAgos (Figure I1.A). AGO proteins consist of two lobes. The first lobe contains the N-terminal domain (N-domain) and PAZ domain, named this way because it is present in PIWI, AGO, and ZWILLE proteins. The second lobe

of AGOs contains MID (middle) and PIWI (P element-induced wimpy testis) domains. The niche in between of two lobes has space for short RNA molecules, which can be bound to Argonaute from both sides by PAZ and MID domains via the 3' OH group and 5'phosphate respectively (Yuan et al. 2005; Schirle and MacRae 2012; Nakanishi et al. 2012). The function of the N-domain is less understood and varies in different Argonautes. The PIWI domain has a similar structure to RNAse H (Nakanishi et al. 2012) and some PIWI domains in Agos, carrying DEDX (where X is H, D or K), possess cleavage activity and can perform slicing of the target directly (Yuan et al. 2005; Schirle and MacRae 2012; Nakanishi et al. 2012).

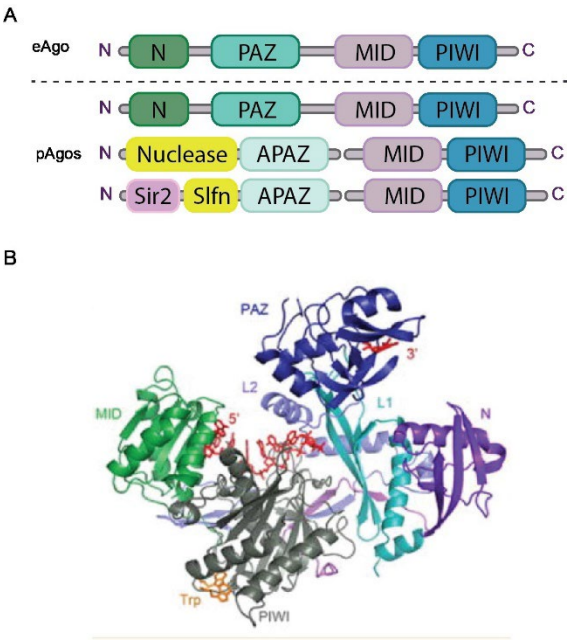

Figure I1. Structure of Argonaute proteins. **A**. Domain organization of eAgo and pAgos. B. Crystal structure of hAgo2 (Schirle and MacRae, 2012).

Most of the eAgos do not have slicing activity, however, RNAi relies not only on the post-transcriptional regulation of translation due to the cleavage of mRNA. In many cases, slicing activity of the Argonautes is not required, as AGOs can engage other pathways to set RNA for degradation via RNA depolyadenylation (Behm-Ansmant et al. 2006); inhibit transcription via heterochromatin formation (Sienski et al. 2012), DNA methylation (Kuramochi-Miyagawa et al. 2008) and translation inhibition (Bazzini et al. 2012).

Different Argonautes in combination with various types of effector ssRNA molecules perform different jobs. Below I will discuss the most common types of small RNAs and their work together with Argonaute proteins.

1.3.5. Argonaute and microRNA pathway.

MicroRNA pathway is mostly involved in the fine regulation of the transcription of host genes. Recondition of the target by miRNA-AGO complex does not require full complementarily of the sequences. The alignment of nucleotides 2-8 of miRNA, also known as a seed sequence, is necessary, while mismatches on the 3' of miRNA can be tolerated. MiRNAs-AGOs usually target 3'UTR of the mRNA, where AGO cannot collide with traveling ribosome. Targeting of the coding sequence or 5'UTR is also happening, but the silencing of the target is much less efficient (Bartel 2018).

Most microRNAs (miRNA) are transcribed by RNA Pol II from defined loci in the genome as pri-miRNA (primary miRNA transcript) in a tissue-specific manner (Lee et al. 2004). These transcripts are polyadenylated and capped, and despite diverse sequences, they possess a common secondary dsRNA stem-loop structure (Nguyen et al. 2015). In the nucleus, this structure is recognized by a microprocessor complex, consisting of Drosha (RNAse III type enzyme) and two molecules of DGCR8, also known as Pasha. Cleavage of pri-miRNA by microprocessor complex leaves the pre-miRNA molecule about 65 nucleotides long (Lee et al. 2003). Pre-miRNA is exported to the cytoplasm by Exportin 5 and RAN GTPases, where it undergoes further processing by Dicer, another protein from the RNAse III family. Dicer processes double-stranded RNA substrate in a sequence-independent manner, leaving the 20-24 nucleotide-long double-stranded RNA fragments 20-24, which are loaded to the Argonaute Proteins (Bartel 2018).

Some of the miRNAs originate from the introns, and instead of Pasha and Drosha, they are processed by the splicing machinery. The secondary structure of these "mitrons" is similar to pre-miRNA enough to be recognized by pre-miRNA exporting machinery and transferred to the cytoplasm, where they undergo the same processing as pre-miRNA (Ruby et al. 2007).

Loading of the dsRNA fragment to the Argonaute proteins happens with the assistance of Hsc70/Hsp90 chaperons, which help to change the conformation of the Argonaut protein to open the access of dsRNA to the RNA binding area (Iwasaki et al. 2010). However, when chaperons are disconnected and Ago returns to the initial conformation, the niche becomes too small to accommodate both strands of the dsRNA and one of the strands leaves the complex. The choice of the strand, which is loaded to AGO not random and depends on the thermodynamic stability of the 5'end of the RNA molecule (Schwarz et al. 2203).

After recognition of the target, the AGO protein attracts the GW182 protein, and this gives a start to the formation of an RNA-induced silencing complex (RISC). GW182 proteins are conserved amount metazoans and characterized by multiple glycine-tryptophan dipeptides, necessary for the interaction with other proteins, including Argonauts (Dexheimer and Cochella 2020). RISC complexes are very diverse and may include various RNA degradation factors. Plants do not encode a direct ortholog of GW182, however, RISC complex assembly may be promoted by another GW-containing protein SUO. GW182 proteins act as a bridge between AGO and other players of the RICS complex, for instance, GW182 has been shown to attract CCR4-NOT and PAN2-PAN3 deadenylation complexes, which also attract decapping factors (Braun et al. 2012). It seems, that silencing of RNA transcripts by miRNAs relies mostly on the RISC complex, and does not require cleavage activity of AGOs. For instance, the human genome encodes four different AGO proteins, and only Ago2 has slicing activity, miRNAs are associated with all four Agronautes and all four are capable of target silencing (Peters and Meister 2007).

1.3.6. Argonaute and siRNA pathway.

Exogenous siRNAs are mostly produced from viral RNA. Endogenous miRNAs and siRNA share some similarities in the processing and further silencing activities, however siRNA pathway is less conserved and the origin of ssRNA may be different in different species. Initially, it had been thought that endogenous ssRNA are present only in those species, that encode RNA-dependent RNA-polymerases (RdRPs), like *C. elegans* or fission yeasts (Smardon et al. 2000). Later, endogenous siRNA was identified in the other species too (Okamura et al. 2008; Czech et al. 2008).

In species, lacking RdRPs, siRNA production is happening independently of the microprocessing complex, but the double-stranded RNA precursor of siRNA is still processed by Dicer. SiRNA pathway is best studied in Drosophila because *Drosophila* genome encodes two homologs of Dicer – Dicer-1 and Dicer-2, and while Dicer-1 is processing pre-miRNA, Dicer-2 is selectively cleaving siRNA precursors. Further, these two pathways are also separated – while miRNA is loaded to Ago1, siRNA is associated with Ago2. There are two sources of double-stranded RNA for siRNA precursors in *Drosophila* – bidirectional transcription from the same locus and transcription from structural loci, which results in a hairpin secondary structure (Okamura et al. 2008). In mammalians in addition to hairpins and bidirectional transcripts were also proposed the possibility of the production of dsRNA precursor for siRNA production from distinct loci in the

genome, like genes and pseudogenes. Studying the processing and functions of siRNA in vertebrates is difficult, as the presence of long dsRNA can trigger interferon response, therefore function of siRNA is not that well studied. SiRNA function has been associated with defense against foreign sequences. In *Drosophila*, compromising the siRNA pathway caused reduced spermatogenesis due to transposon reactivation and subfertility (Lin et al. 2018).

1.3.7. PIWI clade and piRNA.

PIWI-interacting RNA (piRNA) is an animal-specific class of small RNAs, which acts together with PIWI proteins in animal gonads. The main role of the PIWI-piRNA complex is to protect the germline against foreign invaders or transposons. Intact genetic material is crucial for animal propagation, and in most species, deficiency in the piRNA pathway leads to sterility or severe fertility impairment of the organism. PiRNA pathway is present in most animals and relatively conserved, although some variations in piRNA biogenesis exist. In most species, piRNA is produced as a long single-stranded RNA precursor molecule, transcribed from a special cluster in the genome by RNA Pol II and processed further into single piRNA in the Dicer-independent manner. In this Thesis, I focus on piRNA biogenesis in *C. elegans*, and how piRNA biogenesis in *C. elegans* differs from the other organisms, I will also discuss the piRNA pathway in *Drosophila*.

1.3.7. Prokaryotic Argonauts.

Prokaryotic Argonautes (pAgos) have been found in around 30% of Archaea and about 10% of bacterial genomes (Swarts et al. 2014; Makarova et al. 2009; Lisitskaya et al. 2018). PAgos are much more diverse than eAgos and have more functions. In contrast to eAgos, which bind only ssRNA molecules and affect the fate of RNA targets, different pAgos in addition utilize RNA and DNA oligonucleotides and are capable of RNA-dependent DNA interference and DNA-guided DNA interference (Swarts et al. 2014; Bobadilla Ugarte et al. 2023).

1.3.7.1. Domain organization of pAgos and their functions.

PAgos can be divided into three different groups – long-A and long-B pAgos and short pAgos (Ryazansky et al. 2018). Long pAgos have a similar domain organization as eAgos and contain four major domains – N, PAZ, MID and PIWI (Figure I1.B). While almost all A-pAgos are catalytically active, B-pAgos do not contain catalytic DEDX tetrad and therefore do not possess catalytic activity (Makarova et al. 2009; Swarts et al. 2014). Short pAgos often lack an N-terminal

part and contain only MID-domain and catalytically inactive PIWI domain (Koopal et al. 2022). Some short pAgos are fused to the analog of PAZ domain (APAZ), or encoded in the same operon with it. This APAZ domain has no sequence similarity to the PAZ domain of eAgos and is often fused to SIR2, Toll/interleukin-1 receptor (TIR)-like domain, or other nuclease domains (Makarova et al. 2009; Swarts et al. 2014; Ryazansky et al. 2018)(Makarova et al. 2009; Swarts et al. 2014)(Makarova et al. 2009; Swarts et al. 2014)(Makarova et al. 2009; Swarts et al. 2014)(Makarova et al. 2009; Swarts et al. 2014). Interestingly, in some bacteria, APAZ domains contain a Schlafen-domain (see Chapter 1.6.) prior to SIR2 (Swarts et al. 2014).

The main function of pAgos is closest to the function of PIWI Argonaute proteins - defense against invaders such as plasmids and viruses (Swarts et al. 2014). Recently it has been shown, that short pAgos together with TIR-APAZ form a complex, named Sparta, which promotes the elimination of bacteria, containing plasmids with high copy numbers. Recognition of the plasmid DNA is performed by short pAgo in the RNA-guided manner, and the TIR-domain is responsible for the cleavage of the $NAD(P)^+$, and its activation causes the depletion of $NAD(P)^+$ and cell death, helping to eliminate bacteria, containing high copy plasmids from the population(Koopal et al. 2023; Koopal et al. 2022).

1.3.7.2. pAgos application in molecular biology.

The ability of pAgos to target and cleave DNA makes PAgos a promising tool for genome editing, which can be potentially more versatile than CRISPR/CAS9, as it does not require the presence of a PAM sequence (Lisitskaya et al. 2018; Hegge et al. 2018). However, eukaryotic DNA is very compact, and that might prevent the access of pAgo proteins to the desired regions. It had been speculated, that a high level of condensation could be a reason why regulation, performed by eukaryotic Argonautes is performed on the RNA level. There is some evidence that eAgos, for instance, eAgo2, is able to bind DNA oligonucleotides and plant Ago4 is capable of recognizing DNA targets and causing heterochromatin formation in order to repress transcription (Lahmy et al. 2016). Thus, maybe the ability of DNA binding by eAgos was not lost, and the creation of CRISPR/CAS–like tools, based on the prokaryotic Argonautes and DNA molecules is possible.

1.4. PiRNA pathway in *Drosophila.*

Drosophila piRNA pathway is necessary for gonad development and fertility both in male and female flies (Cox et al. 1998; Kalmykova et al. 2005). Three PIWI proteins are expressed in *Drosophila*: PIWI, Aubergene (Aub) and AGO3, the function of these proteins is not redundant, as the knockout of each of them causes sterility (Cox et al. 1998). PIWI localizes to the nucleus, where it acts on the transcriptional level, guiding the placement of repressive methylation marks (Le Thomas et al. 2014; Akkouche et al. 2017; D. N. Cox and A. Chao and H. Lin), while Ago3 and Aub act in the cytoplasm. Nuclear localization of PIWI protein strongly correlates with loading with piRNA (Yashiro et al. 2018).

Most of the studies of the *Drosophila* piRNA pathway had been done in ovaries. *Drosophila* ovaries consist of two kinds of cells – somatic (follicular) and germ (nurse) cells. PiRNA pathway has been well studied in cell lines originating from both types of ovarian cells and some significant differences in piRNA pathway in these two cell types have been found. For instance, PIWI is expressed in the whole ovary, while Aub and Ago3 are germ cells specific (Aravin et al. 2007; Malone et al. 2009).

Drosophila's piRNAs are 23-29 nucleotides long. PIWI and Aub-associated piRNAs are mostly antisense to the transposons and have 5'U bias, while Ago3 bound piRNAs have adenine in the position 10 counting from the 5'end (Aravin et al. 2007; Gunawardane et al. 2007). 5' ends of *Drosophila* piRNAs are phosphorylated, and 3'end carries 2'-*O*-methylation (Horwich et al. 2007).

PiRNA originate from long ssRNA molecules, transcribed by RNA Pol II from two different types of clusters, harboring repetitive elements and transposons (Aravin et al. 2007). Uni-strand clusters are transcribed only in one direction, starting at the promotor. Uni-strand clusters are mostly used to produce piRNA in somatic cells and these clusters are located in the euchromatic regions of the genome. *Flamenco* is one of the most known uni-strand piRNA clusters in *Drosophila* (Goriaux et al. 2014).

Production of piRNA from dual-strand clusters happens mostly in the ovarian cells. These clusters can be transcribed bi-directionally and do not have their own promoter (Mohn et al. 2014). Dual-strand clusters carry repressive methylation marks and possess heterochromatin structures, placed there by PIWI (Akkouche et al. 2017; Le Thomas et al. 2014). Despite the

difference in the biogenesis, mature piRNAs deriving from both cell types have similar structural features: both are 23-30 long and carry 5'P and 3' 2'-*O*-methylation marks. Sequences of piRNAs are also not unique, and as most piRNAs have derived from transposons, they map to more than one region in the genome (Aravin et al. 2007).

Transcription of piRNA from dual-stranded clusters depends on Rhino (Rhi), a homolog of heterochromatin protein HP1a, which recognizes the H3K9 methylation mark and brings alone two co-factors: Deadlock (Del) and Cutoff (Cuff) (Mohn et al. 2014; Ryazansky et al. 2018). Deadlock attracts proteins Moonshiner (Moon) and TRF2, which subsequently recruit transcription factors TFIIA-L and TFIIA-S, which allow transcription by RNA-polymerase II in the promotor-independent manner. As transcription from dual-cluster can start at almost any place, piRNA precursors have various lengths. Cutoff, together with TREX complex, protects PiRNA transcripts from splicing machinery, and as a result, piRNA transcripts are not spliced and not poly-adenylated (Pritykin et al. 2017; Chen et al. 2016).

Cutoff prevents RNA polymerase II from termination, and that allows continuing transcription over new TE insertions, which carry poly-A or termination signals. Thus, dual-strand clusters work as traps for transposons: as soon as insertion of a transposon happens to occur in the piRNA cluster, its sequence integrates into pre-piRNA transcripts and becomes a template for piRNA production (Chen et al. 2016).

After transcription, PiRNA precursors are exported from the nucleus for further processing. Bootlegger (Boot), a protein that can bind UAP56, a component of the TREX complex, brings piRNA precursors to the Nxf3/Nxt1, which ensures its translocation to nuage, where further piRNA biogenesis in germ cells is believed to happen (ElMaghraby et al. 2019).

Uni-strand piRNA clusters are mostly used for piRNA production in somatic ovarian cells. As a product of canonical transcription, these precursors are spliced and polyadenylated, and share processing machinery with regular mRNAs.

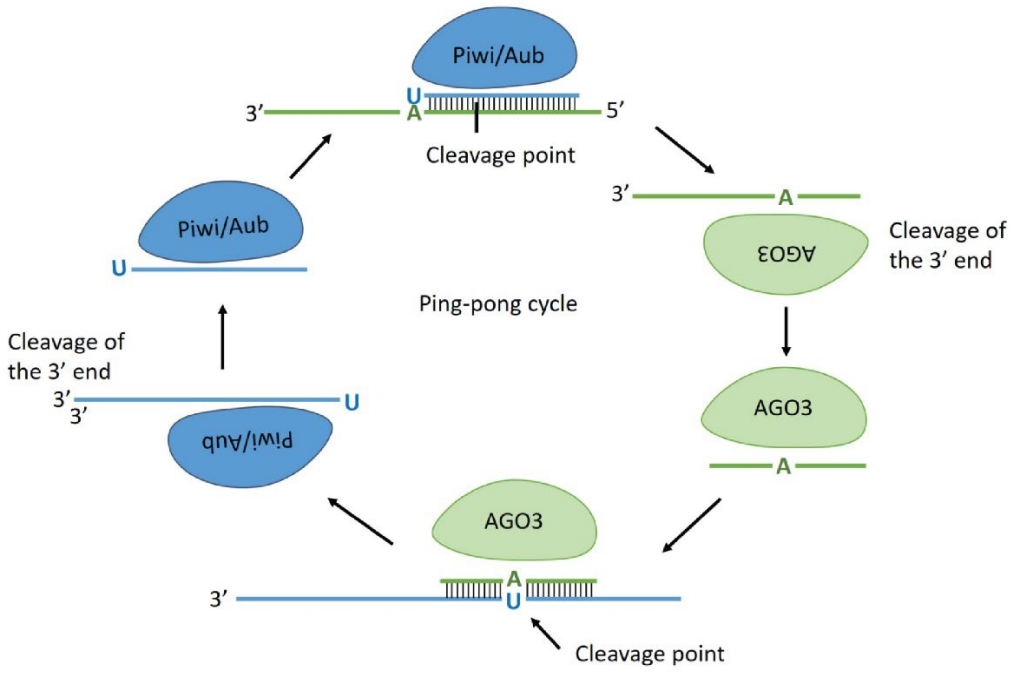

Figure I2. Ping-Pong cycle in *Drosophila*. Piwi/Aub and AGO3 act together to produce piRNA.

In somatic cells, all 5’ end of piRNAs are generated by the endonuclease Zucchini (Zuc). Zuc is located on the membrane of the mitochondria and has been shown to cleave precursors in vitro (Ipsaro et al. 2012). Zucchini did not show substrate specificity cleavage *in* vitro, however, PIWI-bound piRNA have a strong 5’U bias. Thus, it is possible, that *in vivo* piRNA precursor cleavage with generation of 5’U is regulated by an unknown co-factor of Zucchini, or cleavage of precursors by Zuc is indeed non-specific, and selection of 5’U-carrying transcripts depends on PIWI itself. The 3’ end of piRNA can be generated either by Zucchini or as a by-product of the next cleavage event, producing the next 5’end of piRNA, also known as phasing (Mohn et al. 2015). In the other cases, the 3’end of the piRNA still needs to be trimmed, and this job is done by exonuclease Nibbler (Nbr) (Wang et al. 2016). After final adjustments, piRNA is loaded to the PIWI protein, and its 3’end is 2-O-methylated by HEN1 (Horwich et al. 2007).

In addition to Zucchini-mediated cleavage, in germ cells, piRNA can be also generated by the ping-pong mechanism or ping-pong cycle (Aravin et al. 2007; Gunawardane et al. 2007)In this cycle, Aub and Ago3 used their slicing activity to produce new piRNA and amplify the silencing response. Ago3, bound to the sense piRNA, recognizes and cleaves anti-sense piRNA cluster transcript; generated by Ago3 cleavage 5’-product, which mostly starts with U, get loaded to Aub.

After adjustment of the 3’ end of the transcript, Aub, on its turn, recognizes sense pre-piRNA, cleaves it, and this new cleavage product, containing A in position 10, is loaded to Ago3 to produce new piRNA. The ping-pong is evolutionally conserved in many species including humans. The advantage of the ping-pong cycle over the simple cleavage of the cluster transcripts is in an increasing population of piRNAs, acting against active transposons, as transposon mRNA can enter the ping-pong cycle directly and serve as a template for piRNA production.

It is not yet known how exactly mRNA is picked for piRNA production, and how is the choice between translation and piRNA production is made. The most common theory suggests that any mRNA can enter piRNA production if it co-localizes together with piRNA processing machinery.

PiRNA processing nuclease Zuc is anchored to the mitochondrial membrane, and other piRNA biogenesis factors like GASZ and Minotaur are also localizing there (Munafò et al. 2019). The role of mitochondrial localization of piRNA biogenesis factors and mitochondria still needs to be investigated.

1.5. PiRNA pathway in *C. elegans*.

C. elegans genome encodes 27 Argonaute proteins, however, some of them are probably pseudogenes. 18 of those Argonaute proteins belong to the special clade of Argonautes - WAGOs – worm-specific Argonaute proteins. These Argonaute proteins act together with different classes of small RNAs, and most of those pathways are unique for nematodes. Different Argonauts do not have specific targets, and several AGOs can target one RNA molecule at the same time.

In general, Argonaute proteins in worms can be divided into primary Argonautes – those, that participate in recognition of the target transcripts and secondary – those that are binding to 22G RNA, a class of worm-specific RNAs, that are amplified by RNA-dependent RNA polymerases (RdRPs) from the target, recognized by primary Argonautes. Thus, small RNA pathways are working together, defending the integrity of the genome and preventing unwanted gene expression.

1.5.1. 21U RNA biogenesis in *C. elegans*.

PiRNA pathway in nematodes is different from other species and amplification of piRNAs does not happen via the ping-pong cycle, but via the activity of RdRPs, which produce secondary

siRNA from the target. Unlike *Drosophila*, piRNAs are not transcribed as long precursors, which are further cleaved to produce multiple piRNAs, in *C. elegans* they are transcribed individually. The absence of piRNAs does not result in immediate sterility in *C. elegans*, but cases of mortal germline phenotype (Mrt) which appears in reduced fertility, which decreases with generations (Wang and Reinke 2008).

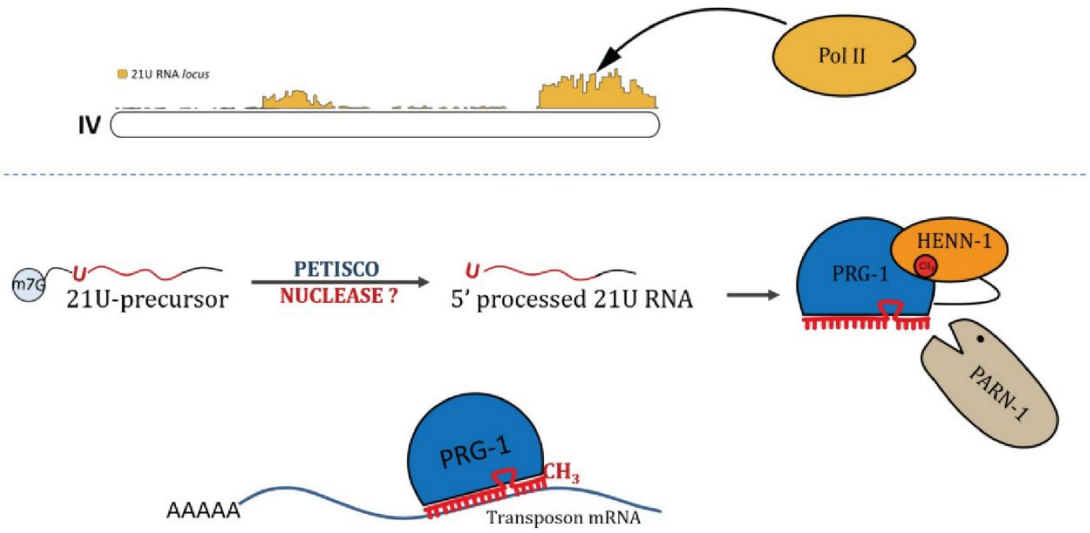

Figure I3. 21U RNA biogenesis in *C. elegans*. 21U RNAs are transcribed by RNA Pol II and processed on the 5'end by unknown nuclease. After the 5'end processing, the RNA is loaded to PRG-1, where the 3'end of it is trimmed by the exonuclease PARN-1. The final step of 21U RNA processing is methylation of the 3'end by HENN-1. Afterwards, PRG-1-21U RNA complex is ready to recognize the target transcript.

The genome of *C. elegans* encodes more than 30,000 individual piRNAs (Batista et al. 2008; Lee et al. 2012). In *C. elegans* piRNAs are called 21U RNA because they are 21 nucleotide long and have a strong 5'U bias (Ruby et al. 2006). 21U RNAs are divided into two groups – type I and II depending on their origin. Type I 21U RNA originates from two clusters on Chromosome IV and is transcribed by RNA Pol II with the help of the UTSC complex. 21U RNA type II is transcribed bi-directionally from the transcription start sites of the coding genes. (Cecere et al. 2012; Ruby et al. 2006; Lee et al. 2012). Mature 21U RNAs have 5' Phosphate and 2-O-methylation on the 3' end.

Promotor of the majority 21U RNA type I contains consensus Ruby motif – CTGTTTCA. Transcription factor SNPC-4, associated with the Ruby motif, together with other factors including PRDE-1, TOFU-4, and TOFU-5 recruits Pol II to perform the transcription (Weng et al. 2019). RNA

Pol II transcribes 21U RNAs as individual mini-genes. Most of the transcripts, generated by RNA Pol II in *C. elegans* are trans-spliced and, therefore, carry a 2,2,7-trimethylguanosine (TMG) cap. Each of the 21U RNA precursors is also capped, however as precursors do not undergo trans-splicing, most probably they are carrying an m^7G cap. Transcription of precursors usually starts at the YRNT motif (where T represents 5'U in 21U RNA), two nucleotides before the sequence of mature piRNA – before the U – starting nucleotide for most of 21U RNA in *C. elegans*, and stops 3 to 5 nucleotides after mature 21U RNAs sequence. Termination of transcription happens due to pausing of RNA Pol II and is not specific, therefore 21U RNA precursors have different amounts of nucleotides on the 3' end (Beltran et al. 2019).

After transcription, 21U RNA precursors are transported to P-granules, perinuclear nuage-like structures, where an unknown nuclease is processing the 5' end, cleaving the cap and two first nucleotides. Knock-out of proteins TOFU-1 and TOFU-2, identified in RNAi screen for factors, participating in 21U RNA precursor processing, showed the absence of mature 21U RNA and accumulation of 21U RNA precursors, suggesting, that these proteins are participating in the 5'end processing of 21U RNA precursors. However, no domains with potential nuclease activity were identified in TOFU-1 and TOFU-2 (Goh et al. 2014). Another mutagenic screen identified the PID-1 protein, the deletion of which affected the maturation of precursors (Albuquerque et al. 2014). PID-1 protein interacts with a complex named PETISCO, and most probably PETISCO it is a platform for 5'end precursor processing (Cordeiro Rodrigues et al. 2019; Zeng et al. 2019).

After 5'end processing, 21U RNA is loaded to PRG-1, a main PIWI protein in *C. elegans. C. elegans* genome encodes two PIWI proteins – PRG-1 and PRG-2, however only PRG-1 is shown to be functional. PRG-1 might have a slicing activity, however, there is no evidence that it performs cleavage of the targets. Precursors, loaded to PRG-1, undergo 3' ends trimming by exonuclease PARN-1 until the length of 21 nucleotides. The last step of 21U RNA formation in *C. elegans* is methylation of the 3'end by HENN-1 (Billi et al. 2012; Bagijn et al. 2012). Mature 21RNAs have 5' phosphate and 3' O-CH_3 group (Figure I3).

1.5.2. Downstream effect of 21U RNAs.

PRG-1 loaded with 21U RNA can recognize the target transcript with an allowance of up to four mismatches between 21U RNA and the target (Bagijn et al. 2012; Lee et al. 2012). Even though PRG-1 has a conserved catalytic center, it seems that the slicing activity of PRG-1 is not

required for target silencing (MP Bagijn, 2012). PRG-1 and 21U RNA pathways are responsible for the initiation of the silencing response, but not for its maintenance. Recognition of the target by PRG-1 loaded with 21U RNA recruits RNA-dependent RNA polymerases (RdRPs) for the secondary class of small RNA production - 22G RNAs, 22 nucleotide long transcripts with G at the 5' end These RNAs, complementary to the target, are loaded to secondary Argonaute proteins, which perform the silencing.

C. elegans genome encodes four different RdRPs, but only RRF-1 and EGO-1 are producing 22G RNAs. There are some slight differences between these two RdRPs, and RRF-1 is a main producer of 22Gs in response to the PRG-1 target recognition, while EGO-1 produces 22G RNA for CSR-1 pathway (Maniar and Fire 2011; Phillips et al. 2012), however, EGO-1 is somewhat compensating for the loss of RRF-1 (Bagijn et al. 2012; Gu et al. 2012).

RRF-1 protein is localizing to the *Mutator* foci, and the amplification of 22G RNAs is dependent on the *Mutator* foci (ketting et al. 1999; Phillips et al. 2012).

22G RNAs act together with worm-specific Argonauts WAGOs to execute the silencing. WAGOs can silence their target at both the transcriptional and post-transcriptional levels. WAGO proteins do not possess slicing activity. WAGO-1 and WAGO-4 are localized to the cytoplasm, where they perform target silencing on the post-transcriptional level (Gu et al. 2009; Yigit et al. 2006), HRDE-1 (WAGO-9) and NRDE-1 perform the silencing on the transcriptional level (Buckley et al. 2012). HRDE-1 is able to initiate the deposition of the heterochromatin marks on the target, by recruiting NRDE-1, NRDE-2, and NRDE-4 proteins (Buckley et al. 2012; Burkhart et al. 2011; Mao et al. 2015).

After establishment, WAGO-mediated silencing becomes independent from PRG-1 and can be inherited for many generations. This effect is called RNA-induced epigenetic silencing (RNAe), and *Mutator* foci, as well as WAGO proteins activity required for it (Ashe et al. 2012; Maartje J Luteijn et al. 2012; Shirayama et al. 2012).

1.6. Schlafen proteins.

Schlafen genes were discovered in mice in 1998, as genes responsible for maintenance of the quiescent state of T cells. Therefore, this gene family has been named "Schlafen", which means "to sleep" in German, reflecting the dormant state of the T cells (Schwarz et al. 1998). Besides being involved in T cell development, it has been shown that these cells are involved in

thymocyte maturation and cell growth regulation (Schwarz et al. 1998). Every analyzed mammalian genome, except platypus, contains at least two copies of Schlafen genes. Besides mammalians, Schlafen proteins were identified in the African clawed frog *Xenopus laevis*, in the elephant "fish" *Callorhinchus milii,* and in the viruses from the orthopoxvirus family. (Bustos et al. 2009).

Most of the research had been focused on human and mouse Schlafen genes. In mice, 9 Schlafen genes had been identified – *Slfn1*, *Slfn2*, *Slfn3*, *Slfn4*, *Slfn5*, *Slfn8*, *Slfn9*, *Slfn10*, and *Slfn14*. These genes are located on chromosome 11 in the Om region, the absence of which is responsible for the DDK syndrome formation (Timothy A. Bell, 2005). *Slfn10* is a pseudogene and does not encode a protein (Bustos et al. 2009).

In the human genome, six Schlafen genes *SLFN5*, *SLFN11*, *SLFN12*, *SLFN12L*, *SLFN13,* and *SLFN14* are clustered on Chromosome 17. Phylogenic analysis showed that *SLFN3/4* are direct orthologues of *SLFN12* and *SLFN12L*, *SLFN5* is an orthologue of *SLFN14*, and mouse *SLFN8/9/10* are orthologues of human *SLFN13*. There are no orthologues of *SLFN11* found in mice, however, it was detected in many other mammals except humans (Al-Marsoummi et al. 2021; Bustos et al. 2009). Besides canonical Schlafen proteins, genes that contain only part of the SLFN-gene, named *Slfn1L* and *SLFN1L* were identified both in mice and humans on chromosomes 4 and 1, respectively (Bustos et al. 2009).

1.6.1. Domain organization of Schlafen proteins.

Based on the domain organization, Schlafen proteins are classified into three groups. Conventional Schlafen proteins contain a unique N-terminal structure called Schlafen-box (SLFN-box). The other domains vary between different Schlafen groups (Figure I4).

1.6.1.1. Schlafen proteins belonging to the group I contain SLFN-box only.

All proteins, belonging to group I, contain SLFN-box only. The human genome does not encode proteins, which belong to this group. In mice, proteins SLFN1 and SLFN2 belong to the group I (Bustos O, 2009). Initially, the SLFN-box had been annotated as an AAA ATPase-associated domain based on sequence similarities (Lupas and Martin 2002; Chaturvedi et al. 2014). However, later studies revealed that the Schlafen box possesses a nuclease domain with a structure unique to the SLFN proteins (Yang et al. 2018). The structure of the Schlafen box consists of two lobes,

however, the catalytic center of the Schlafen nuclease is located in only one of them. This Schlafen box is present in all canonical Schlafen proteins, but in some of the SLFN proteins nuclease center is not conserved, while others have been shown to cleave DNA and RNA (Jo and Pommier 2022).

1.6.1.2. Schlafen proteins belonging to group II contain SLFN-box and SWADL domain.

In addition to the N-terminal SLFN box, Groups II and III contain a conserved SWADL motif, which is located on the C-terminal side of the SLFN box. Human proteins SLFN12 and SLFN12L, and mice SLFN3 and SLFN4 belong to group II. SWADL motif is named this way because it contains conserved Ser-Trp-Ala-Asp-Leu amino acids sequence. It is often referred to as the M (middle)-domain and its function is not yet known (Metzner et al. 2022b; Metzner et al. 2022a). Some studies claim that this domain could be involved in protein-protein interactions and ribosome binding, however, that would not be possible for all Schlafen proteins, as in Schlafen12 for instance, the SWAVDL-domain is not accessible from the surface and located inside of the protein (Chen and Kuhn 2019).

Schlafen14 has a different connection between the M-domain and C-terminal helicase domain (domain, which only Schlafen proteins belonging to Group III have; see 1.1.3) compared to other SLFN proteins, which makes the SWADL motif available for potential interaction with ribosomes (Pisareva et al. 2015). As Schlafen13 and Schlafen11 have been reported to regulate rRNA levels and/or interact with ribosomal proteins respectively, the SWAVDL motif may be involved in the interaction with ribosomes (Jo and Pommier 2022). These variations in the structure of the SWAVDL motif in different Schlafen proteins one more time pinpoint the possible differences between the functions of Schlafen proteins.

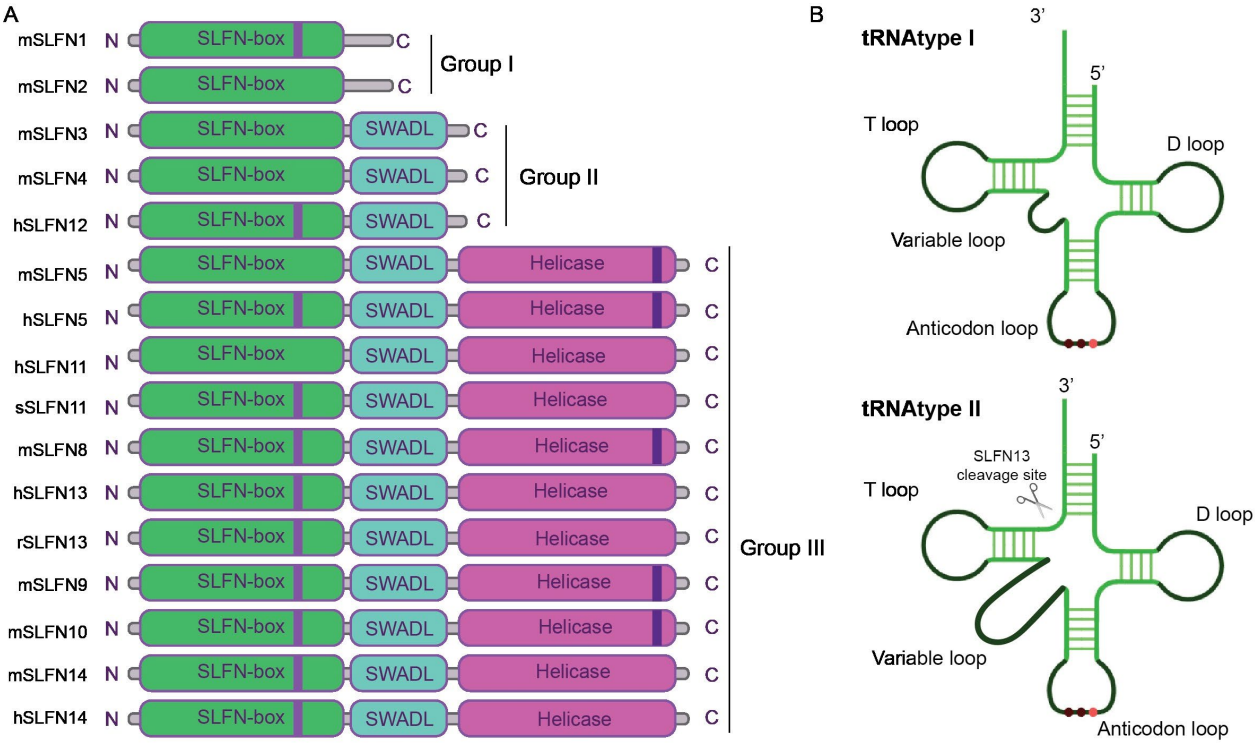

Figure I4. Domain organization of SLFN proteins. **A**. Schematic representation of domain organization of Schlafen proteins. The lines indicate low-complexity regions and the rectangles indicate the predicted folded domains. The purple line represents the conserved nuclease catalytic center, violet line putative helicase domain. **B**. Structure of the tRNA type I and type II. SLFN13 cleavage site is marked on the tRNA type II.

1.6.1.2. Schlafen proteins belonging to group III contain an additional helicase-like domain on the C-terminus.

Schlafen proteins, belonging to group III, possess an additional C-terminal domain, with sequence homology to the family of SF1 DNA/RNA helicases (Geserick et al. 2004) and some of them contain nuclear localization signals. The C-domain of long SLFN proteins contains conserved Walker A and B motifs, suggesting ATP-binding and ATPase activity, and ssDNA binding activity, the latter had been confirmed for Schlafen11 (Metzner et al. 2022b). In Humans SLFN5, SLFN11, SLFN13, and SLFN14 proteins belong to Group III. In mice, genes *Slfn5*, *Slfn8, Slfn9*, and *Slfn14* encode group III Schlafen proteins.

1.6.1.2. V-Schlafen and Schlafen-like proteins.

Besides the three groups of canonical Schlafen-proteins, Schlafen-like proteins, containing only parts of the N-terminal Schlafen box, were identified in many genomes. In the human

genome, for instance, such a gene named *SLFN1L* is located in chromosome 1 and not clustered with other *SLFN* genes. This protein contains an unknown domain on the N-terminus and has a partial SLFN box on the C-terminus.

Even though Schlafen proteins as they are found in humans are present almost exclusively in mammals, partial Schlafen-like structures were identified in many different species including archaea, where Schlafen-like domains are present in several prokaryotic Argonaut proteins (Section 1.3.7.) (Swarts et al. 2014). Interestingly, besides mammals and several exceptions mentioned above, copies of the Schlafen-related *v-slfn* gene were identified in many viruses from the orthopoxvirus family (Bustos et al. 2009).

1.6.1.2.1. *V-slfn* proteins are present in the orthopox viruses.

One of these non-canonical Schlafen proteins is present in the family of orthopox viruses (OPVs) and is called v-schlafen (*v-slfn*) (Schwarz et al. 1998; McLsaght et al. 2003; Gubser et al. 2007). A single copy of the *v-slfn* gene is present in all analyzed genomes of the OPVs, where it is located in the orthologous position (Bustos et al. 2009). In orthopoxviruses the v-slfn domain is attached to the poxin domain, a viral cyclic guanosine monophosphate adenosine monophosphate (cGAMP) nuclease, involved in inactivating the cGAS–STING pathway (Liu et al. 2019; Eaglesham et al. 2019). This pathway is a part of the innate immune system and activates in response to the presence of free DNA in the cytosol, as free DNA can be a sign of viral infection or tumorigenesis. Thus, poxin prevents the early recognition of the infection by the host and delays the immune response, which results in the enhancement of the infection, which had been shown for the Camelpox virus (Hernaez et al. 2020; Gubser et al. 2007). However, the exact role of the *v-slfn* fold in the poxin-v-slfn fusion remains unknown.

1.6.2. Expression and localization of the Schlafen proteins.

In mice and humans, Schlafen genes are clustered together; however, the transcription of Schlafen genes is independent of each other and varies in different cell types. Human SLFN5, SLFN12L, and SLFL13 are mostly expressed in T cells, while SLFN11 is more abundant in monocytes and monocyte-derived cells (Puck et al. 2015). Schlafen genes belong to the interferon-stimulated genes; however, the level of interferon-stimulated upregulation also differs between the proteins (Katsoulidis et al. 2009).

Localization of the Schlafen proteins also differs between the proteins proteins. Schlafen belonging to Group I and II are mostly located in the cytoplasm, while Schlafen proteins from Group III are predominantly located in the nucleus, as most of them possess nuclear localization signals, except SLFN13, which localizes in the cytoplasm (Neumann et al. 2008; Yang et al. 2018).

Interestingly, according to the Alphafold predictions SLFN12L contains a transmembrane domain on the C-terminus right after the SWADL domain (UniProt, Q6IEE8).

Thus, different domain organization, as well as different localization and variations in the expression levels suggest that despite high homology these proteins are not redundant and participate in the different biological processes.

1.6.3. Enzymatic activities of Schlafen proteins.

Schlafen proteins are believed to have two enzymatic functions. Since Schlafen proteins were discovered, the SLFN-box was wrongly identified as an AAA ATPase-associated domain (Lupas and Martin 2002; Chaturvedi et al. 2014). Later, it was found that the Schlafen N-box forms a nuclease domain, which is related to the AlbA_2 domain and unique for Schlafen proteins (Yang et al. 2018). Besides the nuclease domain, Schlafen proteins belonging to group III exhibit homology to the SF1 DNA/RNA helicases, suggesting putative helicase, ATPase, and nucleic acid binding activity (Geserick et al. 2004). I will summarize the knowledge about the Schlafen enzymatic activity and its biological functions below.

1.6.3.1. Nuclease function of Schlafen proteins.

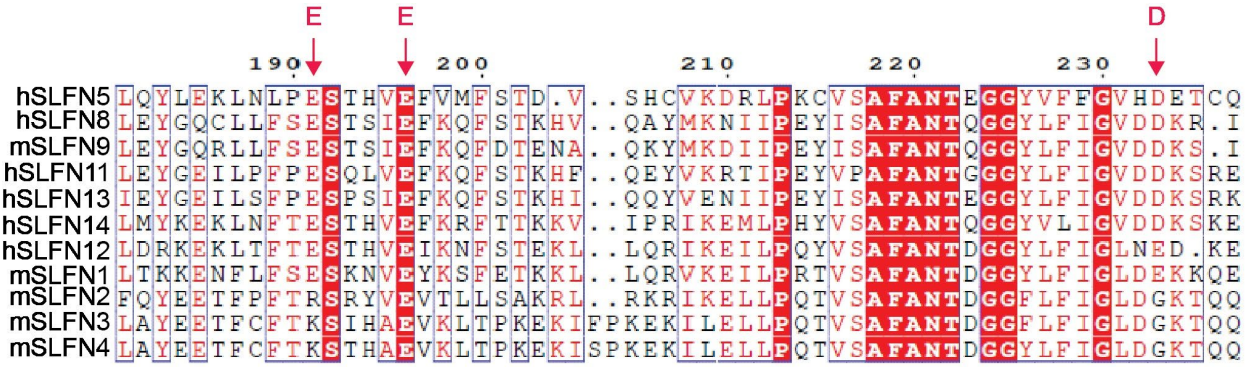

Figure I5. Alignment of the catalytic center area of the canonical mammalian Schlafen proteins. The catalytic center consists of three crucial amino acid residues, which coordinate the metal ion – E, E, and D, residues are annotated.

The first structure of the N-terminal nuclease domain was obtained for rat Schlafen13, later structures obtained for the SLFN-box of SLFN5 and SLFN11 showed high similarities to the Schlafen13 structure (Yang et al. 2018; Hou et al. 2023; Metzner et al. 2022b). Schlafen's N-terminal domain consists of two lobes, N- and C-lobes, which are forming and U-shape. The catalytic center is located in the C-lobe only, and consists of three negatively charged amino acid residues, which coordinate positively charged divalent cation (Yang et al. 2018). Up to this moment, nuclease activity has been detected for SLFN13, SLFN11, SLFN14, and SLFN12. Not all Schlafen proteins have the nuclease active center conserved (Figure I5), and not all of those which have it, show nuclease activity. Interestingly, SLFN5 has conserved the catalytic center and the general structure of the N-box, however, it failed to express RNAse activity. It is possible, that the substrate of Shlafen5 was not yet identified, or that SLFN5 needs a partner for the conformational change (Metzner et al. 2022a).

Despite the structural similarities, different Schlafen proteins have distinct substrate preferences, which still need to be characterized (Seong et al. 2017; Yang et al. 2018; Metzner et al. 2022a; Jo and Pommier 2022). It is still poorly understood how Schlafen proteins recognize their substrates, and no crystal structure of SLFN protein together with the substrate was resolved. In addition to the catalytic center, several Lysine residues were identified on the outer surface of the U-shaped N-terminal of SLFN11, which are crucial for the RNAse activity of SLFN11 and might contribute to the substrate binding (Hou et al. 2023).

1.6.4. Biological role of the Schlafen nuclease activity.

The nuclease activity of Schlafen proteins has a critical function in the immune response to viral infections and in the anti-cancer response. Even though structures of the N-terminal domains have been obtained for most of the Schlafen proteins, little is known about how exactly substrate recognition is happening and all potential substrates have been identified. Schlafen proteins act on tRNAs and rRNAs, and substrate preferences vary between different proteins.

1.6.4.1. The role of tRNA cleavage in the antiviral activity.

Schlafen proteins play a significant role in the innate immune response. Human SLFN11 and SLFN13 contain an active nuclease domain, which can cleave tRNA in reply to the viral

infection (Metzner et al. 2022b). However, the tRNA substrates and mechanisms of the antiviral activity differ between these two proteins.

Schlafen11 is capable of cleaving tRNA of all types with a preference towards tRNA with a longer v-loop (Hou et al. 2023). Besides the tRNA type II preferences, SLFN11 prioritizes tRNAs, serving codons with low GC content and A or U in position 3 (Hou et al. 2023). Many viral genomes exploit rare codon usage, characterized by low GC-level and A or U in position 3. Thus, the SLFN11 antiviral response is based on the recognition of those rare tRNAs and cleaving them to prevent the translation of the viral proteins (Kim and Weitzman 2022; Hou et al. 2023). Schlafen11 has been reported to inhibit HIV replication with the original viral codon usage in HEK293 cells, while there was almost no effect on the replication of the HIV variant with optimized codons (Li et al. 2012). Besides HIV, SLFN11 impairs the replication of DNA viruses like human cytomegalovirus HCMV (Nightingale, 2022), and positive-strand RNA Flaviviruses (Valdez et al. 2019). However, there was almost no effect on the negative strand RNA viruses (Kim and Weitzman 2022).

Schlafen13 recognizes the tRNA substrate based on the secondary structure only. Schlafen13 cleaves only tRNA type II – tRNA, containing a longer variable loop (Figure I4). This secondary structure does not correlate with the rare codon usage, thus, the antiviral functions of SLFN13 and SLFN11 do not completely overlap. SLFN13 performs a single cleavage of tRNA always in the same position - 10 nucleotides from the 3'end (Yang et al. 2018).

Interestingly, the antiviral activity of the N-terminal domains of Schlafen proteins is not restricted by tRNA cleavage. SLFN5 does not possess nuclease activity, however, SLFN5 protein is capable of tRNA binding and affects the translation of viral proteins simply by direct binding of tRNA (Metzner et al. 2022b; Metzner et al. 2022a).

1.6.4.2. The role of tRNA cleavage in the DNA damage response.

In addition to the antiviral response, tRNA cleavage by Schlafen11 is required for DNA damage response. Expression of Schlafen11 prevents the translation of ATR and ATM – serine/threonine kinases, two major components of the DNA damage response (Blackford and Jackson 2017; Cimprich and Cortez 2008; Li et al. 2018). ATM and ATR genes contain leucine and serine codons, served by tRNA type II, which is a preferable substrate for Schlafen11. The inability of expression of ATM and ATR because of SLFN11 activity causes cell death. Therefore, Schlafen11 has been studied as a biomarker in oncology to predict the outcome of the therapy with DNA

damaging agents (DDA), such as inhibitors of topoisomerase I and II, alkylating and crosslinking agents as well as poly-(ADP)-ribose polymerase inhibitors (Zhang et al. 2021). It has been shown, that cancer cells with low Schlafen11 expression levels were resistant to the DDA treatment (Murai et al. 2019b). In chemo-resistant cancer cells expression of SLFN11 is often suppressed by methylation of the SLFN11 promotor (Reinhold et al. 2017). A low expression level of SLFN11 gives more chances for cells to survive the treatment. Reversing DNA methylation with following recovery of expression level of SLFN11 had helped to return sensitivity to the chemotherapy (Gardner et al. 2017; He et al. 2017).

DNA damage response and the viral response of SLFN11 are connected. It has been shown, that some viruses exploit proteins involved in DNA damage response, like ATM and ATR, to enhance their replication (Ariumi and Trono 2006). This, function of SLFN11 in the regulation of DNA damage response can be just an extension of SLFN11's antiviral activity.

1.6.4.3. rRNA and mRNA cleavage.

In addition to tRNA cleavage, SLFN13 can process rRNA in vitro (Yang et al. 2018), however, the biological relevance of this cleavage is not exactly known.

Ribosomal RNA cleavage has been demonstrated for Schlafen14, which is capable of interacting with the ribosomes. Schlafen14 N-box is capable of cleaving different kinds of RNA, like mRNA, rRNA, and tRNA with no preferences, in Mg^{2+}/Mn^{2+} and ATP-dependent manner (Pisareva et al. 2015). However, nuclease activity was observed only for the C-terminally truncated protein. The full-length version of SLFN14 did neither demonstrate nuclease activity nor was able to bind ribosomes (Pisareva et al. 2015). As SLFN14 is expressed in a limited number of cells and its expression is upregulated as a response to viral infections, it seems possible that SLFN14 acts in line with the general immune response and can change the conformation to process RNAs by the interaction with an unknown factor. Thus, inhibition of the nuclease activity of SLFN14 by its own C-terminus may be necessary to avoid unwanted cleavage (Kim and Weitzman 2022).

Nuclease activity has been demonstrated for SLFN12. SLFN12 is capable of binding ribosomes and possesses rRNA-cleaving activity. SLFN12 has been demonstrated to form a complex with phosphodiesterase PDE3A, and complex formation increased the nuclease activity of the SLFN12 (Garvie et al. 2021). It seems, that SLFN12 might have a similar role as SLFN14. Both

proteins are activated in the RNAse L manner and may be necessary for the immune response (Chakrabarti et al. 2011; Kim and Weitzman 2022).

1.6.5. Functions of the C-terminus of SLFN proteins.

Schlafen proteins belonging to Group III have predicted helicase-like domains at their C-terminus with homology to SF1 DNA/RNA helicases, including Walker A and B motifs, suggesting that the C-terminal domain can be involved in DNA binding as well as ATP binding and hydrolysis (Geserick et al. 2004; Jo and Pommier 2022).

Helicase activity has been demonstrated for SLFN11. SLFN11 has been shown to bind to both sides of stress replication forks, open chromatin, and block replication. ATPase activity of SLFN11 is required for this process. (Murai et al. 2019b). SLFN11 selectively binds ssDNA and does not bind dsDNA. SLFN11 interacts with replication protein A (RPA), which specifically binds to single-stranded breaks, and this interaction results in the destabilizing of the RPA-ssDNA complex and the interaction of SLFN11 with DNA. ATPase activity of SLFN11 is required for chromatin opening and blocking replication fork progression. Cells, expressing SLFN11 exhibit checkpoint failure and inability to enter homologous recombination, and therefore cells with higher expression levels of SLFN11 are much more sensitive to the treatment with DNA-damaging agents. Thus, SLFN11 uses its C-terminus to irreversibly bind replication forks and block DNA damage response (Murai et al. 2019a; Murai et al. 2020). Thus, low expression levels of SLFN11 are associated with a negative prognosis of anti-cancer therapy(Murai et al. 2019a).

Despite the *in vivo* data described above, the ability of SLFN11 to bind and hydrolyze ATPs is still a discussion. Purified SLFN11 was neither binding nor hydrolyzing ATPs and the crystal structure of the C-terminus domain showed that the protein is present in the autoinhibiting state (Metzner et al. 2022b). This may be because SLFN11 needs an unknown yet interacting partner to activate helicase activity.

In addition to DNA-damage response, the putative ATPase activity of SLFN11 is also necessary for its antiviral functions. It has been shown, that mutations in the catalytic part of the Walker B domain, which is responsible for ATP hydrolysis, impair the ability of SLFN11 to reduce the replication of the prototype foamy virus (PFV) (Guo et al. 2021).

In contrast to SLFN11, the conformation of SLFN5 allows ATP binding, however, ATP hydrolysis was not observed (Metzner et al. 2022a). A similar situation is observed for SLFN14. C

terminus of the Helicase domains of both proteins are able to bind DNA, but DNA binding does not stimulate ATPase activity (Metzner et al. 2022b; Metzner et al. 2022a). SLFN11 is at the point the most studied protein in the whole SLFN family, thus, further investigation of the helicase activity of other SLFN proteins is required.

1.6.6. Functions of SLFN proteins beyond the enzymatic functions.

Schlafen proteins are upregulated in response to the interferon, therefore there were many observations of upregulation of SLFN proteins in response to various infections, for instance, in mice, expression of Schlafen proteins is upregulated during the Brucella or Listeria infection (Geserick et al. 2004). However, the exact function of SLFN proteins during bacterial infections remains unclear.

It had been reported, that mouse Slfn1 and Slfn2 negatively regulated cellular growth by repressing cycling D1 (Katsoulidis et al. 2009; Brady et al. 2005), however, these results were not always reproducible (Zhao et al. 2008).

Immune functions of SLFN proteins are not restricted by their nuclease or helicase activities. SLFN2 belongs to group I of Schlafen proteins, consisting only of SLFN-box, and the nuclease catalytic center is not conserved in this protein (Figure IX). Interestingly, despite the absence of catalytic activity, Schlafen2 plays an important role in mice's immune response. SLFN2 mutant mice exhibit the immunodeficient “Electra” phenotype, which is caused by the inability of SLFN2-deficient T cells to maintain the quiescence state (Berger et al. 2010).

SLFN2, SLFN3, and SLFN4 have been shown to regulate T cell activation and differentiation (Berger et al. 2010; Geserick et al. 2004), but further investigations of the exact mechanisms are required.

Deletion of the whole *SLFN* cluster or Om variant (ovum mutant) in mice, causes early embryonic lethality for the progeny, generated with an Om-mutant mother and wild-type father. Progeny, obtained from parents, homozygous for Om mutation, is viable. This effect has been named DDK syndrome (Bell et al. 2006).

1.6.7. Evolutional origin of SLFN proteins.

According to the evolutionary studies of the Schlafen protein family, *SLFN* genes underwent positive selection, moving towards more specialized functions (Bustos et al. 2009).

Interestingly, there were no signs of positive selection for the OPV *v-slfn* genes. Analysis of those *v-slfns* showed close homology between *SLFN1* and *SLFN2* and *v-slfn*. This observation, together with the fact that SLFN genes are found only in OPV viral family, suggests that the appearance of *v-slfn* could have happened via horizontal transferring of the *SLFN* gene from rodents to OPVs. Interestingly, *v-slfn* is present in the Monkeypox virus, which caused an outbreak in 2021-2020 (Bustos et al. 2009; Schwarz et al. 1998; Gubser et al. 2007).

Summarizing all mentioned above, it seems clear, that SLFN proteins play an important role in the immune system functioning. Despite all structural similarities, SLFN proteins do not act redundantly, however, some functions overlap between different proteins. As SLFN genes are suspected to undergo positive selection (Bustos et al. 2009), the fact that in different organisms, homologous SLFN genes show different activity, that many of those proteins are evolving towards more specialized functions. To understand the full spectra of the function of SLFN proteins, more systematic studies are required.

1.7. The mystery of the PETISCO complex in C. elegans.

As described above, identifying PID-1 led to the discovery of the novel protein complex PETISCO, which had been predicted to serve as a platform for 21U RNA precursor processing (Cordeiro Rodrigues et al. 2019). PETISCO consists of four core proteins: PID-3, TOFU-6, IFE-3, and ERH-2. PID-3 contains an RRM domain, followed by a MID domain, which theoretically can bind 5’ phosphate of RNA. Besides PID-3, MID domains are also found in the Argonaut proteins. TOFU-6 has an RRM domain, a Tudor domain, and an eIF4E binding motif, which is involved in IFE-3 binding. IFE-3 is an orthologue of eIF4E, and it has been shown to bind m7G-cap (Keiper et al. 2000). Yeast two-hybrid study of PETISCO complex revealed potential interactions between the components of PETISCO complex: ERH-2 is capable of forming a dimer and is responsible for binding PID-3 and PID-1. PID-3 is connected to TOFU-6 most probably via RRM domains, and TOFU-6 is binding IFE-3 via the eIF4E binding motif, located right after the TUDOR domain (Cordeiro Rodrigues et al. 2019).

IP of components of the PETISCO complex, coupled with mass spectrometry showed the presence of another interactor of the PETISCO complex. This protein has been named TOST-1, and the structure of TOST-1 is similar to PID-1 – a small unstructured protein, sharing the conserved area, which is potentially involved in binding ERH-2. TOST-1 and PID-1 are interacting

with PETISCO in a mutually exclusive manner and these two proteins are probably competing for their place on PETISCO.

Knockout of PETISCO protein showed two different phenotypes. Knock out of TOFU-6, IFE-3, PID-3, ERH-2 and PID-1 resulted in 21RNA deficiency. Knock out of core PETISCO proteins and TOST-1 caused maternal embryonic lethal phenotype, meaning that eggs laid by mutant mothers will not develop. Furthermore, this phenotype cannot be rescued from the paternal side. Embryos, laid by *pid-1* mutant worms mostly develop normally, while *tost-1* mutants die after several cell divisions and deeper analysis showed the formation of chromosome bridges (Cordeiro Rodrigues et al. 2019).

Thus, PETISCO is involved in both 21URNA formation and embryonic development, and its function is determined by which protein – PID-1 or TOST-1 is bound to the PETISCO complex.

1.8. Aim of the thesis.

Discovery of PETISCO raised a number of follow-up questions. How is it possible, that one protein complex is involved in two different processes? What are the exact functions of PETISCO in 21U RNA precursor processing and in embryonic development and do they have something in common? Do PID-1 and TOST-1 compete for the binding position on ERH-2, and if yes, how does the selection between these two proteins happen? At last, the most important question is, what protein is actually processing 21U RNA precursors in *C. elegans*, and how this protein is related to PETISCO? In my thesis, I will try to address these questions and identify and characterize the enzyme, which is responsible for the 5'end maturation of 21U RNA precursors.

2. Materials and methods.

Parts of the methods included in this chapter are published in the following scientific papers:

Podvalnaya, N., Bronkhorst, A.W., Lichtenberger, R. *et al.* piRNA processing by a trimeric Schlafen-domain nuclease. *Nature* **622**, 402–409 (2023). https://doi.org/10.1038/s41586-023-06588-2

Perez-Borrajero, C. et al. Structural basis of PETISCO complex assembly during piRNA biogenesis in *C. elegans*. *Genes Dev.* **35**, 1304–1323 (2021).

2.1. Protein purification.

2.1.1. Purification of ERH-2, IFE-3, TOST-1 and PID-1 proteins.

Plasmids pET28a, carrying sequences 6xHis-ERH-2 and 6xHis-IFE-3 and plasmids SV272 carrying 6xHis-MBP-TOST-1 or 6xHis-MBP-PID-1 were transformed into the BL21 (DE3) cells (Thermo Fisher); after recovering cells were transferred to 50 ml of LB Media, containing kanamycin and grown overnight at 37 °C. Afterwards, the culture was diluted from 1 to 100, in the total volume of 2L for ERH-2 and 4L for other proteins. Cultures were grown till OD 600 0,7, cooled down on the ice for 15 min, and expression of the proteins was induced by adding IPTG till 1mM. Cultures were grown at 16 °C, harvested by centrifugation at 4000g for 20 minutes, and the pellet was resuspended in the 100 ml (ERH-2) or 200 ml (IFE-3) of IMAC A buffer (50 mM TRIS-HCl, 300 mM NaCl, 10% (v/v) glycerol, 20 mM imidazole, 1 mM DTT, pH 8.0), additionally containing protease inhibitor cOmplete™ Mini (Sigma Aldrich),. Cells were opened using Cell Disruptor TS2 at 4 °C, and centrifuged at 19500 g for 30 min, supernatant was collected and loaded to the 5 ml HisTrap™ (Cytiva) column using Äkta Prime at 1 ml/min. After loading, the column was washed with 50 ml of IMAC A, elution was done with gradient washes of IMAC B (50 mM TRIS-HCl, 300 mM NaCl, 10% (v/v) glycerol, 500 mM imidazole, 1 mM DTT, pH 8.0). Collected fractions were analyzed on the SGS PAGE gel.

Fractions, containing the protein, were concentrated on Amicon® Ultra spin columns, and loaded to the gel-filtration column Superdex 75 16/60. Fractions were collected and analyzed on the SDS PAGE gel. 200 mg of each protein was used for the antibody production.

2.1.2. Purification of TOFU-6 for antibody production.

Plasmid pET28a, carrying sequences 6xHis-TOFU-6 was transformed into the BL21 (DE3) cells (Thermo Fisher); after recovering cells were transferred to 25 ml of LB Media, containing kanamycin and grown overnight at 37 °C. Afterwards, the culture was diluted from 1 to 100, in the total volume of 2L. The culture was grown till OD 600 0.8, cooled down on the ice for 15 min, and expression of the proteins was induced by adding IPTG till 1mM. Cultures were grown at 16 °C, harvested by centrifugation at 4000g for 20 minutes, and the pellet was resuspended PBS, with benzonaze 1:5000. Cells were opened using Cell Disruptor TS2 at 4 °C, and centrifuged at 19500 g for 30 min, pellet was briefly washed in cold IMAC A buffer (PBS, 8M Urea, 20 mM

Imidazole), then resuspended in 20 ml of IMAC A and brand incubated for 1h on ice, with occasional vortexing. After the incubation, the pellet was dissolved by sonication and the solution was cleared by centrifugation at 4000 g for 30 min. The supernatant was loaded to the 5 ml HisTrap™ (Cytiva) column and purification was carried on as described for ERH-2 and IFE-3. Gel-filtration step was not performed. Urea concentration was reduced to 4M by the gradient buffer exchange, while protein was bound to the column and purified protein was sent for the antibody production.

PID-3 was purified in the same way to be used for the antibody purification.

2.1.3. Purification of TOFU-6 in the native conditions.

Plasmid pET28a, carrying sequences 6xHis-TOFU-6 was transformed into the T7 Shuffle strain (NEB), after recovering cells were transferred to 50 ml of LB Media, containing kanamycin, and grown overnight at 37 °C. Afterwards, the culture was diluted from 1 to 100, in a total volume of 2 L. Expression and purification were carried on as for ERH-2, the gel-filtration step was not performed.

2.1.4. Interaction studies between IFE-3 and TOFU-6.

Interaction analysis was performed using Äkta Micro. 200 solutions of 10µM TOFU-6, IFE-3, or IFE-3 and TOFU-6 in PBS, 10% v/v glycerol were prepared and incubated on ice for 15 minutes. After brief spinning, the proteins were loaded to the Superdex 200 increased gel-filtration column. Chromatograms were analyzed, and samples from the fractions were taken and resolved on the SDS PAGE.

2.1.5. Antibody testing.

The obtained serum was purified by Martin Möckel (Protein Production core facility). To test antibodies, 200 µL aliquots of synchronized young adult worms, tagged or wild type were used. Worms were open by sonication in the 500 µL of the lysis buffer (25 mM Tris HCl pH7.5, 150 mM NaCl, 1.5 mM MgCl2, 1 mM DTT, 1 mM Triton X100, cOmplete Mini). After centrifugation, 10 µL of the worm lysate, samples, containing 15 µg of protein, were mixed with 2x gel loading buffer (2x Novex NuPage LDS sample buffer (Invitrogen), supplemented with 200mM DTT). Samples were heated at 95 °C for 10min before resolving on a 4-12% Bis-Tris NuPage NOVEX gradient gel (Invitrogen) in 1x Novex NuPAGE MOPS SDS Running Buffer (Invitrogen) at 150 V. Separated

proteins were transferred to nitrocellulose membrane (Amersham) 1h at 120V using 1x NuPAGE Transfer Buffer (Invitrogen) supplemented with 10% methanol. The membrane was incubated for 30 min in 1x PBS-Tween (0.05%) supplemented with 5% skim milk, cleaved, and incubated overnight with 5 µg of primary antibodies diluted in PBS-Tween with the addition of 1% of skim milk. Subsequently, the membrane was washed 5 times for five minutes in PBS-Tween, before incubation with the secondary antibody, using 1:10,000 goat anti-rabbit HRP-linked antibodies (#7074, Cell Signaling) and imaged using SuperSignal™ West Pico Plus (TermoFischer) kit.

2.1.8. Purification of labeled ERH-2 on the minimal media.

BL21 (DE3) cells were transformed with pET28a plasmid, containing 6xHis-ERH-2 coding sequence. Cells were grown overnight in 20 ml of LB media, washed with minimal media, and resuspended in 1 L of the minimal media (M9 salts, trace elements, 30 mg/L thiamin-HCl, 100 mM $CaCl_2$, 2 mM $MgSO_4$, pH 7.0), containing 0.5 g of $^{15}NH_4Cl$. Cells were grown till OD600 0.6, cooled down on ice for 15 min, expression was induced by adding IPTG till concentration 1 mM and carried on at 18 °C. Further purification has been performed as described above.

2.1.9. Purification of TOFU-1 and TOFU-2 from the SF9 cells.

Tofu-1 and *tofu-2* coding sequences were cloned to the pFastBac backbones, containing 6xHis tag, by Gibson assembly. Obtained plasmids were transformed to the DH10BAC competent cells (in-house), recovered for 4 hours, and seeded to the plate, containing 50 µg/ml kanamycin, 10 µg/ml tetracycline, 7 µg/ml gentamicin, 40 µg/ml IPTG, and 100 µg/ml X-gal. After 48 hours of incubation, two white clones were selected for further bacmid isolation. Purified bacmids were transfected to the SF9 cells with Cellfectin II (Termo Fischer) according to the manufacturer's protocol. V0 was collected in 60 h dpa, and 25 ml of SF9 cells with a concentration of 0.5 mln/ml were infected with the mixed V0 virus for the virus amplification. V1 was collected in 72 hours dpa, and 5 ml of the viruses per 1L of cells were used for protein production. TOFU-1 and TOFU-2 were expressed in 1L of SF9 cells, following purification had been done as described above for ERH-2, pH of the IMAC buffers was 7.2.

2.1.10. Purification of TGS1 catalytic domain.

TGS1 sequence (618–853aa) was obtained from cDNA made from HEK293 cells and cloned to the pGEX-6a-1 plasmid, containing a GST tag. Expression was done in BL21 (DE3) cells in 2 L of

LB media according to the protocol described in (Monecke et al. 2009). The GST tag was removed with TEV 1:500 (in-house). 1 µl of 2.2 µg/µl protein was used for further enzymatic reactions.

2.2. Worm-related methods.

Strain name	Genotype
RFK204	*mjSi22 [Pmex-5::mCherry::his-58::21UR-1_as::tbb-2(3'UTR)] I.*
RFK851	*mjSi22 [Pmex-5::mCherry::his-58::21UR-1_as::tbb-2(3'UTR)] I prg-1(n4357) I.*
RFK 1057	*tofu-6(xf229[tofu-6::3xMYC]) II.*
RFK1059	*tofu-2(xf231[E216A]]) V.*
RFK1095	*mjSi22 [Pmex-5::mCherry::his-58::21UR-1_as::tbb-2(3'UTR)] I, tofu-2(xf231) V.*
RFK1214	*tofu-2(xf231[tofu-2(E216A)]) V; tofu-6(xf229[tofu-6::3xMYC]) II.*
RFK1242	*pid-1(xf35) II; tofu-2(xf231) V.*
RFK1246	*mjSi22 [mex-5p::mCherry::his-58 + 21UR-1_as + tbb-2(3'UTR)] I; mut-7(xf125) III.*
RFK1269	*tofu-2(xf245[tofu-2::HA]), V.*
RFK1273	*tofu-2(xf246[E216A]::HA]), V.*
RFK1344	*pid-3(xf271[pid-3::gfp]) I.*
RFK1346	*pid-3(xf272[pid-3[A220E]::gfp]) I.*
RFK1348	*pid-3(xf273[pid-3[I182A/V186A]::gfp]) I.*
RFK1349	*pid-3(xf274[pid-3[I182E/V186E]::gfp]) I.*
RFK1442	*erh-2(xf305[ERH-2::GFP]), III.*
RFK1480	*erh-2(xf309[ERH-2[T13E]::GFP]), III.*
RFK1481	*slfl-3(xf248) I.*
RFK1506	*tofu-6(xf312[V266E]::3xMYC) I.*
RFK1580	*tofu-1(xf337[L88R;L92R]) V.*
RFK1605	*tofu-1(xf337) V; tofu-6(xf312) I.*
RFK1639	*slfl-4(xf351) IV.*
RFK1640	*slfl-3(xf248) I, slfl-4(xf351) IV.*
RFK1689	*slfl-3(xf356) I; slfl-4(xf351) IV.*
RFK1692	*tofu-1(xf358[tofu-1::3MYC]), V.*
RFK1693	*tofu-1(xf363[3MYC::tofu-1[L88R&L92R]]) V.*

Table M1. List of strains, used in the study. Strains RFK204, RFK850. RFK1246 were obtained from the database, the rest were created in this study.

2.2.1. Worm culture.

C. elegans strains were cultured on OP50 plates according to standard laboratory conditions. For IP/MS experiment worms were grown on high-density egg OP50 plates and transferred to the standard OP50 plates for the last generation. The Bristol N2 strain was used as a reference wild-type strain.

2.2.2. CRISPR/CAS9 mediated genome editing.

All protospacers were designed using CRISPR (http://crispor.tefor.net) and afterward confirmed with Integrated DNA Technologies "CRISPR-Cas9 guide RNA design checker". Protospacers were cloned to the pRK2412 by SLIM (site-directed ligase-independent mutagenesis). The Bristol N2 strain was used for microinjections unless stated otherwise. ssDNA oligonucleotides (IDT) were utilized as repair templates. Each of the repair templates has 35 nucleotides long homology arms. The injection mix contained 50ng/µl guide RNA coding plasmid for the gene of interest; 50ng/µl of plasmid, harboring CAS9 and *day-10* (cn64) or *unc-58(e665)* co-conversion guide RNA[3]; 750nM of ssDNA oligonucleotide (repair template for gene of interest) and 750nM of co-conversion ssDNA oligonucleotide. Used protospacers and repair templates are listed below.

Strain name	Allele	Genotype	Protospacer	Injected strain
RFK 1057	*xf229*	*tofu-6(xf229[tofu-6::3xMYC]) II.*	(CCG) TGAGCTCGAGCTGTAAAT	N2
RFK1059	*xf231*	*tofu-2(xf231[E216A]]) V.*	TACTTCAGTGAGCATCAGTT(TGG)	N2
RFK1269	*xf245*	*tofu-2(xf245[tofu-2::HA]), V.*	GCTACCATAGGCACCACGAG(CGG)	N2
RFK1273	*Xf246*	*tofu-2(xf246[E216A]::HA]), V.*	TACTTCAGTGAGCATCAGTT(TGG)	RFK1269
RFK1344	*xf271*	*pid-3(xf271[pid-3::gfp]) I.*	(CCT)CTTAAatgtgccccaatttg	N2
RFK1346	*xf272*	*pid-3(xf272[pid-3[A220E]::gfp]) I*	(CCC)ATATCGTGTCAATGGGATC	RFK1344
RFK1348	*xf273*	*pid-3(xf273[pid-3[I182A/V186A]::gfp]) I.*	CTTCAACAGCGAAGACATCA(AGG)	RFK1344
RFK1349	*xf274*	*pid-3(xf274[pid-3[I182E/V186E]::gfp]) I.*	CTTCAACAGCGAAGACATCA(AGG)	RFK1344

RFK1442	*xf305*	*erh-2(xf305[ERH-2::GFP]), III.*	TTGTTGGGTATTGAttacat(tgg)	N2
RFK1480	*xf309*	*erh-2(xf309[ERH-2[T13E]::GFP]), III.*	(CCA)CGTCTTGATTCACGGACTTG	RFK1442
RFK1481	*xf248*	*slfl-3(xf248) I.*	(CCG)TGTTCGAACAGGTGAAAGCC AGACGACGAGAATGATAACG(TGG) (CCT)CTTGTTAAATAGACAATCTG (CCA)GATCTCTCTCCCAATCGCTT	N2
RFK1506	*xf312*	*tofu-6(xf312[V266E]::3xMYC) I.*	ATCATTTGATAATATTCGCG(AGG)	RFK1057
RFK1580	*xf337*	*tofu-1(xf337[L88R;L92R]) V.*	TGCTCTTCTTGGAAAAGCAT (TGG)	N2
RFK1639	*xf351*	*slfl-4(xf351) IV.*	TGAGGAGTACGATCTTACTC(AGG) CAGTTTTTCAGAGTCACTAC(CGG) (CCT)GCTGTATTTATAATAACGCC TACAGATTCTATTTCACAAG(AGG)	N2
RFK1689	*xf356*	*slfl-3(xf356) I; slfl-4(xf351) IV.*	(CCT)CTTGTTAAATAGacaatctg CTTCTCGACATCCAGAACAT(CGG)	RFK1639
RFK1692	*xf358*	*tofu-1(xf358[tofu-1::3MYC]), V.*	CTCAGTTATTTTTGGAACAA(TGG)	N2
RFK1693	*xf363*	*tofu-1(xf363[3MYC::tofu-1[L88R&L92R]]) V.*	CTCAGTTATTTTTGGAACAA(TGG)	RFK1580

Table M2. List of alleles, obtained by CRISPR/CAS9 genome editing.

Allele	Repair template
xf229	
xf231	GAAAATTCAAAAACCTTTATTTTCCACTTTAAATTACTTCAGTGAGCATCAGTTTTGTATGCATGCGGAA AGCATTATGAACATCTTCTTCTACTGCAAACGGTAGTTTC
xf245	GTATTCGTCGATCTTCCTGACGAGGCCCGTGCCGAAATTGACAGTTTTTCAAAGTTTGACGGAGCTGG AGCAGGATACCCATATGATGTCCCGGATTACGCTTAATTAAGTTCGAAAAGTTTCAATAAATTTGTTGA
Xf246	GTATTCGTCGATCTTCCTGACGAGGCCCGTGCCGAAATTGACAGTTTTTCAAAGTTTGACGGAGCTGG AGCAGGATACCCATATGATGTCCCGGATTACGCTTAATTAAGTTCGAAAAGTTTCAATAAATTTGTTGA
xf271	PCR repair template with GFP and flanking 35nt of PID-3 gene
xf272	GTGGATATCCGGGGATGCTCAACACGTTCGGCATCGAGCAATTGCTCACTCCATATCGTGTCAATGGG ATCACCATCACCGGCGCCCAG

xf273	CGACGTTCGACAAGGTTCTCTTCAACAGCGAAGACGCTAAAGATTCTGCCTTCAAAGTTCTGCATGCC GAAGAAGAGCCGAGAGGTGCGGATCAGGAGAA
xf274	CGACGTTCGACAAGGTTCTCTTCAACAGCGAAGACGAGAAAGATTCTGAGTTCAAAGTTCTGCATGCC GAAGAAGAGCCGAGAGGTGCGGATCAGGAGAA
xf305	PCR repair template with GFP and flanking 35nt of ERH-2 gene
xf309	TGTCTACTTCTTCACATACCGTACTTCTGATTCAAGAGTCACCACGcCTcGAcTCgaGGACTTGGGGAGA TTATGAGAGTGTAACTGACGCTC
xf312	CTAGATGGTCTTACCTGGTCACCAGTCGCCATTCCAAGCTTTGATAATATTCGCGAAGAGGTTAAAAAA TGGGGACAAATGGAAAATTCGACGCT
xf337	CAAAATATTTTCCAGCCGAAGTGGAGGAATCACTTCGTGATGCTCTTCGTGGAAAAGCTTTAGCAGGA GATCAGATGAATTCTCGAATTGAAGGACTC
xf356	aagaaaacaaaacgaaaaatgatacagattgtCTACTGGATGTCGAGAAGATTGAGAAGCGATTGGGAGA
xf358	cgaaatcgaatattctgtctcagttatttttggaacaATGTCCGAACAAAAACTTATTTCTGAAGAGGATCTTGAGCA AAAGCTCATCTCCGAGGAGGACCTCGAGCAGAAGTTGATCAGCGAGGAAGACTTGGGATCCGGAGG TGGAGGTGCTGCGCTTTTTGAAGAGAATGATTCTTATTCGgtg
xf363	cgaaatcgaatattctgtctcagttatttttggaacaATGTCCGAACAAAAACTTATTTCTGAAGAGGATCTTGAGCA AAAGCTCATCTCCGAGGAGGACCTCGAGCAGAAGTTGATCAGCGAGGAAGACTTGGGATCCGGAGG TGGAGGTGCTGCGCTTTTTGAAGAGAATGATTCTTATTCGgtg

Table M3. List of repair templates for the strains obtained by CRISPR/CAS9 genome editing.

2.2.3. Crosses with 21U RNA sensor

RFK1059 (*tofu-2*[E216A]) and RFK1481 (*slfl-3(xf248)*) mutant hermaphrodite worms were crossed with males of the RFK1246 strain, which carries a *mut-7* deletion as well as the 21U RNA sensor (Bagijn et al. 2012). Worms, carrying 21U RNA sensor and *tofu-2[E216]* or *slfl-3(xf248)* mutation, and wild type for *mut-7* were selected by genotyping.

2.2.4. Microscopy.

Images of 21U RNA sensor-carrying strains were obtained using a Leica DM6000B. Young adults and adult worms were washed in a drop of M9 (22mM KH_2PO_4, 42mM Na_2HPO_4, 85mM NaCl, 1mM $MgSO_4$) and immobilized with 30mM sodium azide in M9 buffer. Imaging of Bm4 cells was done using Leica TCS SP5.

Microscopy Young adult worms or young gravid adults grown at 25°C were washed in M9 buffer and paralyzed with 60 mM sodium azide in M9. As soon as the worms stopped moving, they were imaged on a TCS SP5 Leica confocal microscope at 25°C. Images were processed using Fiji and Adobe Illustrator.

2.2.5. Mel phenotype scoring.

Mel phenotype For RFK1344, RFK1346, and RFK1348, 30 young adult worms were singled and left for 24 h at 20°C or overnight at 25°C. Afterward, mothers were removed and the eggs were counted, and after 2 d, developed animals were scored. RFK1349 is Mel (maternal embryonic lethal) at 20°C; therefore, the progeny of heterozygous mothers were singled and genotyped. At least 2200 eggs from at least 23 worms were counted for each condition.

2.2.6. Mass spectrometry.

2.2.6.1. Worm pellet preparation.

All IP/MS experiments were performed in quadruplicates. Worms, grown on the OP50 plates were bleached (2% NaClO, 666 mM NaOH) into high-density egg plates, grown until the gravid adult stage, and bleached again. The embryos were left to hatch in M9 buffer (22 mM KH_2PO_4, 42 mM Na_2HPO_4, 85 mM NaCl, 1 mM $MgSO_4$), L1 stage worms were seeded on standard OP50 plates and harvested at the young adult stage. Worms were washed three times with M9 buffer and one time with cold sterile water. 200µl worm aliquots were pelleted and frozen in liquid nitrogen and stored at –80°C.

2.2.6.2. Lysis preparation.

200 µl of synchronized young adult worms were thawed on ice and resuspended in 250 µl of 2x Lysis Buffer (50 mM Tris HCl pH7.5, 300 mM NaCl, 3 mM $MgCl_2$, 2 mM DTT, 2 mM Triton X100, 2x cOmplete Mini, EDTA-free) and 50 µl of sterile water. The Bioruptor Plus (Diagenode) sonicator was used to lyse worms (10 cycles 30/30 seconds, high energy, 4°C). After pelleting, the supernatant was accurately removed without the lipid phase. Finally, the protein concentration of the lysate was determined using the Pierce BCA Protein Assay Kit (ThermoFisher Scientific, 23225).

2.2.6.3. Immunoprecipitation.

For anti-HA IPs, 550 µL of worm lysate containing 0,75 mg protein was resuspended in a final volume of 550 µL of 1x Lysis Buffer. Anti-HA IPs were performed with 2 µg of in-house-made anti-HA antibodies (mouse, clone 12CA5). The lysate was incubated with the antibodies for two hours at 4 °C. For each sample, 30 µl of protein G magnetic beads (Dynabeads, Invitrogen) were washed three times in washing buffer (25 mM Tris HCl pH7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT, 1 mM Triton X100, cOmplete Mini, EDTA-free). Subsequently, equilibrated beads were added to the lysis and incubated for an additional hour at 4° C by end-over-end rotation. Finally, beads were washed 6 times with Wash Buffer, resuspended in 2xNuPAGE LDS Sample Buffer (containing 200 mM DTT), and boiled for 15 min at 95°C.

2.2.6.4. Mass Spectrometry.

Performed by Emily Nischwitz at the Falk Butter Lab

To identify TOFU-2::HA and TOFU-2::HA(E216A), samples were separated on a 4%–12% NOVEX NuPAGE gradient SDS gel (Thermo) for 10 min at 180 V in 1×MES buffer (Thermo). Proteins were fixated and stained with Coomassie G250 Brilliant Blue (Carl Roth). The gel lanes were cut, minced into pieces, and transferred to an Eppendorf tube. Gel pieces were destained with a 50% ethanol/ 50mM ammonium bicarbonate (ABC) solution. Proteins were reduced in 10mM DTT (Sigma-Aldrich) for 1h at 56°C and then alkylated with 5mM iodoacetamide (Sigma-Aldrich) for 45 min at room temperature. Proteins were digested with trypsin (Sigma) overnight at 37°C. Peptides were extracted from the gel by two incubations with 30% ABC/acetonitrile and three subsequent incubations with pure acetonitrile. The acetonitrile was subsequently evaporated in a concentrator (Eppendorf) and loaded on StageTips[5] for desalting and storage.

For mass spectrometric analysis, peptides were separated on a 20-cm self-packed column with 75 µm inner diameter filled with ReproSil-Pur 120 C18-AQ (Dr.Maisch GmbH) mounted to an EASY HPLC 1000 (Thermo Fisher) and sprayed online into a Q Exactive Plus mass spectrometer (Thermo Fisher). We used a 94-min gradient from 2 to 40% acetonitrile in 0.1% formic acid at a flow of 225nl/min. The mass spectrometer was operated with a top 10 MS/MS data-dependent acquisition scheme per MS full scan. Mass spectrometry raw data were searched using the Andromeda search engine (Cox, J. & Mann, 2011) integrated into MaxQuant suite 1.6.5.0[7] using the UniProt *C. elegans* database (August 2014; 27,814 entries). In both analyses,

carbamidomethylation at cysteine was set as a fixed modification, while methionine oxidation and protein N-acetylation were considered variable modifications. The match-between-run option was activated. Before bioinformatic analysis, reverse hits, proteins only identified by site, protein groups based on one unique peptide, and known contaminants were removed.
For further bioinformatic analysis, the LFQ values were log2-transformed and the median across the replicates was calculated. This enrichment was plotted against the – $\log_{10}$- transformed P value (Welch's t-test) using the ggplot2 package in the R environment.

2.2.7. Western blot from the worm lysis.

Worms were grown and lysed as described in the Mass spectrometry section. Lysis of both RFK1269 and RFK1280 worms, containing 15µg of protein, were mixed with 2x gel loading buffer (2x Novex NuPage LDS sample buffer (Invitrogen), supplemented with 200 mM DTT) and were heated at 95 °C for 10 min before resolving on a 4-12 % Bis-Tris NuPage NOVEX gradient gel (Invitrogen) in 1x Novex NuPAGE MOPS SDS Running Buffer (Invitrogen) at 150 V. Separated proteins were transferred to nitrocellulose membrane (Amersham) 1 h at 120 V using 1xNuPAGE Transfer Buffer (Invitrogen) supplemented with 10% methanol. The membrane was incubated for 30min in 1x PBS-Tween (0.05%) supplemented with 5% skim milk, cleaved, and incubated overnight with primary antibodies diluted in PBS-Tween (1:1,000 monoclonal anti-HA (12CA5, in-house); 1:1,000 anti-h3 (H0164, Sigma) rabbit polyclonal antibodies. Subsequently, the membrane was washed 5 times for five minutes in PBS-Tween, before incubation with the secondary antibody, using 1:10,000 horse anti-mouse HRP-linked antibody (#7076, Cell Signaling) and goat anti-rabbit HRP-linked antibodies (#7074, Cell Signaling) and imaged using SuperSignal™ West Pico Plus (TermoFischer) kit.

For strains RFK1057, RFK1506, RFK1692, and RFK1693 50 young adult worms were picked into 13 µl of M9 buffer, 5µl of 4x Novex NuPage LDS sample buffer (Invitrogen), and 2 µl of 1M DTT, boiled for 30 minutes at 95°C and loaded to the 4-12% Bis-Tris NuPage NOVEX gradient gel (Invitrogen). Gel run, transfer, staining, and imaging were performed as described above, and anti-MYC (1:1000, mouse anti-MYC (9B11), #2276S, Cell Signaling) antibodies were used.

2.2.7. RNA isolation and RNA sequencing.

Worms were grown at 20°C, synchronized by bleaching (2% NaClO, 666 mM NaOH), and were left to hatch overnight in M9 buffer. Next, L1-stage worms were seeded onto OP50 plates and harvested as young adults. For RNA extraction 500 µL of TRISOL LS (ThermoFisher Scientific, 10296-028) was added to the 50 µL worm aliquot, and five cycles of freezing in liquid nitrogen/thawing in the 37° C water bath were performed. Samples were centrifuged for 5min at 21x*g* at RT, and the supernatant was collected. 1 volume of 100% EtOH was added to the supernatant, before proceeding with the RNA extraction using the Direct-zol™ RNA MicroPrep (Zymo) kit. RNA was eluted into 13 µl of Nuclease Free water (Ambion® Invitrogen™) and each sample was divided into two aliquots for piRNA-precursor and mature piRNA library preparation.

2.2.7.1. CIP/RppH treatment and library preparation (for precursors).

CIP treatment on 1,5 µg of isolated RNA was performed in rCutSmart™ Buffer (B6004S) using 3µL of Quick CIP (M0525L) in a 40 µL reaction. The reaction was incubated at 37 °C for 20 minutes, followed by heat inactivation for 2 min at 80 °C. The CIP-treated RNA was subjected to another round of purification using the Direct-zol™ RNA MicroPrep (Zymo) kit. RppH (NEB) treatment was performed with a starting amount of 500 ng.

2.2.7.2. Library preparation and sequencing.

Performed by the IMB Genomics Core facility and Emil Karaulanov

NGS library prep was performed with NEXTflex Small RNA-Seq Kit V3 following Step A to Step G of Bioo Scientific`s standard protocol. Amplified libraries were purified by running an 8% TBE gel and size-selected for 15-40 nt. Libraries were profiled in a High Sensitivity DNA Chip on a 2100 Bioanalyzer (Agilent Technologies), quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 2.0 Fluorometer (Life Technologies), and sequenced on Illumina NextSeq 500/550.

The raw sequence reads in FastQ format were cleaned from adapter sequences and size-selected for 18-35 nt inserts (plus 8 random adapter bases) using cutadapt v.4.0 (http://cutadapt.readthedocs.org) with parameters "-a TGGAATTCTCGGGTGCCAAGG -m 26 -M 43". Data quality was assessed with FastQC v.0.11.9 (https://github.com/s-andrews/FastQC) and MultiQC v.1.9 (https://multiqc.info/). Read alignment to the *C. elegans* genome (Ensembl WBcel235/ce11 assembly) with concomitant trimming of the 8 random bases was performed using Bowtie v.1.3.1 (http://bowtie-bio.sourceforge.net) with parameters "-v 1 -M 1 -y --best --

strata --trim5 4 --trim3 4 -S" and the SAM alignment files were converted into sorted BAM files using Samtools v.1.10 (http://www.htslib.org). *C. elegans* WBcel235/ce11 gene annotation in GTF format was downloaded from Ensembl release 96 (ftp://ftp.ensembl.org/pub/). Aligned reads were assigned to small RNA loci and classes using Samtools, GNU Awk, and Subread featureCounts v.1.6.2 (http://bioinf.wehi.edu.au/featureCounts/). Structural reads aligned in sense orientation to rRNA, tRNA, snRNA, and snoRNA loci were excluded from further analysis. Mature 21U-RNAs were stringently defined as reads of length 21 nt starting with T and fully overlapping with annotated piRNA (21ur) genes in sense orientation. Because 21ur gene annotation corresponds to mature piRNA sequences, 21U precursors were stringently defined as reads of length 23-35 nt starting 2 bp upstream of the annotated 5' end of (mature) piRNAs in sense orientation. The relative abundance of mature and precursor 21U-RNAs was normalized to the number of non-structural 18-35 nt reads in each sample. Coverage tracks of aligned reads overlapping in sense with piRNA genes were produced using Bedtools v.2.27.1 (http://bedtools.readthedocs.io) and kentUtils v.385 (https://github.com/ucscGenomeBrowser/kent). The tracks were normalized based on all non-structural reads in each sample and visualized on the IGV genome browser v.2.15.4 (https://igv.org/).

2.2.8. RNA preparation of the RNA-MS.

RFK1059 and RFK1214 were synchronized by bleaching and seeded to the OP50 plates. Young adults were harvested in three biological replicates in the 250 µL aliquots, each aliquot was split into two – 50 µL for total RNA extraction (described above) and 200 µL for the IP with anti-myc antibodies (1:1000, mouse anti-MYC (9B11), #2276S, Cell Signaling). IP was performed as described in section (2.2.7). The beads, carrying the immune-precipitated proteins, were used for RNA isolation by Zymo Direct-zol RNA Miniprep Kit, according to the manufacturer's protocol.

Ribonucleosides (A, C, m^6A) standards, ammonium acetate, and LC/MS grade acetonitrile were purchased from Sigma-Aldrich. (Am, Im, m1A, Gm, Cm, Um) from Trilink, (m1I, m22G,m7G,m1G) from Carbosynth, 13C10-A, 13C9-C was purchased from Silantes, GmbH (Munich, Germany). 2H3- m6A was obtained from TRC, Inc. (Toronto, Canada). All solutions were prepared using ultrapure water (Barnstead GenPure xCAD Plus, Thermo Scientific). A 1–2 µg of RNA was degraded to nucleosides with 0.003 U nuclease P1 (Roche), 0.01 U snake venom

phosphodiesterase (Worthington), and 0.1 U alkaline phosphatase (Fermentas). Separation of the nucleosides from the digested RNA samples was performed with an Agilent 1290 UHPLC system equipped with ReproSil 100 C18 column (Jasco) (15cm) with a gradient of 5 mM ammonium acetate (pH 7, solvent A) and acetonitrile (solvent B). Separations started at a flow rate of 0.5 ml/min. Then, washing and re-conditioning were done at 1 ml/min from 24 min until 34, and linearly decreased to 0.5 ml/min during the last minute. The gradients were as follows: 100% solvent A for the first 8 min. then B linear increase from 0 to 15% for the next 16 min, followed by isocratic elution at 15% solvent B for another 1 min; then gradient increase from 15% to 60 % solvent B for another 4 min. Then switch to 0% solvent B for the last 6 min, to recondition the column. Quantitative MS/MS analysis was performed with an Agilent 6490 triple quadrupole mass spectrometer in positive ion mode. Details of the method and instrument settings are described elsewhere (PMID: 26751644).

2.2.9. Conservation Heat map for the PUCH and PETISCO complexes.

Perfomed by Peter Sarkies.

The proteome of *C. elegans* WS235 was used as the test file for the reciprocal blastp searches against all other species, listed in the heatmap, recording only the best hits. The heat map was generated in R by using a heatmap.2 function.

2.3. *In vitro* assays.

Preparation of the IP-PUCH and mini-PUCH proteins was performed by Walter Bronkhorst (IP-PUCH) and Sebastian Falk lab (mini-PUCH) and described in Podvalnaya, N., 2023.

Precursor	Sequence	Labeling
m^7G-AAU precursor	m7G-rArArUrUrCrCrArCrUrGrUrUrUrUrArGrUrUrUrGrUrCrUrGrUrUrUrArA	3'end
m^7G-AAC precursor	m7G-rCrArUrUrCrCrArCrUrGrUrUrUrUrArGrUrUrUrGrUrCrUrGrUrUrUrArA	3'end
m^7G-CAU pecursor	m7G-r**C**rArUrUrCrCrArCrUrGrUrUrUrUrArGrUrUrUrGrUrCrUrGrUrUrUrArA	3'end

TMG-AAU pecursor	TMG- rArArUrUrCrCrArCrUrGrUrUrUrUrArGrUrUrUrGrUrCrUrGrUrUrUrArA	3'end
5'P precursor	P-rArArUrUrCrCrArCrUrGrUrUrUrUrArGrUrUrUrGrUrCrUrGrUrUrUrArA	5'end
5'OH precursor	OH-rArArUrUrCrCrArCrUrGrUrUrUrUrArGrUrUrUrGrUrCrUrGrUrUrUrArA	3'end

Table MX. The sequence of synthetic RNA precursor used in the studies. m^7G-AAU, m^7G-AAC, m^7G-CAU and TMG-AAU were synthesized by Bio-synthesis, the rest is produced by IDT.

2.3.1. 3' RNA radioactive labeling.

3'-end labeling of substrate RNA was performed in a 13µL reaction containing 2.5µL f DMSO, 2.5µL of T4 ligase buffer (NEB), 1µL of T4 ligase (NEB), 2.5µL 10mM ATP (NEB), 1µL of synthetic RNA precursor (5pmol/µL). The reaction was mixed and 2.5µL of [5'-^{32}P]pCp (SCP-111, Hartmann analytic) was added before overnight incubation at 16°C. Finally, the labeled RNA was purified using G25 columns (Cytiva) according to the manufacturer's protocol. The 3'-end labeled synthetic RNA precursor was used for *in vitro* cleavage assays and EMSAs.

2.3.2. 5' RNA radioactive labeling.

5 pmol synthetic RNA oligonucleotide (OH-precursor) was labeled with ATP, [γ-^{32}P] (PerkinElmer) using T4 PNK (NEB), according to the manufacturer's protocol.

2.3.3. *In vitro* cleavage assay.

The PUCH complex used for *in vitro* cleavage assays was obtained in two different ways. The full-length PUCH complex was obtained from GFP-IPs using BmN4 cell lysates (see above), whereas the minimal catalytic complex (mini-PUCH) was purified from *E. coli.*

For the *in vitro* cleavage assays performed with IP material from BmN4 cells, beads were washed in the cleavage buffer (CB) containing 40mM Tris-HCl, pH 8.0, 20 mM KCl, 11 mM $MgCl_2$, and 2 mM DTT. Beads were subsequently resuspended in 10µL of CB and incubated with 0.2pmol of the labeled RNA substrate for 1h at room temperature.

For cleavage assays with mini-PUCH purified from *E.coli* 0.2pmol of labeled RNA substrate was incubated in 10µL CB buffer with 27nM mini-PUCH protein complex (final concentration) at 20°C for 30 min.

The cleavage reaction was terminated by adding 1µL of 20mg/ml proteinase K. 1 volume of the 2xRNA Gel Loading Dye (Thermo Scientific™, R0641) was added and the RNA was resolved on a 15% TBE-UREA gel (Novex™) for 90min at 180V with 1xTBE as the running buffer.

2.3.4. Substrate specificity test of PUCH complex.

Capped RNA oligonucleotides were labeled at the 3' end, 0.2 pmol (1µL) of RNA per sample, was used in the cleavage reaction. For reaction with IP material, to obtain 5'P-containing 21U RNA precursor oligonucleotide, 5' OH-21U RNA precursor had been labeled on the 3'end as described above. After labeling, 5'P was created by T4 PNK treatment (NEB, M0201S), done according to the NEB T4 PNK protocol. For the reaction with mini-PUCH 5'OH-21U RNA precursor oligonucleotide had been labeled on the 5'end as described above.

2.3.5. Analysis of divalent cations as cofactor of PUCH complex.

For the metal assay, beads were washed with CB, but 100 mM $MgCl_2$ was replaced by $ZnCl_2$, $MnCl_2$ or $CaCl_2$. Cleavage reaction was done with full-length PUCH obtained from GFP-IP material from BmN4 cell lysates.

2.3.6. Ligation of small RNA oligo to the cleavage product to prove the formation of 5'P on the cleaved RNA precursor.

2pmol of labeled RNA was incubated in 35µL of CB containing mini-PUCH (or mutated mini-PUCH) at a final concentration of 40nM and was incubated at 20°C for 1h. Afterward, 3 volumes of Trisol LS reagent (ThermoFisher Scientific, 10296-028) were added, and RNA was purified using Direct-zol™ RNA MicroPrep (Zymo) according to the manufacturer's protocol. Next, the RNA was ligated to 10pmol of 5'OH-rGrUrCrUrGrUrUrUrArA-OH3' oligonucleotide using T4 RNA ligase according to the manufacturer's protocol. After 16h of incubation at 16°C, the reaction terminated by proteinase K and RNA was resolved on a 15% TBE-UREA gel (Novex™) for 90 min at 180 V with 1xTBE as the running buffer.

2.3.7. PUCH complex cleavage activity in the presence of PETISCO.

The assay has been done with the PUCH complex, obtained from Bm4 cells. Per sample: 1µL of 3'end labeled 21U RNA precursor (0.2pmol) was incubated with five times access of PETISCO protein complex on ice for 1h in 10µL of CB buffer. PUCH-IP-containing beads were resuspended in 10µL of RNA-PETISCO mix and incubated at 20°C.

2.3.8. Comparison of m7G-CAU and m7G-AAU substrate processing.

18µL of 0.2pmol/µl RNA substrate (AAU or CAU) were added to the 162µl of CB buffer, containing recombinant mini-PUCH at the final concentration of 27 nM. Samples were transferred to 20 °C and samples for each time point were taken. The reaction was stopped by adding proteinase K. Images were processed with Fiji.

2.3.9. Experiment with cold RNA.

To the 2µL mix of 0.2 pmol of labeled AAU substrate and 0.4 pmol of cold RNA of choice 10µl of CB buffer with 27 nM mini-PUCH were added. The cleavage reaction was incubated at 20 °C for 15 minutes and then stopped by adding protease K.

2.3.10. PUCH complex cleavage activity in the presence of PETISCO.

16 µL of 3'end labeled piRNA precursor (0.2 pmol/µl) were incubated with five times molar access of PETISCO protein complex on ice for 1 h in 160 µL of CB buffer. After the incubation, PUCH-IP-containing beads were added, and the samples were split into two tubes. The same procedure had been performed in parallel for RNA incubated without PETISCO. Reactions were incubated at 20 °C with mild shaking, and 10 µL samples were taken for each time point. The same experiment had been performed with recombinant mini-PUCH at the concentrations described for cleavage reactions.

2.3.11. Modification of TMG.

Two reactions of 10 µl of 3’ labeled m^7G-AAU RNA precursor (0.2 pmol/µl) were incubated with 0.5µM of $TGS1_{618–853}$ in the presence and absence of 25 pM of SAM (Sigma Aldrich) at 37 °C for 30 minutes in total volume of 20 µl. The reaction was stopped by adding protease K and purified using a Zymo mini RNA purification kit. 2 µl of each reaction was used for further cleavage reactions.

2.3.12. EMSA.

0.2 pmol of capped 21U RNA precursor m^7G-AAU, 5'P 21U RNA precursor, and 5'OH-RNA precursor were incubated with recombinant proteins of PETISCO complex, containing IFE-3, TOFU-6, ERH-2 and PID-3 (Perez-Borrajero, 2021) in a concentration range from 75 pM to 1.44 µM, in 10 µL of binding buffer (20 mM HEPES pH 7.5, 150 mM NaCl) for 1h at the room temperature. After the incubation, each sample was mixed with 15% Ficoll with bromophenol

blue. Native 6% TBE gel was pre-run for 30 min at 180 V at room temperature in 1xTBE, and samples were resolved for 2 h.

3. Results I. Investigation of the structure of PETISCO.

Parts of the figures included in this chapter are published in the following scientific paper:

Perez-Borrajero, C. et al. Structural basis of PETISCO complex assembly during piRNA biogenesis in *C. elegans*. *Genes Dev.* **35**, 1304–1323 (2021).

In the first part of my thesis, I have focused on the structure of PETISCO complex and its function in 21U RNA formation (Cordeiro Rodrigues et al. 2019; Zeng et al. 2019). The PETISCO complex consists of four proteins, ERH-2, TOFU-6, PID-3, and IFE-3 (Figure R1.A). In addition, PETISCO interacts with two proteins, TOST-1 and PID-1, which determine the function of the complex (Figure R1.A and B). In combination with TOST-1, PETISCO has a role that is crucial for embryonic development and unrelated to 21U RNAs, while interaction with PID-1 directs the complex toward the production of 21U RNAs in *C. elegans*, and in this form, it does not affect embryonic viability. This implies that PETISCO mutants show two phenotypes – 21U RNA deficiency and maternal embryonic lethality (Mel). In contrast, loss of TOST-1 only causes a Mel phenotype, but does not affect 21U RNA levels (Cordeiro Rodrigues et al. 2019). Worm strains lacking PID-1 are viable but have reduced amounts of 21U RNAs (Albuquerque et al. 2014; Cordeiro Rodrigues et al. 2019). Thus, the aim of the first part of the project is to understand how PETISCO can combine these two functions. To answer this question, structural studies of PETISCO complex were performed.

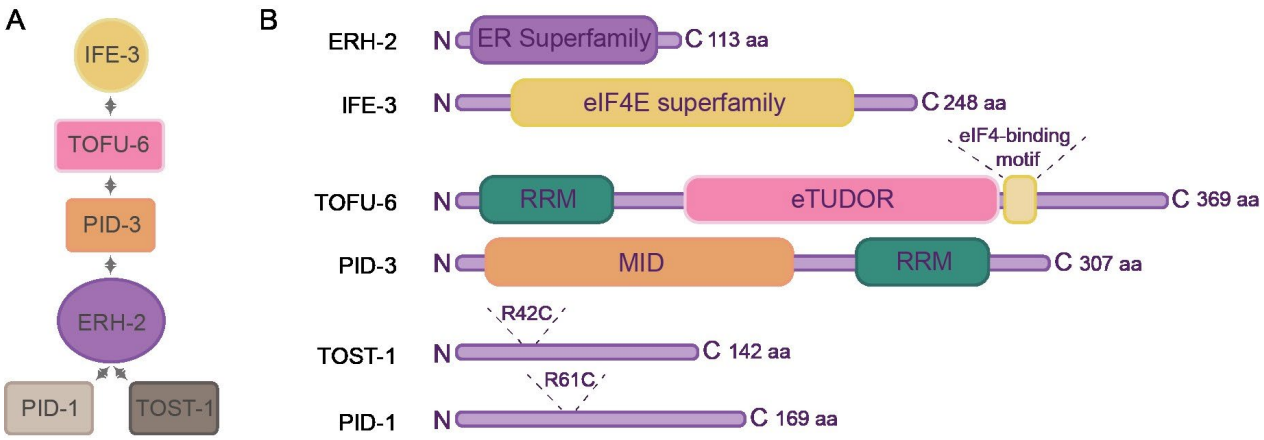

Figure R1. **Schematic representation of domain organization PETISCO and its interactors TOST-1 and PID-1**. **A**., Assumption of the PETISCO organization, based on the Y2H data (Rodrigues, 2019). **B**., Schematic representation of domain organization of ERH-2, IFE-3, TOFU-6, PID-3, PID-1, and TOST-1. The lines indicate low-complexity regions and the rectangles indicate the predicted folded domains. R42C in TOST-1 and R61C in PID-1 represent mutations, leading to the disruption of the interaction with ERH-2 (Rodrigues, 2019).

3.1. ERH-2, PID-3, IFE-3, and TOFU-6 can be purified from the *E. coli*.

The first goal of the project was to purify components of PETISCO complex from bacterial cells for various applications. First, purified proteins could be used for antibody production of PETISCO complex components. Good antibodies would allow to perform IP-MS experiments, Western blots, and immunostainings. Second, if purifications in the native conditions are successful, purified proteins could be used in the structural studies, and resolving the structure of PETISCO would help to understand the dual nature of PETISCO functions. Besides that, we envisaged that we may need PETISCO to perform the *in vitro* cleavage reactions of 21U RNA precursors, once we have identified the 21U RNA-processing nuclease (see part 2 of the Results section).

3.1.1. Purification of ERH-2 from *E. coli*.

ERH-2 belongs to the Enhancer of Rudimentary Homolog family of proteins, which is widely conserved in eukaryotes (Weng and Luo 2013). ERH proteins have been shown to take part in exosome-mediated RNA degradation (Sugiyama et al. 2016) and help microprocessor complex formation through the assistance of the dimerization of the DGCR8 protein in human cells (Kwon et al. 2020).

There are two ERH proteins present in *C. elegans*, ERH-1 and ERH-2, but only ERH-2 is a component of PETISCO. ERH-2 is a small well-structured protein with the size of 13 kDa, predicted to form a dimer. *Erh-2* was amplified from worm cDNA, and cloned into the pET28a plasmid, encoding N-terminal 6xHis tag. Protein expression was performed in BL21 (DE) cells. Purification was done in two steps: first 6xHis-tagged ERH-2 was isolated from the lysis by IMAC (Immobilized metal affinity chromatography) (Figure R2.A and R2.B), and the rest of the impurities were removed by size exclusion chromatography (SEC) (Figure R2.A and R2.C), all in native conditions. 6xHis-ERH-2 protein was eluted from the SEC column with the volume, corresponding to the 28 kDa, suggesting, that ERH-2 forms a dimer, as it was shown for its homologs in the different

organisms.

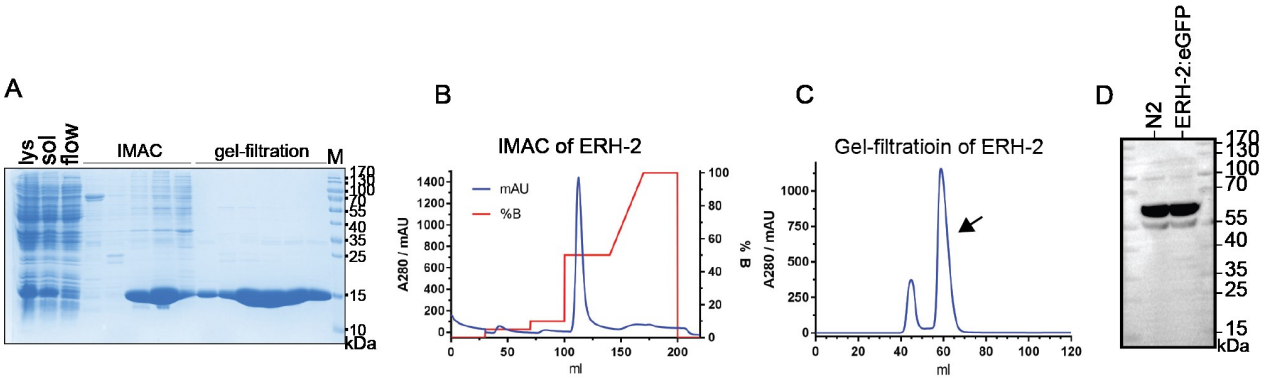

Figure R2. Purification of ERH-2. A., SDS PAGE, contains samples from both IMAC and SEC ERH-2 purification steps. **B**., Chromatogram of IMAC purification step of ERH-2. ERH-2 eluted as one peak, which was concentrated and used for the SEC. The right axis shows the % of the elution buffer (% B), containing imidazole in the concentration 500mM. **C**., Chromatogram of SEC ERH-2 purification step. The first peak corresponds to the aggregated proteins; samples from the second peak (marked with an error) were loaded into the gel (shown in **A**) and then used for antibody production. **D**., Western blot for wild type and ERH-2:eGFP worms with anti-ERH-2 antibodies, obtained from two different rabbits (one of the two similar Western blots is shown).

Purified protein was sent for antibody production to the Speedy program in Eurogentec; obtained sera were purified with the help of the IMB core facility and their specificity was tested on the Western blot. To test the specificity of the antibodies wild-type worms and worms, carrying a deletion of ERH-2, rescued by ERH-2:eGFP transgene were used. Unfortunately, there was no ERH-2-specific band detected on the Western blot (Figure R2.D). These sera were not used any further.

3.1.2. Purification of IFE-3 from E.coli.

There are five eIF4E proteins encoded in the *C. elegans* genome and IFE-3 is one of them. IFE-3 is a 28 kDa protein which contains a single eIF4E domain. Most of the mRNA in *C. elegans* undergo trans-splicing and carry 2,2,7-trimethylguanosine (TMG) cap. 21U RNA precursors are transcribed by RNA Pol II and not trans-spliced and therefore carry an m^7G cap. IFE-3 has been shown to interact exclusively with 5' m^7G caps (Keiper et al. 2000), suggesting it may interact with the 5' end of 21U RNAs. IFE-3 interacts with TOFU-6 in the Y2H assays, most probably via an eIF4E-interacting motif present in TOFU-6 (Cordeiro Rodrigues et al. 2019). Besides PETISCO, IFE-3

interacts with components of the SMN complex and several Gemin proteins (Cordeiro Rodrigues et al. 2019).

IFE-3 protein has three isoforms and sequences encoding them were detected in the cDNA, obtained from wild-type worms. I chose to work with the isoform b, as it is the longest one. *Ife-3b* sequence was cloned into the pET28a plasmid, 6xHis-IFE-3 protein was expressed in BL21 (DE3) cells and then purified by IMAC (Figure R3.A and R3.B) and SEC (Figure R3.C and R3.D) in native conditions. A relatively big amount of protein aggregated and eluted from the SEC column too early for IFE-3's molecular weight. A small portion of the protein was eluted at the volume, corresponding to 43 kDa size, and this fraction was used in further experiments. As for ERH-2, two rabbits were immunized for antibody production; however, antibody isolation was successful only for one of the sera. Antibody specificity was tested by Western blot using wild-type worms, and worms with IFE-3 knockout, rescued by insertion of 3xFLAG:mCherry:IFE-3 encoding transgene. On Western blot, clear bands corresponding to both IFE-3 and 3FLAG:mCherry:IFE-3 were observed (Figure R3.E). Thus, the generated anti-IFE-3 antibodies were good, however, I did not use them in further experiments.

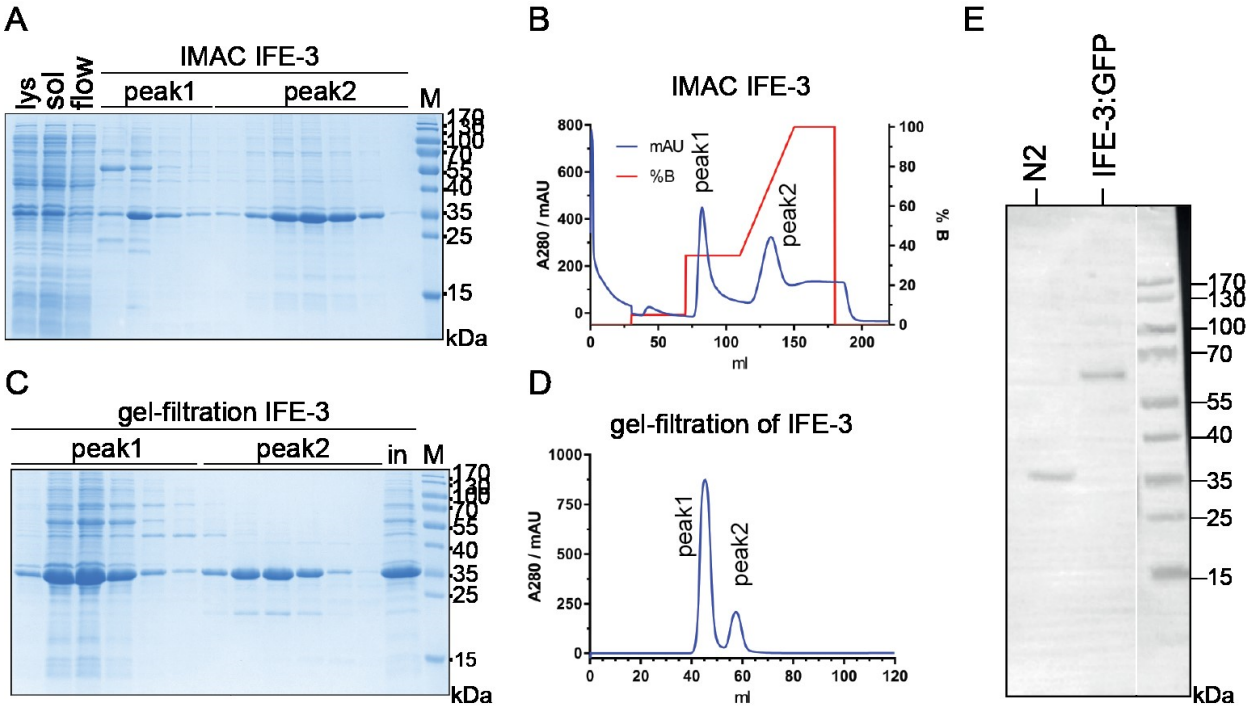

Figure R3. **Purification of IFE-3. A.** SDS PAGE of the IMAC purification step of 6xHis-IFE-3. Fractions containing the second peak were concentrated and used in the SEC purification step. **B.** Chromatogram of the IMAC purification of IFE-3. **C**. SDS PAGE of the SEC purification step of IFE-3. **D**. Chromatogram of the SEC purification step of IFE-3. Protein, eluted in a second peak was used for antibody production. **E.** Western blot for wild-type worms and worms carrying 3xFLAG:mCherry:IFE-3 transgene with anti-IFE-3 antibodies.

3.1.3. Purification of TOFU-6 from *E. coli.*

TOFU-6 is a 41 kDa protein, specific for the nematodes. It contains an RRM domain and an extended Tudor domain (eTudor), followed by the small stretch that forms a conserved eIF4E (IFE-3) interaction motif (Figure R1.A.).

Initial expression tests had shown that TOFU-6 is insoluble, and therefore in addition to the pET28a I cloned TOFU-6 expressing gene into plasmids pGex-6a-1, containing GST-tag, and plasmid SV272, with MBP tag, hoping that it will improve the solubility of the protein, however, it did not solve the problem. Expression in the Rosetta (DE3) pLysS strain and Arctic BL21 (DE3) strain at 12 °C also did not result in soluble protein (data not shown). Therefore, IMAC purification of 6xHis-TOFU-6 from inclusion bodies under denaturing conditions using 8M urea was performed (Figure R4.A). As the purified protein did not contain too many impurities, further purification steps were not required. The concentration of urea was reduced to 4M and 6xHis-TOFU-6 protein was sent for antibody production.

It was not clear if the obtained antibodies would work, therefore tagging of the *tofu-6* gene with the *3xmyc* sequence was performed. The tag was added tag on the C-terminus of the endogenous loci of the *tofu-6* gene by CRISPR/CAS9 genome editing.

As for IFE-3, we were able to purify antibodies only from one of the rabbit sera. Specificity test was performed and tested using the wild-type worms and a newly generated TOFU-6:3xMYC expressing strain. Clear bands corresponding to both wild-type TOFU-6 and TOFU-6:3xMYC were observed (Figure R4.B). Thus, TOFU-6 antibodies showed great specificity and a low level of background binding and can be used in further experiments.

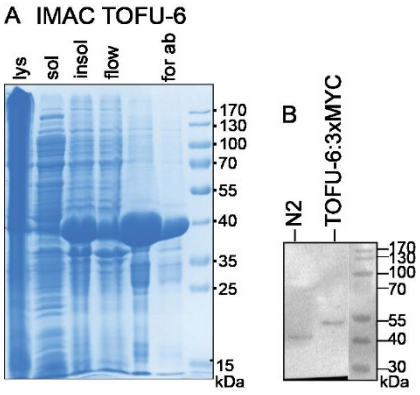

Figure R4. **Purification of TOFU-6 in the denaturing conditions. A.** IMAC purification of TOFU-6 from inclusions bodies. **B.**

Western blot for wild-type worms and worms expressing TOFU-6:3xMYC protein with anti-TOFU-6 antibodies.

3.1.4. IFE-3 interacts with TOFU-6 in a 1-to-1 ratio.

As TOFU-6 was planned to be used for the interaction studies with the other PETISCO proteins, optimization of TOFU-6 purifications was required. Several conditions were tested, including purification from SF9 cells. The best results were obtained from the expression of 6xHis-TOFU-6 protein at 16 °C in the SHuffle T7 strain and further on bench purification (Figure R5.A). Thus, soluble TOFU-6 protein was obtained and used in the interaction studies with IFE-3 proteins (purification shown above). TOFU-6 was incubated together with IFE-3 and loaded into the analytical gel filtration Superose® 6 Increase column to test the interaction between TOFU-6 and IFE-3. TOFU-6 and IFE-3 eluted in one peak (Figure R5.B). Elution peaks were analyzed by SDS-PAGE and the presence of both TOFU-6 and IFE-3 was detected in the shifted peak (Figure R5.C).

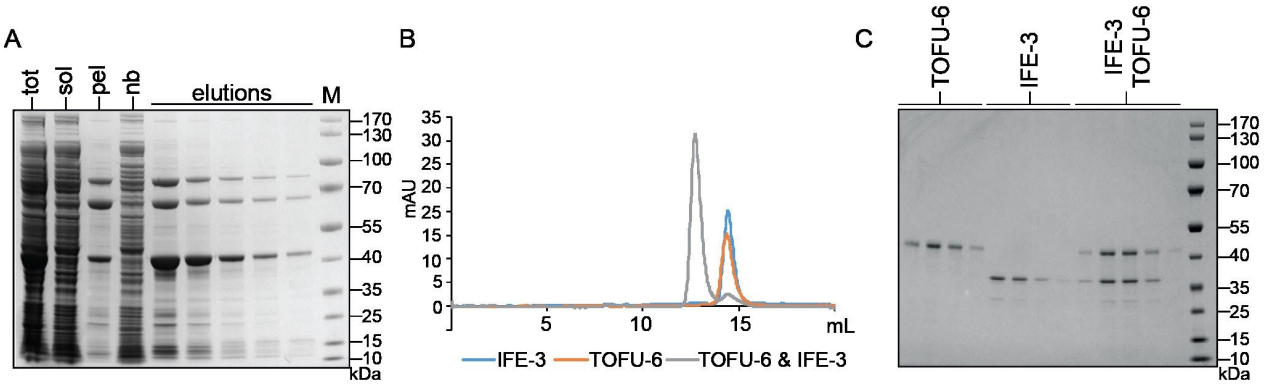

Figure R5. TOFU-6 interacts with IFE-3. A. SDS PAGE of on-bench purification of 6xHis-TOFU-6. **B. Chromatogram** of the SEC for IFE-3 (blue), TOFU-6 (orange), and TOFU-6 and IFE-3 incubated together (grey). **C.** SDS PAGE of the samples taken from the peak fractions shown in **B**.

The complex was eluting at the size, corresponding to 107 kDa. The sum of the masses of TOFU-6 and IFE-3 together would be 73.2 kDa. Despite the possibility of the presence of two molecules of IFE-3, in the complex, I assume that proteins are interacting in a 1-to-1 ratio. SEC predicted molecular mass of the complex depends on the shape of the protein, and two interacting globular proteins might appear bigger than they are. The peak of IFE-3 alone corresponded to 42 kDa, while the molecular mass of IFE-3 is only 28.2 kDa. Besides that, TOFU-6 and IFE-3 were mixed in a 1-to-1 molar ratio and most of the loaded protein appeared in the

shifted peak suggesting complete interaction. More precise analyses by SEC coupled to multiangle light scattering (MALS) confirmed the 1-to-1 interaction ratio (Perez-Borrajero et al. 2021).

3.1.5. Anti-PID-3 antibodies can be used for Western blots.

PID-3 is another PETISCO subunit. It contains two domains: an RRM domain and a MID domain. MID domains are also present in Argonaute proteins and it is assumed to be able to bind 5'P-ends of RNA (MacRae et al. 2008). Previously in the lab, PID-3 was purified under denaturing conditions and this protein was used for antibody production. These antibodies were, however, not yet tested.

I repeated the purification of PID-3 in order to obtain protein for antibody purification, performed by the Protein Production core facility. Purified PID-3 antibodies were tested on the Western blot. While I was able to see bands corresponding to PID-3 and PID-3:eGFP with sera from both rabbits, the serum, obtained from rabbit 884 showed many non-specific bands. The serum of rabbit 885 was much cleaner (Figure R1.L). These antibodies I also have not used in further experiments, but they have been used in the other projects in the lab.

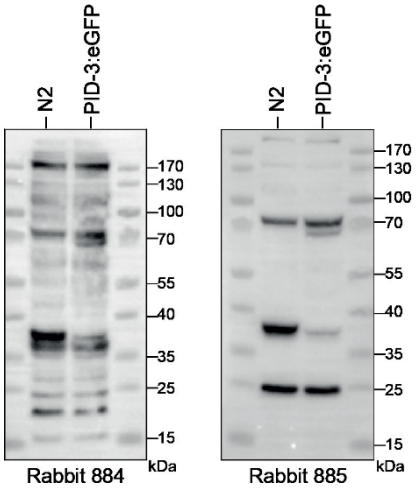

Figure R6. Anti-PID-3 antibodies work on the Western blot. Western blot for wild-type worms and worms expressing PID-3:eGFP transgene with anti-PID-3 antibodies.

3.1.6. Purification of PID-1 and TOST-1 from *E. coli.*

TOST-1 and PID-1 are both small and disordered proteins. PID-1 is predicted to be more unstructured, than TOST-1. Both proteins share a short common stretch, which was shown before to mediate interaction with ERH-2: Mutations R42C in TOST-1 and R61C in PID-1 prevent interaction of these proteins with ERH-2 in Yeast-2-hybrid (Y2H) assays (Cordeiro Rodrigues et al.

2019). Purification of TOST-1 and PID-1 was similarly unsuccessful. I have started with the purification of His-TOST-1, expressed from pET28a plasmid in BL21 (DE) cells. While after IMAC elution was happening in the one good peak, most of the TOST-1 protein did not survive the storage at -80 and degraded (data not shown). IMAC purification was repeated (Figure R7.C) and the gel-filtration step was performed in order to remove the contaminants, which could potentially cause degradation of TOST-1. However, purification of TOST-1 with gel-filtration was not possible either due to the relatively disordered structure of the protein or due to its aggregation (Figure R7.D and E). Purification of His-MBP tagged TOST- was also unsuccessful. Degradation of the protein was observed already on the IMAC step; after cleavage of MBP and gel-filtration, most of TOST-1 was degraded. A similar situation was observed with PID-1, but degradation of the PID-1 was much more intense.

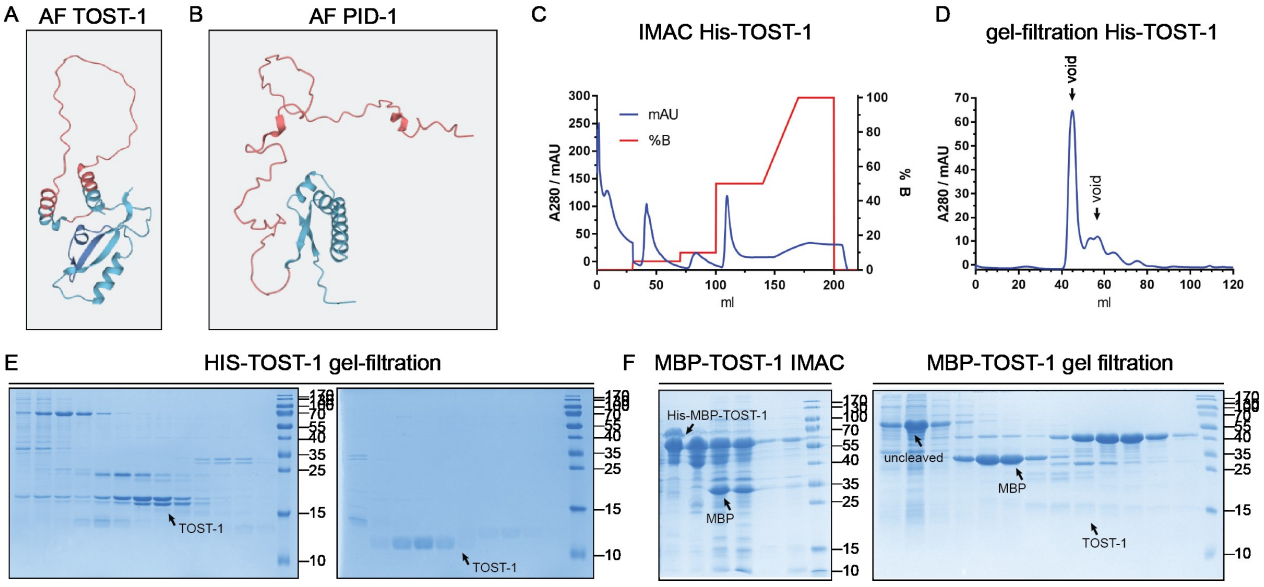

Figure R7. Purification of TOST-1 from *E. coli.* A-B. Alphafold prediction of TOST-1 (**A**) and PID-1 (**B**) structures. Color code – from blue to red, where red is disordered and blue organized. **C**. Chromatogram for the IMAC purification step of His-TOST-1. The right axis is showing the % of the elution buffer (B), containing 500mM of Imidazole. **D**. Chromatogram of the gel-filtration of His-TOST-1. E. SDS page of the gel-filtration step of His-TOST-1 purification, shown in **D**. Samples from all of the fractions were loaded to the gel from left to right. **F**. Purification of His-MBP-TOST-1: IMAC (on the left) and SEC (on the right).

3.2. Structural analysis of PETISCO complex.

As purification of most PETISCO proteins was possible from *E.coli*, further structural studies were planned. At this point, a collaboration was started with the laboratories of Sebastian

Falk (Max Perutz, Vienna) and Janosch Hennig (EMBL, Heidelberg). Analysis of PID-3, TOFU-6, IFE-3, and ERH-2 interactions as well as the crystallization of their sub-complexes were performed with proteins purified by the Falk laboratory by the Falk laboratory. Protein NMR was performed by the Hennig laboratory. I will focus on the results I obtained myself, and will only briefly mention the outcome of the purification and structural studies where needed. The complete data are published in Perez-Borrajero, 2021.

3.2.1. Stoichiometry of PETISCO complex.

For investigation of PETISCO complex stoichiometry, PETISCO proteins were purified from bacterial cells and analyzed by gel-filtration, coupled to multiangle light scattering (MALS). This approach allows more precise measurements of the molecular mass. PETISCO formed a single peak during the gel-filtration with 231kDa size. This corresponds to the double mass of all components, suggesting that PETISCO complex forms an octamer, IFE-3:TOFU-6:PID-3:ERH-2 form 2:2:2:2 (Perez-Borrajero et al. 2021). For further structural studies, PETISCO was split into three sub-complexes, which will each be discussed in the next three sections.

3.2.2. The PID-3:TOFU-6 sub-complex.

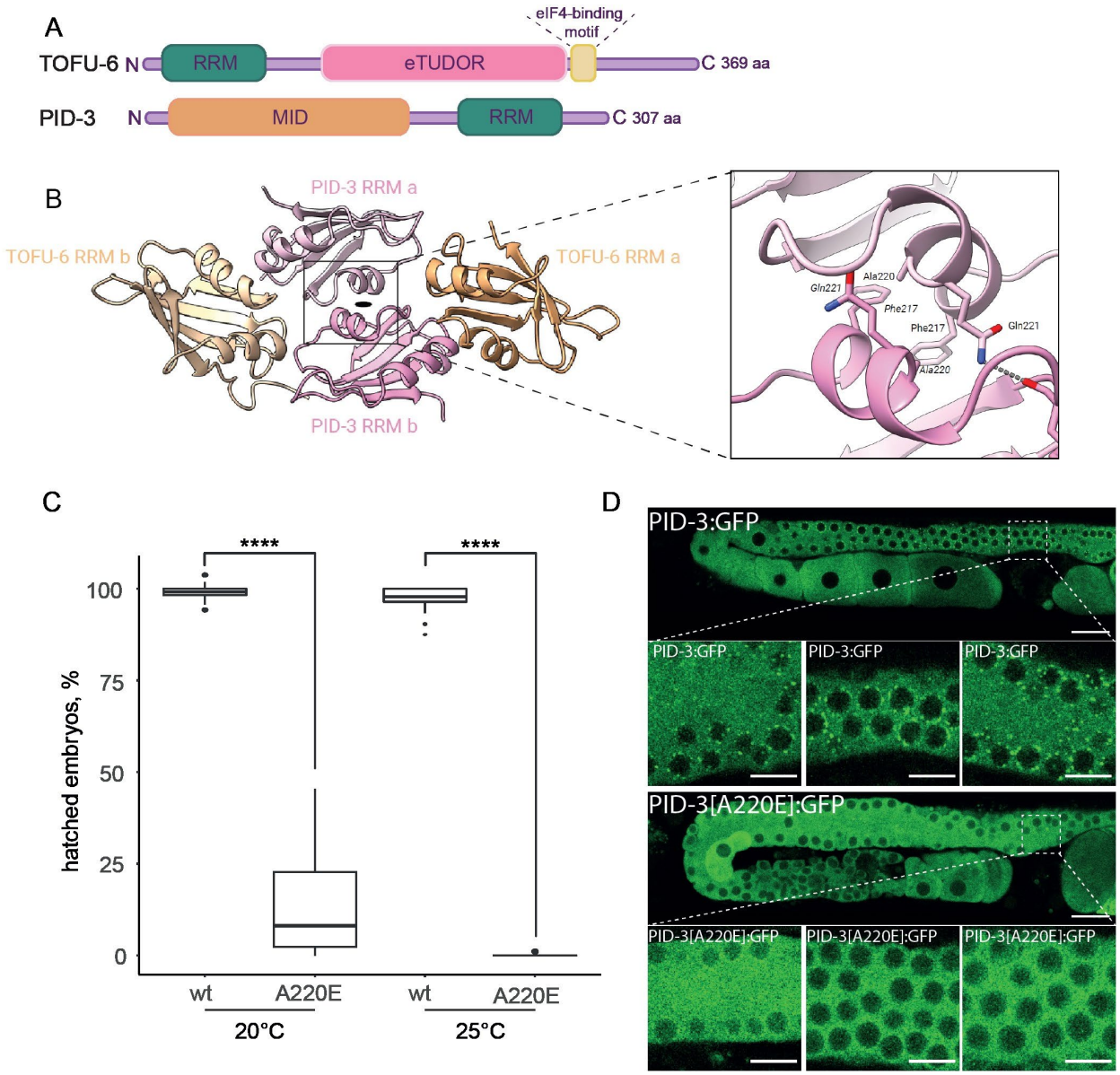

Figure R8. PID-3 RRM interacts with TOFU-6 RRM and they form a tetramer. A. Schematic representation of domain organization of TOFU-6 and PID-3. **B**. Purification of the proteins and crystallization were performed in the Falk Laboratory. On the left: Crystal structure of the PID-3^{RRM}/TOFU-6^{RRM} complex shown in cartoon representation. The two PID-3^{RRM} domains are shown in different shades of pink, and the two TOFU-6^{RRM} protomers are shown in different shades of orange. On the right: Zoom in on the dimerization surface of PID-3 RRM protomers. **C**. Box plot showing the percentage of hatched embryos of *pid-3::*gfp (wt) and *pid-3(a220e):gfp* animals grown at 20°C and 25°C. The progeny of 30 different mothers was analyzed for each condition, and the development of at least 2600 eggs was scored. A two-sided t-test was used to assess significance. **D**. Single-plane confocal micrographs of PID-3::GFP(WT) (top) and PID-3[A220E]::GFP (bottom) at 25°C. The boxes indicate the regions (above the spermatheca) from which three zoomed-in examples are shown below. Scale bars: overview, 20 µm; zoom-in, 8 µm.

PID-3 could not be purified in a good enough quality to perform the study with full-length protein, therefore the RRM domain and MID domains of PID-3 were purified separately. A combination of pull-downs and size exclusion chromatography helped to narrow PID-3/TOFU-6 interaction to RRM domains of both proteins (Figure R8.A). The PID-3 RRM domain forms a homodimer, probably contributing to the dimerization of PETISCO (Figure R8.B). Structural analysis of the crystal structure of co-crystalized RRM domains of PID-3 and TOFU-6 showed that dimerization of PID-3 happens via an alpha-helix by a combination of hydrophobic and polar interactions, and that dimerization of PID-3 is necessary for the interaction between TOFU-6 and PID-3 (Figure R8.B) (Perez-Borrajero et al. 2021).

Based on the structure, mutations were designed to interfere with PID-3 dimerization. Mutation A220E successfully prevented PID-3 dimerization *in vitro* (Perez-Borrajero et al. 2021). To confirm the importance of dimerization of PID-3 *in vivo* I first endogenously tagged PID-3 with eGFP on the C-terminus by CRIPSR/CAS9, so it can be detected using fluorescent microscopy. Afterward, I introduced the A220E mutation into the RRM domain of eGFP-tagged PID-3 by CRISPR/CAS9 and scored the resulting animals for a Mel phenotype and PID-3:eGFP localization. First, the PID-3[A220E]:eGFP strain showed a strong, but incomplete Mel phenotype at 20°C. This phenotype worsened with rising temperature, as none of the eggs hatched when worms were kept at 25°C (Figure R2.B). Second, the PID-3(A220E) point mutant did not localize to perinuclear granules anymore, while it was expressed at similar levels as the wild-type protein (Figure R2.C). Taken together, dimerization of PID-3 is necessary for the localization of PID-3 into P-granules and is required for PETISCO function. Collection of the worms for the small RNA sequencing was not possible, due to the severity of the Mel phenotype, as not enough worms hatched after the bleaching. However, I assume that PETISCO function, related to 21U RNA precursor processing was affected as well as the function in the embryonic development.

3.2.3. The PID-3:ERH-2 sub-complex

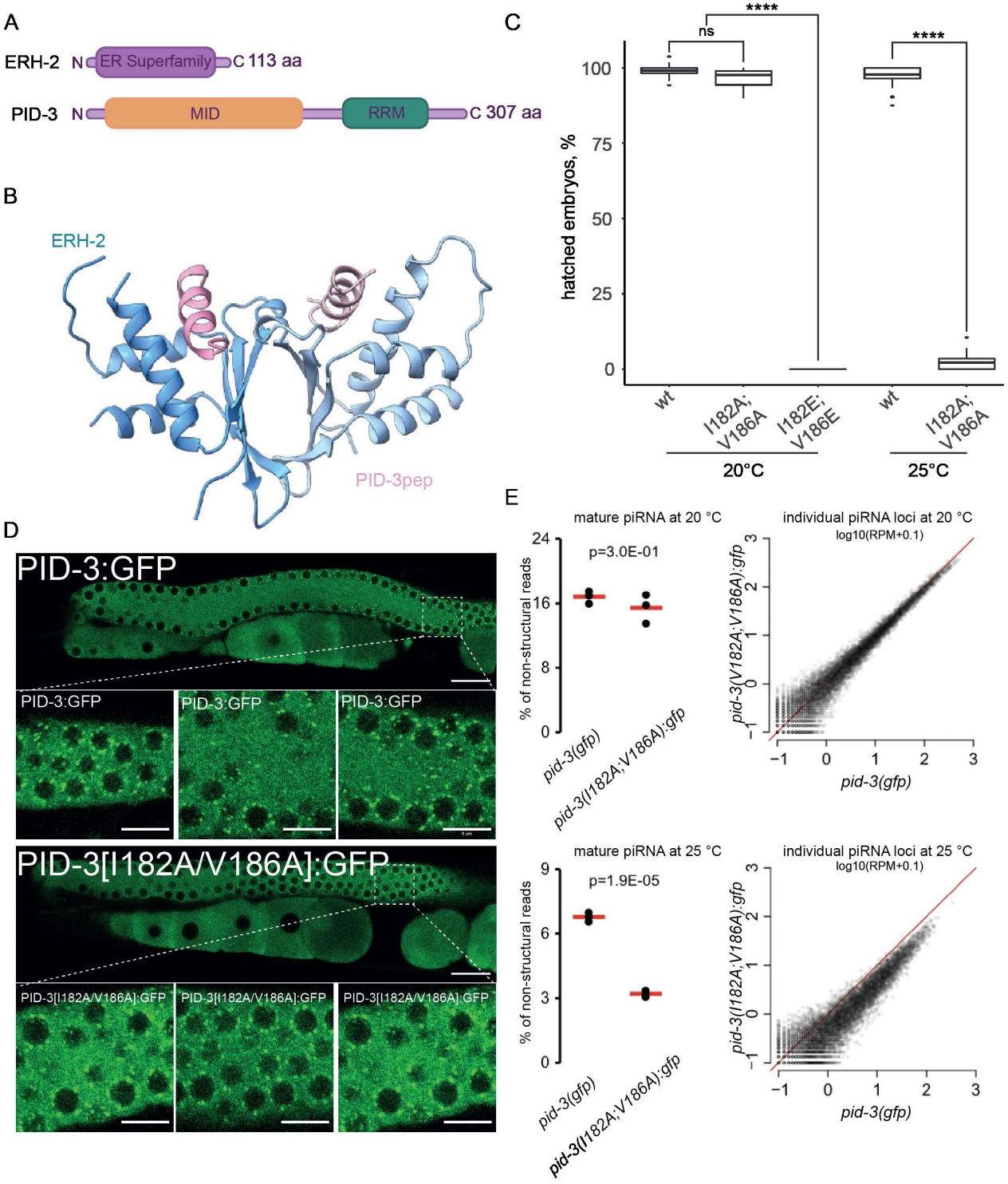

Figure R9. PID-3 interacts with ERH-2. A. Schematic representation of domain organization of ERH-2 and PID-3. **B**. Purification of the proteins and crystallization were performed in the Falk Laboratory. Crystal structure of the PID-3^{pep}/ERH-2 complex shown in cartoon representation. The two ERH-2 protomers are shown in different shades of blue, while the two PID-3 peptides are in different shades of

pink. **C**. Box plot showing the percentage of hatched embryos of *pid-3:gfp (wt)* and *pid-3(i182a;v186a):egfp* and *pid-3(i182e;v186e):gfp* animals grown at 20 and 25°C. The progeny of 30 different mothers was analyzed for each condition, and the development of at least 2600 eggs was scored. A two-sided t-test was used to assess significance. P-values are indicated in the graph. **D**. Single-plane confocal micrographs of PID-3::GFP(WT) (top) and PID-3[V182A;V186A]::GFP (bottom) at 25°C. The boxes indicate the regions (above the spermatheca) from which three zoomed-in examples are shown below. Scale bars: overview, 20 µm; zoom-in, 8 µm. **E**. Total 21U levels in wild-type and pid-3[i182a/v186a]-mutant embryos grown at 20°C (top left) and 25°C (bottom left) from three biological replicates. Group means are depicted by red lines and P-values are calculated using a two-tailed unpaired t-test. On the right: Scatter plots depicting the relative abundance of individual 21U loci in pid-3(i182a/v186a)-mutant versus wild-type embryos grown at 20°C (top) and 25°C (bottom).

Using pull-down experiments and SEC, PID-3 was found to interact with ERH-2 via a short region located in the N-terminus of the RRM domain (aa 177-193) (Figure R9.A). A crystal structure of a peptide covering this region bound to ERH-2 showed two things: the PID-3 peptide binds ERH-2 through the formation of a hydrophobic interaction surface between PID-3 and ERH-2, and ERH-2 forms a dimer (Figure R3.B). Indeed, ERH was found to be a stable dimer also in other species, and similar interactions as we find between PID-3 and ERH-2 have been described for Mmi1 and ERH in fission yeast, and for DGCR8 and ERH in humans. Based on the structure, I182 and V186 of PID-3 were identified as important residues for interaction. Mutation of residues I182 and V186 to alanine or glutamic acid indeed disrupted the interaction *in vitro.* I introduced these mutations *in vivo* into the PID-3:eGFP strain described above. Mutations of residues I182 and V186 to alanine did not cause a severe phenotype at 20°C, but caused a full MEL phenotype when temperature was elevated to 25°C. Mutations of residues I182 and V186 to glutamic acid resulted in a 100% MEL phenotype already at 20°C (Figure R3.C). We did not observe changes in PID-3 localization in *pid-3(V182A;V186A):gfp* strain (Figure R3.D).

On the side of ERH-2 T13 was identified as important for interaction *in vitro*: a T13E mutation in ERH-2 abrogated the interaction between ERH-2 and PID-3 peptide *in vitro*. The introduction of the analogous mutation *in vivo* did not, however, cause a MEL phenotype, suggesting that *in vivo* this mutation can somehow be compensated (data not shown).

To test the effect of the interaction of PID-3 with ERH-2 on the level of 21U RNA production, young adult animals of the PID-3(I182A;V186A):eGFP strain were collected for small RNA sequencing. The temperature-dependent phenotype of the PID-3(I182A;V186A):eGFP

animals allowed testing of two conditions: animals were first grown at 20°C, and eggs were isolated and subsequently hatched. The resulting L1 larvae were then plated and grown at either 20°C or 25°C, until the young adult stage. This revealed a significant reduction of 21U RNAs only at 25°C (Figure R3.D), consistent with the temperature-dependent Mel phenotype. Thus, the disruption of PETISCO affects both PETISCO functions.

3.2.4. The ERH-2:TOST-1/PID-1 sub-complex.

The binding of PID-1 or TOST-1 to ERH-2 determines the function of PETISCO complex. TOST-1 directs PETISCO towards embryonic development while PID-1 associated PETISCO takes part in 21U RNA precursor formation. TOST-1 and PID-1 interact with ERH-2 via a conservative motif, present in both proteins (Figure R4.A) (Cordeiro Rodrigues et al. 2019). Thus, the next question was, if PID-1 and TOST-1 share an interaction surface with ERH-2 and if TOST-1 and PID-1 compete with each other for the interaction with ERH-2.

Obtaining PID-1 and TOST-1 that were suitable for structural studies was a very hard task, as these proteins are unstructured and prone to degradation (Figure R7). ERH-2 is a well-expressed protein, which can be purified in high concentrations, and therefore had been decided to study the interactions between ERH-2 and TOST-1/PID-1 using a two-dimensional protein NMR, using synthetic PID-1 and TOST-1 peptides. Purification of ERH-2 on the minimal media, containing ^{15}N and ^{13}C was successful (Figure 4. B). Next, synthetic peptides for wild-type TOST-1 and PID-1 were designed, covering the conserved region, as well as a mutant TOST-1 peptide, containing the R42C mutation that prevents binding to ERH-2 (Cordeiro Rodrigues et al. 2019) (Figure R4.A).

NMR analysis was done in collaboration with the group of Janosch Hennig. Initial analysis showed that chemical shifts for ERH-2 occurred in the same residues when ERH-2 was incubated with either TOST-1 or PID-1, suggesting, that PID-1 and TOST-1 bind the same interface on ERH-2. Chemical shifts were stronger for ERH-2 – TOST-1 than for ERH-2-PID-1 interactions, indicating that most probably TOST-1 binds ERH-2 stronger than PID-1. Measurements done by isothermal titration calorimetry (ITC) of PID-1/TOST-1 peptide interactions with ERH-2 confirmed this observation: TOST-1 could effectively compete with PID-1 from ERH-2 (Perez-Borrajero et al. 2021).

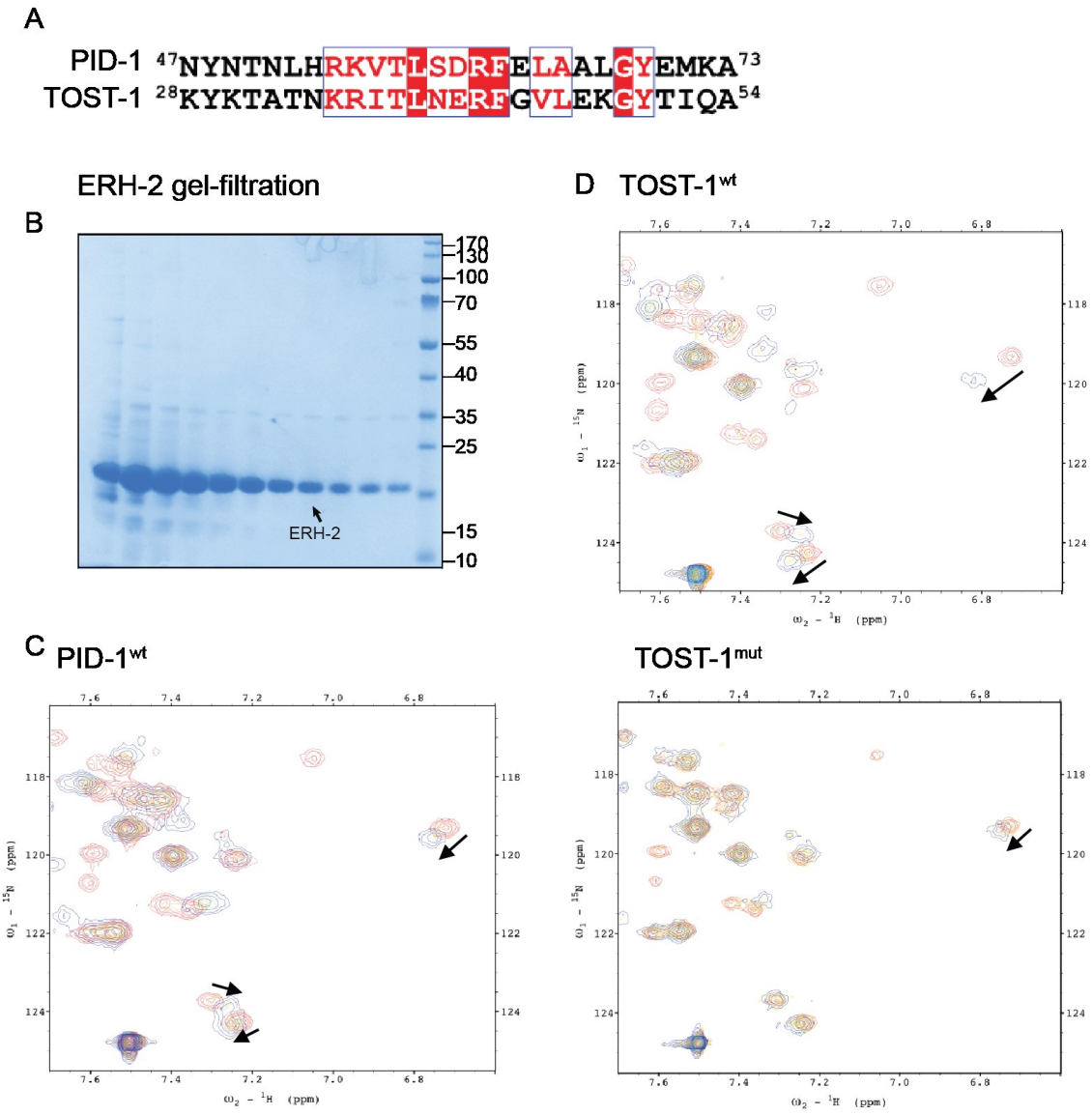

Figure R10. TOST-1 and PID-1 share the interaction surface on ERH-2. **A**. Amino acid sequence of the PID-1 and TOST-1 synthetic peptides used for the NMR studies. **B**. SDS-PAGE of purification of ERH-2 on the minimal media with ^{15}N. **C**. Chemical shifts of labeled ERH-2 protein incubated with PID-1 peptide. **D.** Chemical shifts of labeled ERH-2 protein incubated with TOST-1 wild-type (top) and TOST-1 mutant (bottom) peptides. Signals of free ERH-2 are shown in pink, and signals of ERH-2 incubated with peptides are colored in blue.

4. Results II. PUCH is a novel SLFN-based nuclease

Parts of the figures included in this chapter are published in the following scientific papers:

Podvalnaya, N., Bronkhorst, A.W., Lichtenberger, R. *et al.* piRNA processing by a trimeric Schlafen-domain nuclease. *Nature* **622**, 402–409 (2023). https://doi.org/10.1038/s41586-023-06588-2

4.1. TOFU-1 and TOFU-2 are potential nucleases processing 21U RNA precursor.

There were no signs of the presence of a potential nuclease domain within PETISCO, nor among PETISCO-interacting proteins, detected by mass spectrometry (Rodrigues et al, 2019). Thus, the current understanding of PETISCO role in 21U RNA biogenesis is that PETISCO binds 21U RNA precursors via the IFE-3 cap-binding domain, to assist with 21U RNA precursor cleavage.

Proteins, shown to be involved in the 21U RNA pathway were analyzed for the presence of potential nuclease. An RNAi screen, performed in 2014 by the Hannon Lab, identified proteins, that are involved in 21U RNA biogenesis in *C. elegans* (Goh et al. 2014). From this screen, TOFU-1 and TOFU-2 had the strongest effect on mature 21U RNA levels. Furthermore, a single knockout of the two corresponding genes resulted in the accumulation of 21U RNA precursors, suggesting that TOFU-1 and TOFU-2 are involved in 21U RNA precursor processing. However, at that point in time, no nuclease domains were detected in these two proteins.

Protein sequences of TOFU-1 and TOFU-2 were reanalyzed with structure-based homology search HHPred and Alphafold2. Both proteins showed homology to the Schlafen-protein family, in particular to rat Schlafen13 and human Schlafen12 and Schlafen5. Besides their Schlafen-like domain, both proteins have additional domains: TOFU-1 contains an acidic N-terminus and TOFU-2 accommodates a SPRY domain with yet unknown function on the N-terminus (Figure R11.A). Many proteins, containing the SPRY domain have been shown to participate in the innate immune response and RNA processing, SPRY domain itself may drive protein-protein interactions (D'Cruz et al. 2013).

Many proteins from the Schlafen family possess nuclease activity; however, in the well-characterized cases, the nuclease's active center is formed by two SLFN folds. Curiously, both TOFU-1 and TOFU-2 are predicted to have only a single SLFN fold. There is a possibility, that TOFU-1 and TOFU-2 interact to form an active nuclease that would catalyze piRNA precursor processing (Figure R11.B).

To test if TOFU-1 and TOFU-2 interact, TOFU-2 was endogenously tagged with an HA-tag using CRISPR/CAS9, animals were synchronized and collected at the young adult stage for an IP-MS experiment. TOFU-1 was identified among the TOFU-2 interactors (Figure R5.C). Thus, TOFU-1 and TOFU-2 may indeed interact to form an active nuclease.

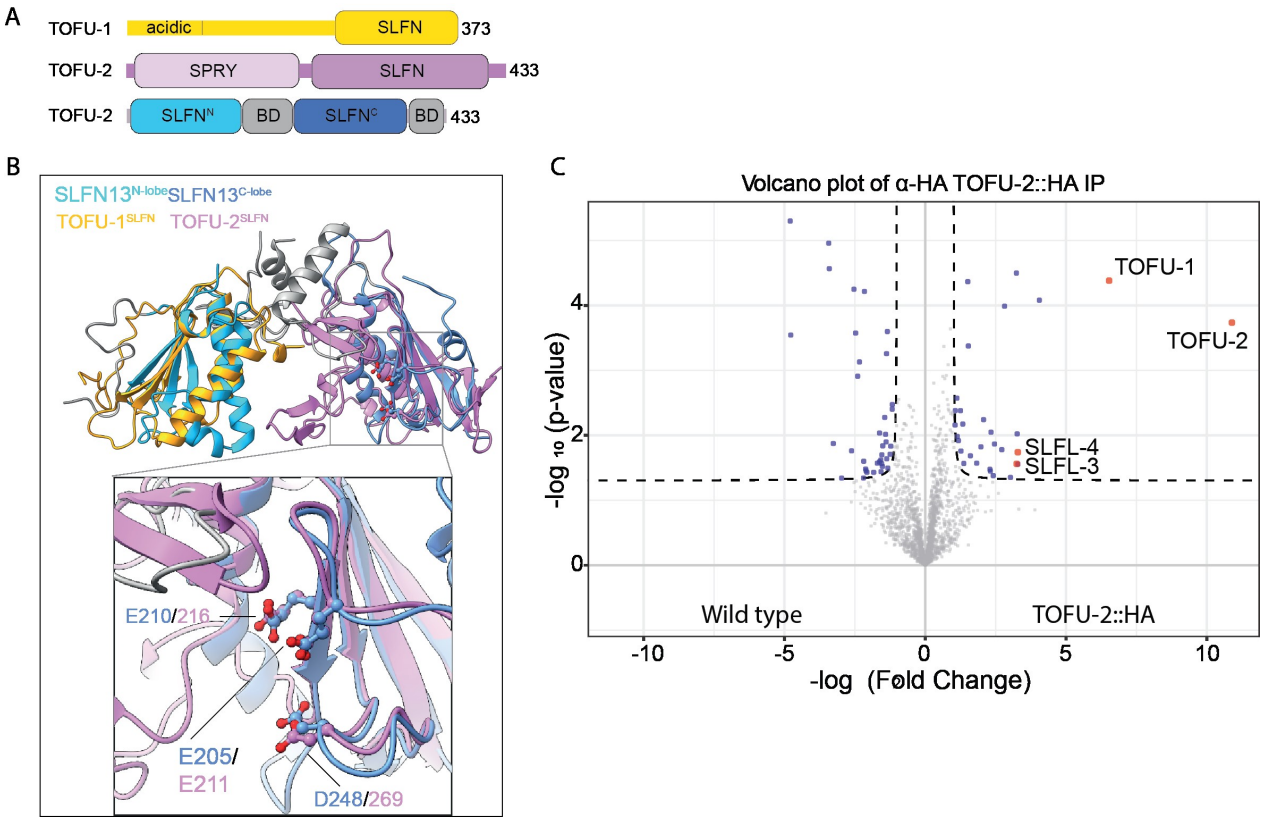

Figure R11. TOFU-1 and TOFU-2 form putative nuclease complex. A. Schematic representation of domain organization of TOFU-1, TOFU2, and SLFN13. **B**. Structural alignment Alphafold 2 predicted TOFU-1 and TOFU-2 SLFN domains onto the crystal structure of the N-terminal SLFN13 endoribonuclease domain (Protein Data Bank (PDB): 5YD0). Domains are colored as in **A**. **C**. Label-free proteomic quantification of TOFU-2–HA and wild-type immunoprecipitates from young adult extracts. n = 4 biological replicates. The x-axis shows the median fold enrichment of individual proteins, and the y-axis shows $-\log^{10}[P]$. P values were calculated using Welch two-sided t-tests. The dashed lines represent enrichment thresholds at P = 0.05 and fold change > 2, curvature of enrichment threshold c = 0.05. The dots represent enriched (blue/red) or quantified (grey) proteins.

4.2. TOFU-2 contains a catalytic center of Schlafen-like nuclease.

To identify the potential catalytic center of the TOFU-1-TOFU-2 complex, AlphaFold2 predicted TOFU-2 and TOFU-1 structures were aligned to the structure of the rat Schlafen-13. This alignment showed that TOFU-2 contains a potential catalytic center, and TOFU-1 does not contain it (Figure R11.B). Glutamic acid residues, crucial for the Schlafen13 nuclease function, because they coordinate a divalent metal ion (Yang et al. 2018), are conserved in TOFU-2 (Figure R12.A). To test if mutation of this potential catalytic center in TOFU-2 affects 21U RNA biogenesis, glutamic acid residue in position 216 of TOFU-2 was mutated to alanine, by CRISPR/CAS9 genome

editing of the *tofu-2* gene. To resolve the effect of the mutation, *tofu-2(e216a)* mutant worms were crossed with a 21U RNA sensor-carrying strain (Figure R12.B). These 21U RNA sensor-carrying animals produce mCherry, fused to H2B histones, controlled by a 21U RNA recognition site (21UR1 to be precise). In the presence of 21U RNAs, the mCherry transcript will be recognized and silenced. If the 21U RNA pathway is compromised, the 21U RNA sensor will escape silencing and be expressed; in this case, mCherry fluorescence can be observed in the nuclei, on the chromosomes, due to the H2B fusion.

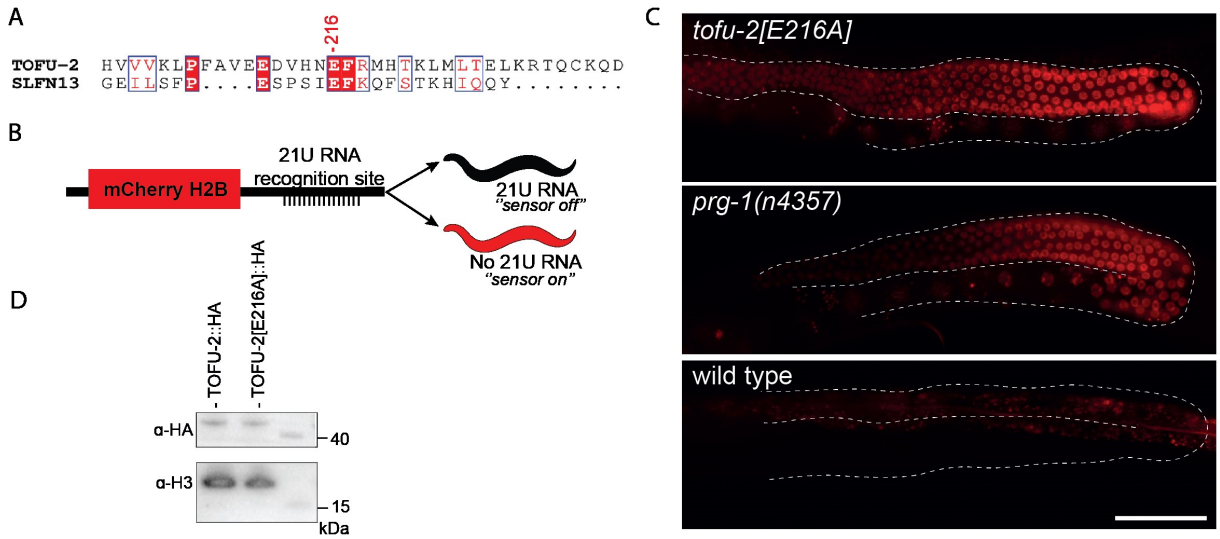

Figure R12. **TOFU-2 contains a catalytic center of 21U RNA processing nuclease complex. A**. Sequence alignment of the catalytic center of rSLFL-13 and TOFU-2. Glutamic acid residue used for the creation of *tofu-2(e216a)* strain is annotated. **B**. Schematic representation of 21U RNA sensor strain. **C**. Wide-field fluorescence microscopy images of adult hermaphrodites carrying the 21U RNA sensor in the following three genetic backgrounds: *tofu-2(e216a)* (top), *prg-1(n4357)* (middle), and wild type (bottom). Germlines are marked with white dashed lines. Scale bar, 50 µm. **D**. Westen blot performed on *tofu-2:ha* and *tofu-2(e216a):ha* strains with anti-HA antibodies, anti-H3 antibodies are used as a control.

As mentioned above (see introduction), recognition by 21U RNA can cause permanent, or rather a form of 21U-independent silencing of the target, known as RNAe. To avoid this RNAe effect, the sensor strain was crossed in via a strain carrying a *mut-7* deletion, which prevents RNAe. During the crosses, and selfing of the offspring, the *mut-7* deletion was removed, allowing me to score the effect of the *tofu-2(e216a)* mutation. Wide-field fluorescent microscopy on the obtained strain revealed bright red fluorescence in the nucleus, indicating that mutation *tofu-2(e216a)* disables the 21U RNA pathway (Figure R12.C).

To confirm, that the expression level of the TOFU-2 protein was not affected, I also introduced the *e216a* mutation into the *tofu-2::ha* allele and compared TOFU-2 levels in *tofu-2::ha* and *tofu-2(e216a)::ha* strains by Western-blot. The *e216a* mutation did not affect levels of TOFU-2 protein levels (Figure R12.D).

4.3. Mutation of the putative TOFU-2 catalytic center causes the accumulation of 21U RNA precursors.

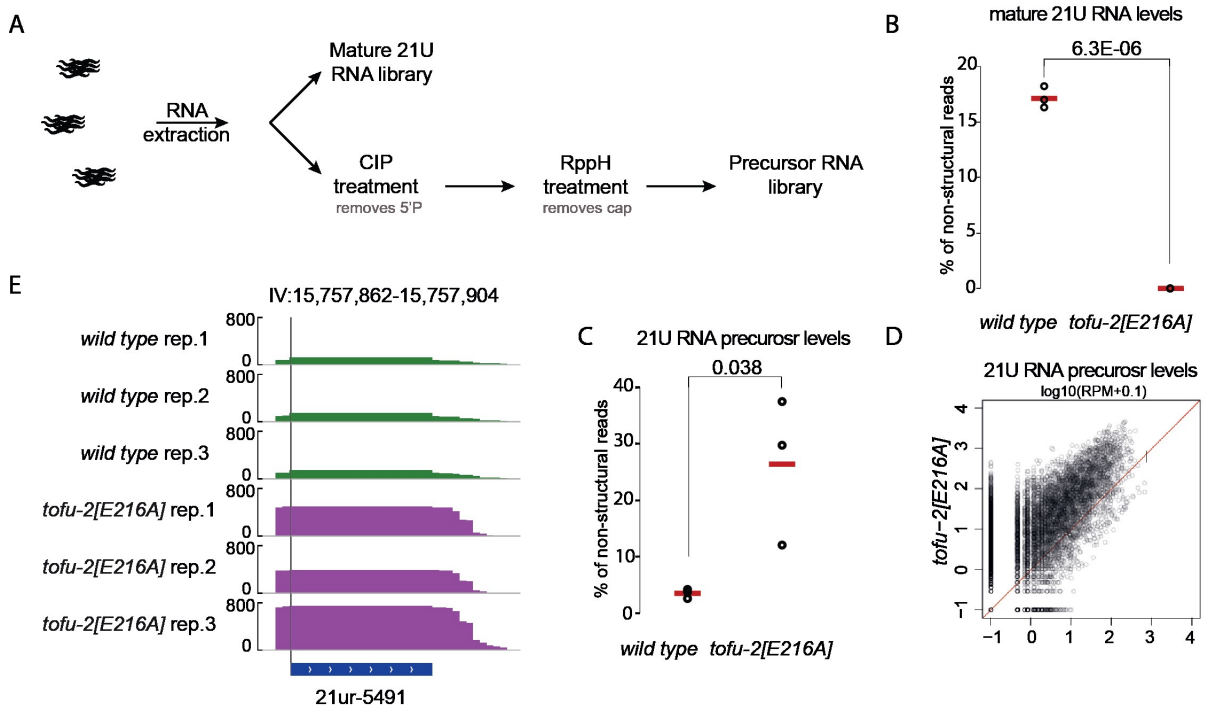

Figure 13R. Mutation in TOFU-2 catalytic center causes accumulation of 21U RNA precursors. A. Scheme of 21U RNA precursor library preparation. **B**. Total mature piRNA levels (type 1) in wild-type and *tofu-2*(*E216A*)-mutant young adult hermaphrodites. *n* = 3 biological replicates. The red lines show the group means. *P* values were calculated using two-tailed unpaired *t*-tests. **C**. Total 21U RNA precursor levels in wild-type and *tofu-2*(*e216a*)-mutant young adult hermaphrodites. *n* = 3 biological replicates. The red lines show the group means. *P* values were calculated using two-tailed unpaired *t*-tests. **D**. The relative abundance of piRNA precursors from individual loci in *tofu-2*(*e216a*)-mutant versus wild-type young adult hermaphrodites. *n* = 3 biological replicates. RPM reads per million non-structural small RNA reads. **E**. Genome browser tracks individual 21U RNA loci, displaying normalized read coverage in piRNA precursor libraries. The top three tracks (green) are derived from a wild-type background, and the bottom three tracks (purple) are from a *tofu-2(e216a)* mutant background. Mature piRNAs are severely depleted from precursor libraries, thus most of the depicted read coverage derives from piRNA precursors starting 2 nucleotides upstream of the 5'-ends of mature piRNAs which are indicated by a vertical line.

Next, small RNA sequencing was performed for both mature 21U RNA and 21U RNA precursors. Separate libraries had to be created for mature 21U and for precursor RNAs, as precursor sequences are much less abundant and carry a 5' cap, which interferes with cloning. Worms were grown in triplicates till the young adult stage, and RNA was isolated and split to prepare two libraries – for precursors and for mature 21U RNA. RNA for precursor libraries was treated with Quick CIP enzyme to remove 5' phosphates in order to prevent cloning any other types of RNA and enhance the possibility of cloning 21U precursors. After that, the decapping of RNA was performed with the RppH enzyme to remove the cap from the precursors and enable cloning by ligation of adapter sequences to the 5' and 3' ends (Figure R13.A). Libraries for mature 21U RNAs, which carry a 5' monophosphate, were made without the Quick CIP and RppH treatments.

In *tofu-2(e216a)* samples, the almost complete absence of mature 21U RNA reads was observed (Figure R13.B). Furthermore, levels of 21U RNA precursors were 3 to 5 times higher in the mutant than in the wild-type worms (Figure R13.C and R13.D). Reads obtained from *tofu-2(e216a)* precursor started two nucleotides upstream of the 5' end of mature piRNAs (Figure R13.E), as expected for 21U RNA precursors (Ruby et al. 2006) . Thus, by introducing a single mutation into the putative TOFU-2 catalytic center the complete *tofu-2* deletion phenotype was reconstituted, suggesting that TOFU-2 indeed harbors catalytic activity that is important for 21U RNA formation *in vivo*.

4.4. 21U RNA precursors could carry RNA modifications important for their processing.

21U RNA precursors may carry modifications, which could be affecting their processing. 21U RNA precursors do not share a consensus motif, however, 78% of precursors contain adenine in position 1 or 2, or both. To test if there are any modifications present on the 21U RNA precursor, samples of *tofu-6:3xmyc* (wild-type) and *tofu-2(e216a);tofu-6:3xmyc* (mutant) young adult worms were collected and split for two sets of samples. From the first sample, total RNA was isolated; from the second PETISCO proteins were extracted by anti-MYC IP, and from that IP RNA was isolated. Finally, RNA samples were analyzed by mass spectrometry for the presence of RNA modifications. Results were normalized to total adenine levels. We observed the enrichment of Am, m7G, Cm, m6A, m1I, and, possibly, Um enrichment in the IP-derived RNA, compared to the total RNA level (Figure R6). 21U RNA precursors are not the only kind of RNA bound by

PETISCO (Rodrigues, 2019), and therefore, more precise isolation of 21U RNA precursors is required for the identification of modification, specific to the precursors. However, we observe a slight difference between levels of m7G in *tofu-2(e216a)* mutants and wild-type worms. As most of the total RNA is ribosomal RNA, and most of the mRNAs in *C. elegans* are trans-spliced and therefore do not carry m7G cap, it is possible, that most of the visible m7G signal belongs to 21U RNA precursors. However, besides the presence in the cap structure, m7G can occur on the Gs inside of the RNA, such modifications are common, for instance, on the tRNA (Tomikawa 2018). Applied RNA mass spectrometry protocol includes dephosphorylation, therefore it would not be possible to distinguish between the m7G in the cap and internal m7G. To address whether precursors are responsible for the all m7G signal, better isolation of 21U RNA precursors will be required.

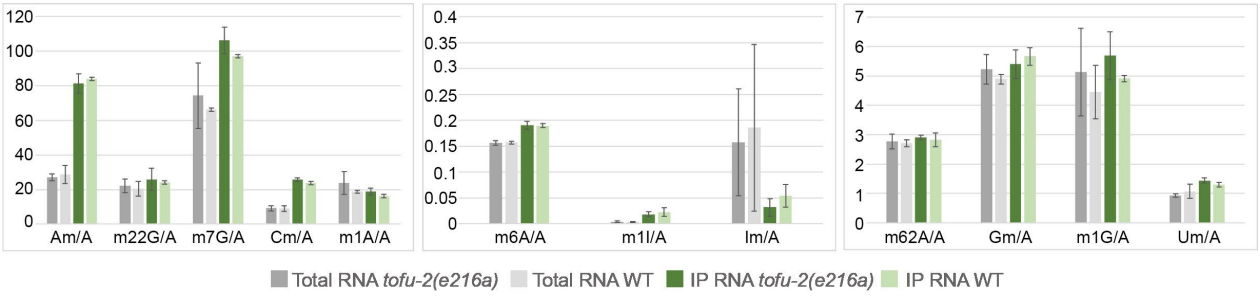

Figure R14. Investigation of various modifications on 21U RNA precursors. The presence of Am, m22G, m7G, CM, m1A, m6A, m1I, Im, m62A, Gm, m1G, and Um was analyzed by RNA mass spectrometry. The total level of RNA modifications found in total RNA obtained from *tofu-2(e216a)* and wild-type worms and in RNA isolated from tofu-6:3xMYC IP material from *tofu-2(e216a)* mutant background and wild-type worms. All values are normalized to the total signal of adenine. Plots were separated based on the y-axis. All measurements were done in triplicates, except the IP RNA WT, which is done in duplicate, as the sample was destroyed during the mass spectrometry.

4.5. SLFL-3/4 are SLFN-fold proteins that interact with TOFU-1 and TOFU-2.

After identification of the catalytic center in TOFU-2, I made several attempts to confirm the activity of the TOFU-1/TOFU-2 nuclease complex *in vitro*.

Purification of TOFU-1 and TOFU-2 from *E. coli* was unsuccessful, as TOFU-1 was expressed at low levels and TOFU-2 was well-expressed, but not soluble (data not shown). Therefore, I attempted to co-purify TOFU-1 and TOFU-2 from SF9 cells, which resulted in a very little amount of the protein, which we used in the cleavage reaction with synthetic RNA oligonucleotide. We were not able to observe any cleavage.

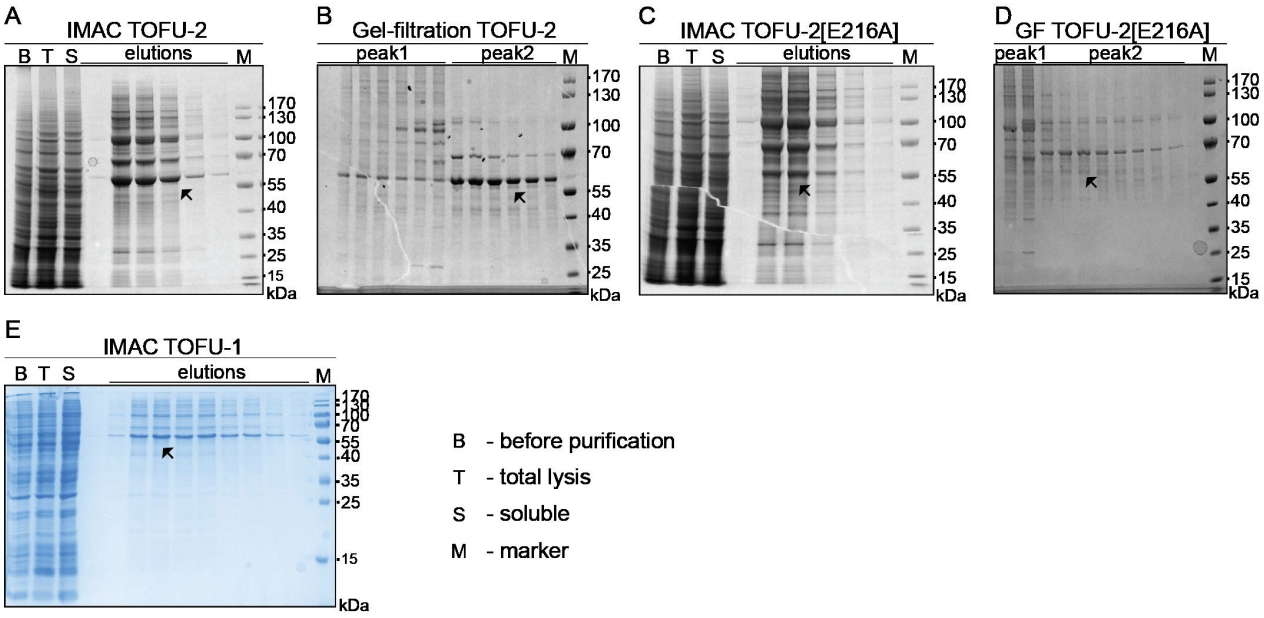

Figure R15. Purification of TOFU-1 and TOFU-2 from SF9 cells. A. SDS PAGE of IMAC purification step of TOFU-2. **B**. SDS PAGE of gel-filtration purification step of TOFU-2. **C**. SDS PAGE of IMAC purification step of TOFU-2[E216A]. **D**. SDS PAGE of gel-filtration purification step of TOFU-2[E216A]. **E**. SDS PAGE of IMAC purification step of TOFU-1. **A-E.** The expected protein band is marked with an error.

After many unsuccessful attempts to achieve cleavage of 21U RNA precursors, concerns were raised, that there might be a missing component in the cleavage complex. Among IP-MS interactors of TOFU-2 two other proteins containing a single SLFN-fold like TOFU-1 and TOFU-2 were identified (Figure R11.C; gene names C35E7.8 and F36H12.2; Figure R16.A). We named these two proteins SLFL-3 and SLFL-4, where SLFL stands for Schlafen-like, and TOFU-1 and TOFU-2 received the alternative names SLFL-1 and SLFL-2 respectively.

In fact, SLFL-3 had appeared in the RNAi screen for 21U biogenesis factors, where TOFU-1 and TOFU-2 were first identified (Goh et al. 2014). However, the reduction in 21U RNA levels in *slfl-3* mutants was below the threshold that was set to be further investigated. SLFL-3 and SLFL-4 are 90% identical on the amino acid level (Figure R16.B) and most probably act redundantly, and therefore RNAi against *slfl-3* (C35E7.8) did not show a strong phenotype.

To better understand the structure of potential nuclease complexes we started a collaboration with the laboratory of Sebastian Falk (Max Perutz, Vienna). They obtained Alphafold prediction of the interaction between SLF3/4, TOFU-1, and TOFU-2 and found, that these proteins form a stable trimer, in which TOFU-1, TOFU-2, and SLFL3 or SLFL4 are present (Figure R16.C).

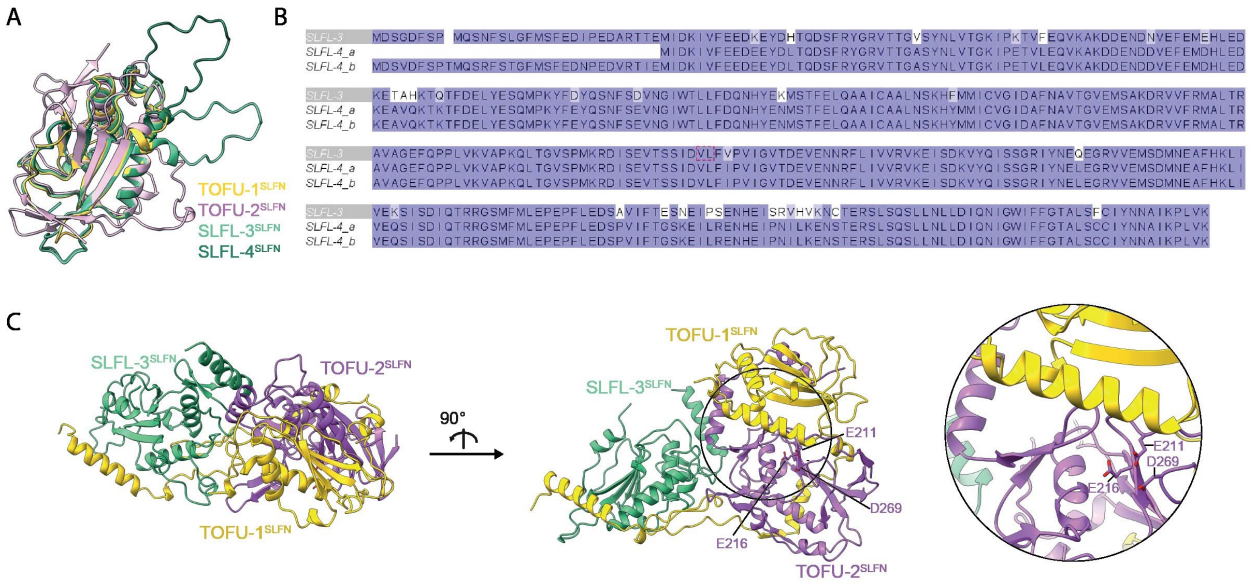

Figure R16. TOFU-1, TOFU-2, and SLFL-3/4 form a PUCH complex. A. Alignment of the Alphafold predicted SLFN-domains for TOFU-1, TOFU-2, SLFL-3, and SLFL-4. **B**. Sequence alignment of SLFL-3 and SLFL-4. Two proteins are 90% identical on the amino acid level. Alignment performed in Crustal Omega. **C**. AlphaFold2-predicted structure of TOFU-1–TOFU-2–SLFL-3 SLFN-folds interaction, shown in two different orientations. TOFU-1 is shown in yellow, TOFU-2 in purple, and SLFL-3 in green. The TOFU-2 active-site residues are shown as a stick representation and are magnified in the circle on the right.

4.6. SLFL-3 and SLFL-4 act redundantly in 21U RNA precursor processing.

To test if SLFL-3 and SLFL-4 are involved in the 21U RNA pathway, deletion mutants of *slfl-3* and *slfl-4* were created by CRISPR/CAS9 and crossed to the 21U RNA sensor strain (described above in Figure R12.B). In *slfl-3* single mutants, a weak red fluorescent signal in the nucleus was observed, which disappeared within a few generations (Figure R17.A). Such a behavior has been observed before in *henn-1* mutant (Luteijn et al. 2012); here it was explained by a partial disruption of the 21U RNA pathway, followed by activation of the RNAe response. In the case of *slfl-3* this may be similarly true, but activation of *slfl-4* expression, in case these two proteins act redundantly, can also not be excluded.

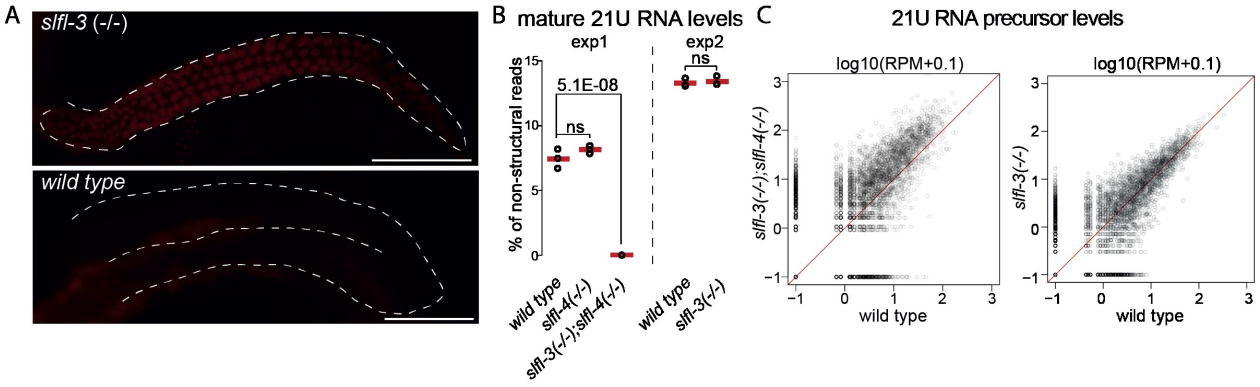

Figure R17. SLFL-3 and SLFL-4 participate in 21U RNA precursor processing. A. Wide-field fluorescence microscopy images of adult hermaphrodites carrying the 21U RNA sensor: *slfl-3(-/-)* (top) and wild type (bottom). Germlines are marked with white dashed lines. Scale bar, 50 µm. **B**. Total mature 21U RNA levels in young adult hermaphrodites of the indicated genotypes. The experiment is done in three biological replicates. The red lines depict group means; *P* values were calculated using one-way analysis of variance (ANOVA) followed by Tukey's honest significant difference (HSD) test (left) and two-tailed unpaired *t*-tests (right). The plot is based on two independent experiments (exp. 1 and 2). NS, non-significant. **C**. The relative abundance of 21U RNA precursors from individual loci in *slfl-3(-/-)* and *slf-3(-/-) ;slfl-4(-/-)* mutant versus wild-type young adult hermaphrodites. Performed in three biological replicates.

To exclude the compensatory effects of s*lfl-4*, *slfl-3; slfl-4* double mutants were created, and animals were synchronized and grown to the young adult stage. Small RNA sequencing was performed for single deletions of *slfl-3* and *slfl-4*, as well as for the double mutant. While the single mutants did not have strong effects on the mature 21U RNA levels (Figure R17.B) and did not show accumulation of 21U RNA precursors (Figure R17.C), deletion of both *slfl-3* and *slfl-4* resulted in the complete absence of mature 21U RNAs and increased numbers of 21U RNA precursors (Figure R17.B and C). Thus, simultaneous loss of SLFL-3 and SLFL-4 has the same effect as mutation of the catalytic center of TOFU-2. Therefore, the conclusion is that all proteins, TOFU-1, TOFU-2 and SLFL-3/4 are involved in 21U RNA precursor cleavage. We named the potential nuclease complex PUCH, which stands for "precursor of 21U RNA 5'-end cleavage holoenzyme".

4.7. Immuno-purified PUCH complex can process 21U RNA precursors.

In order to test the activity of PUCH *in vitro* m^7G-capped synthetic 21U RNA precursor was ordered from the Biosynthesis™. The sequence of the synthetic precursor corresponds to the sequence of 21U RNA-7794, one of the most abundant sequences found in small RNA sequencing datasets described above. The synthetic precursor was radioactively labeled by ligation of [5'-

P32]pCp to the 3'end for visualization. I will refer to this precursor as the "m^{7}G-AAU" precursor (Figure R18.A).

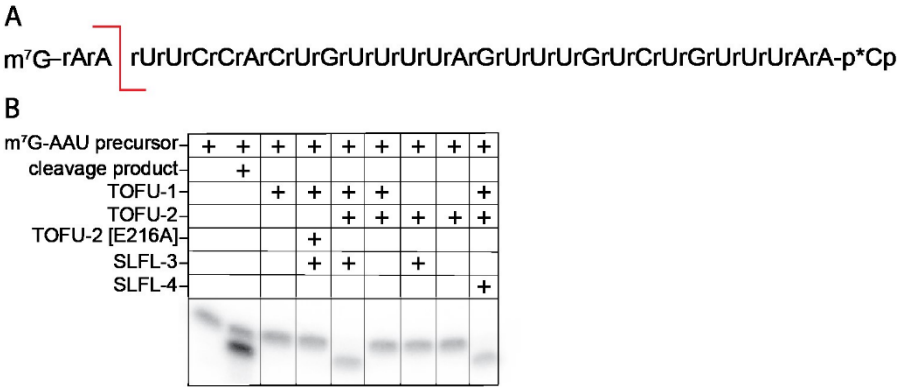

Figure R18. IP-PUCH can cleave 21U RNA precursor. A. Sequence of 21U RNA precursor ("m^{7}G-AAU" precursor) used in most assays if not stated otherwise. The red line indicates the position of the expected cleavage. The labeled precursor was used as a marker in lines 1 and 2 of every gel, except R8.N, as well as the labeled cleavage product. **B.** *In vitro* cleavage assay of the 21U RNA precursor using anti-GFP immunoprecipitated material from BmN4 cell extracts. Cells were transfected with eGFP–TOFU-2, eGFP-TOFU-2[E216A], TOFU-1, SLFL-3 or SLFL-4 at various combinations, as indicated.

Purification of TOFU-1 and TOFU-2 full-length TOFU-1 and TOFU-2 proteins was relatively complicated, therefore different approach was chosen. Instead of purification of separate proteins of PUCH complex from *E. coli*, BmN4 cells were transfected with plasmids encoding TOFU-1-3xFLAG-mCherry, eGFP-TOFU-2, or HA-SLFL-3/4 in different combinations. GFP-IP was performed on the transfected cells and IP-material on the beads was used in cleavage reactions with the m^{7}G-AAU precursor. Later I will write about cleavage reactions with recombinant PUCH complex obtained from *E. coli* (mini-PUCH), therefore I will refer to the PUCH complex obtained from BmN4 cells as IP-PUCH. In short, there was no difference between IP-PUCH or mini-PUCH detected. Cleavage reactions were resolved on 15% TBA-Urea gels. Synthetic RNA oligonucleotide with the sequence corresponding to the expected cleavage product and untreated capped product were used as molecular mass markers (Figure R18.B – first two lines). Cleavage of the 21U RNA precursor was observed only in samples, containing all three proteins – TOFU-1, TOFU-2, and SLFL-3 or SLFL-4 (Figure R18.B). As cleavage of 21U RNA precursor occurred in both samples containing SLFL-3 and SLFL-4, further experiments were performed using SLFL-3 only. Interestingly, an E216A mutation in TOFU-2 prevented the cleavage reaction completely, resembling the result observed *in vivo* (Figure R18.B, line 4).

4.8. PUCH activity requires the presence of Mg-, Mn- or Ca- divalent cation.

Cleavage buffers containing Mg^{2+} cations were chosen for the cleavage reactions, as characterized by SLFN-nucleases acting only in the presence of divalent cations (Jin-Yu Yang. 2018). After having established the enzymatic activity of IP-PUCH, the effect of the presence of different divalent cations was tested. IP-PUCH-containing beads were washed with EDTA and then resuspended in the buffer containing different cations (Mg^{2+}, Mn^{2+}, $Zn^{2+,}$ or Ca^{2+}) or none of them (-). Afterward, these beads were used in cleavage reactions. PUCH cleavage happened in the presence of Mg^{2+} at all tested concentrations. Besides that, cleavage in the presence of Mn^{2+} was observed. Interestingly, for Mn^{2+} lower concentrations showed better cleavage results than higher concentrations (Figures R19A and B). Ca^{2+} also supported cleavage, but only at high concentrations (Figures R19.A and B). There was no cleavage observed in the presence of Zn^{2+}, and likewise, precursor processing was inhibited, when no metal ions were provided at all, suggesting that the presence of divalent cations, notably Mg^{2+}or Mn^{2+}, is crucial for PUCH activity (Figures R19.A and B).

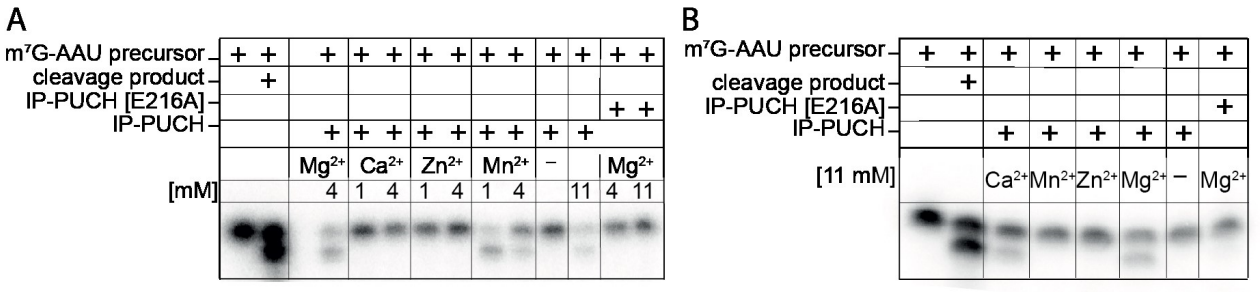

Figure R19. PUCH activity in the presence of different divalent cations. *In vitro* cleavage assay of the 21U RNA precursor using anti-GFP immunoprecipitated material from BmN4 cell extracts (IP-PUCH) in combination with different divalent cations in different concentrations, 1mM and 4mM (A) or 11mM (B).

4.9. PUCH is very specific and requires Uracil in position 3.

Almost all annotated 21U RNA precursor sequences contain Uracil at position 3, which represents the 5' end of the corresponding mature 21U RNA sequence (Figure R20.A). Therefore, it was reasonable to test if Uracil in position 3 at this position is a requirement for PUCH cleavage. Synthetic oligonucleotide, in which Uracil in position 3 was replaced with Cytosine: "m^7G-AAC" was ordered (Figure R20.B). This oligonucleotide also contained U in position 4, which stayed unchanged. The new precursor was radioactively labeled on the 3'end. A cleavage reaction was performed with new "m^7G-AAC" RNA with PUCH material, obtained from BmN4 cells. It turned

out, that the replacement of Uracil with Cytosine abolished the cleavage completely. Thus, Uracil in position three is a requirement for the PUCH-mediated cleavage of 21U RNA precursors.

Another substrate requirement that I tested was the length. Shortened to 10 nucleotides m^{7}G-AAU substrate was efficiently processed by mini-PUCH.

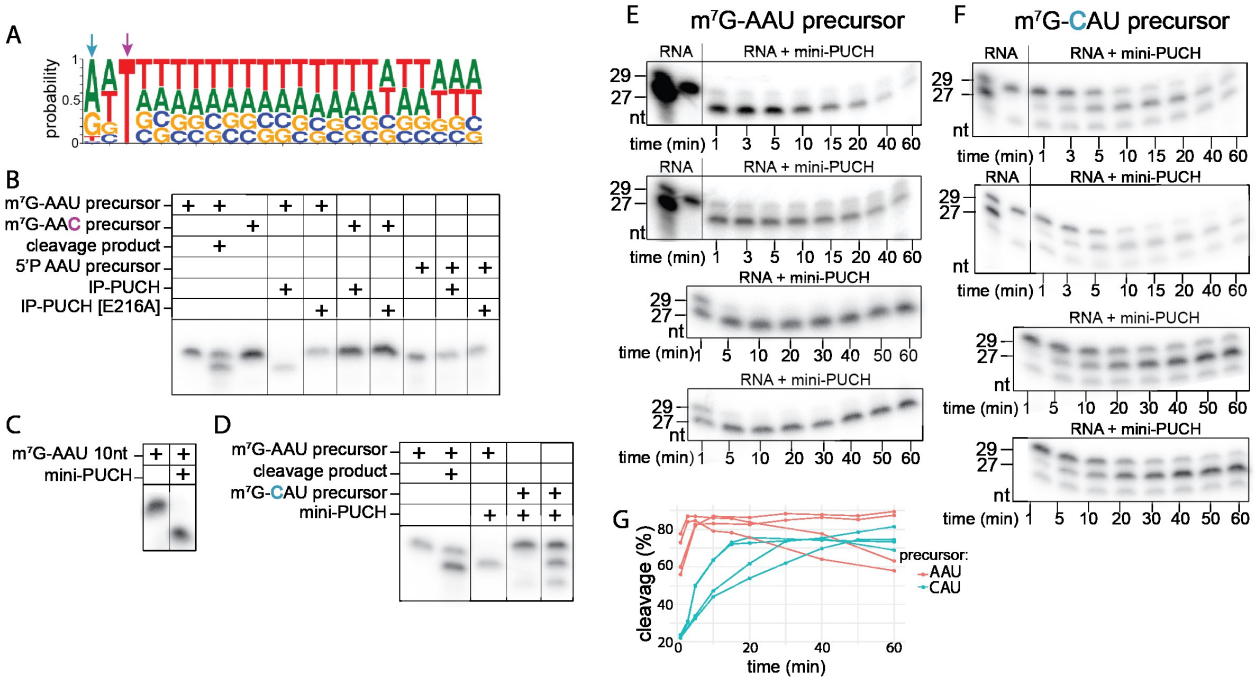

Figure R20. Substrate requirements of PUCH complex related to the RNA sequence. A. Sequence logo of piRNA precursors in *C. elegans*. T in position three of 21U RNA precursor represents U in position 1 of mature piRNA. **B**. *In vitro* cleavage assay with IP-PUCH and "m^{7}G-AAC" and 5'P-AAU precursors. **C**. *In vitro* cleavage assay with a short substrate – 10-nucleotide long m^{7}G-AAU substrate. **D**. *In vitro* cleavage assay with mini-PUCH and "m^{7}G-CAU". **E-F**. *In vitro* cleavage assay with the recombinant mini-PUCH and m^{7}G-AAU (**E**) and m^{7}G-AAU (**F**) substrate in time series. **G**. Quantifications of the signal, presented in **E** and **F**.

4.10. 21U RNA precursors starting with cytosine are processed less efficiently.

Less than 1% of 21U RNA precursors have C in position 1. To investigate the relevance of this position an oligonucleotide identical to the "m^7-AAU" precursor sequence was designed, in which the first adenine in "m^{7}G-AAU" was replaced with cytosine, resulting in "m^{7}G-CAU" substrate. Cleavage reactions were performed with mini-PUCH (see below). Surprisingly, in the first experiment only half of the m^{7}G-CAU substrate was cleaved during the usual reaction time, while all m^{7}G-AAU control substrate was completely cleaved (Figure R20.D). To investigate it further, cleavage time series were performed using mini-PUCH (Figure R20.E and F). The time

series confirmed, that cleavage of the m^{7}G-CAU substrate happens slower than with the canonical m^{7}G-AAU substrate (Figure R20.G).

4.11. PUCH cleavage requires a m7G-cap.

The vast majority of mRNAs in *C. elegans* undergo trans-splicing and therefore contain a 5' TMG cap. 21U RNA precursors are transcribed by RNA Pol II and carry an m^{7}G cap. The importance of the presence of the m^{7}G cap for the cleavage of precursors by PUCH was investigated. First, a cleavage reaction with an oligonucleotide containing a 5'phosphate group, "P-AAU" precursor, was performed. In order to obtain the "P-AAU" precursor, a synthetic oligonucleotide containing a 5'OH group with the sequence corresponding to the standard precursor used above was phosphorylated by T4 polynucleotide kinase using ^{32}P ATP. The product was purified on G25 columns and used in reactions with IP-PUCH (Figure R20.B). Cleavage of the "P-AAU" precursor was not observed.

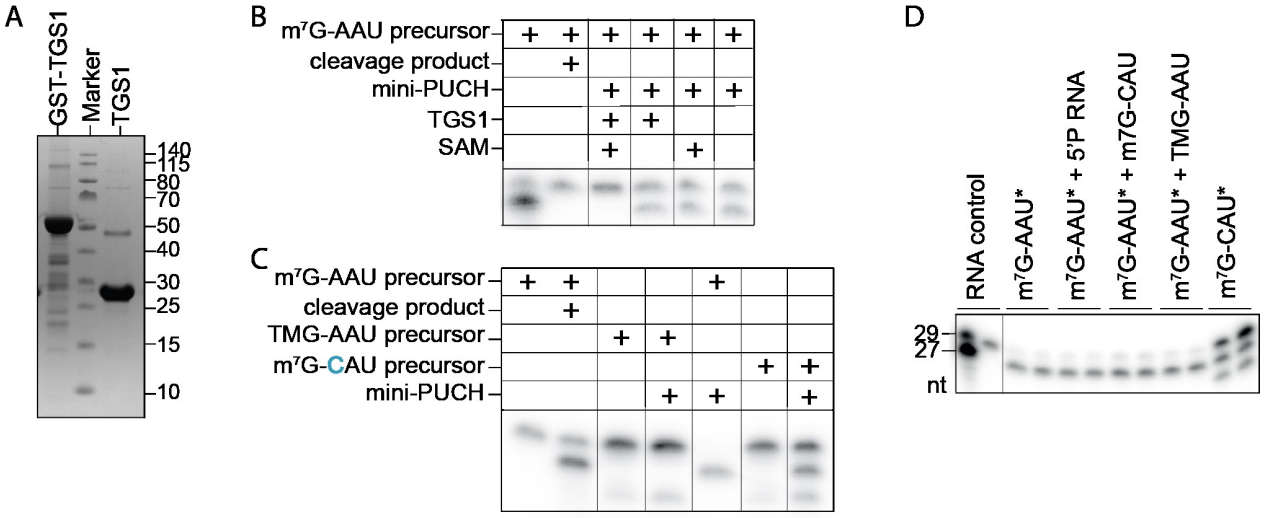

Figure R21. M^{7}G-cap is necessary for PUCH-mediated cleavage. A. Affinity- and SEC-purified GST-tagged TGS-1 protein before (left line) and after (right line) GST-cleavage. **B**. *In vitro* cleavage reaction with m^{7}G-AAU substrate, modified with TGS1 protein, purified in H, with the presence or absence of SAM. **C**. *In vitro* cleavage reaction between mini-PUCH and m^{7}G-CAU and TMG-AAU synthetic precursors. **D**. *In vitro* cleavage assay performed with mini-PUCH and m^{7}G-AAU substrate in the presence of the cold substrate, indicated on the panel and m^{7}G-CAU substrate as a control.

Next, I tested if the TMG-capped precursor, "TMG-AAU" can be processed by PUCH. To obtain a TMG-tagged precursor, the catalytic domain of the TGS1 protein was purified according to the protocol described in Thomas Monecke et al (Figure R21.A). TGS1 protein is shown to modify the m7G cap structure to TMG (Monecke et al. 2009).

M^{7}G-AAU RNA precursor was incubated with purified TGS-1 in the presence and absence of S-Adenosylmethionin (SAM), co-substrate, and source of the methyl groups to be added to the m^{7}G cap. Treated precursors were used in cleavage reactions with mini-PUCH, as well as untreated m^{7}G-AAU precursor. Only in samples containing both TMG and SAM, the substrate was not cleaved, suggesting that PUCH cannot process TMG-capped precursors. In the absence of TGS1 or SAM, the precursor substrate was cleaved to the same percentage as the untreated control precursor (Figure R21.B).

To confirm the result coming from the enzymatically modified substrate, a synthetic oligonucleotide, carrying a TMG cap with the same sequence as used in the m^{7}G-AAU substrate with mini-PUCH was obtained. Again, the TMG-AAU substrate was not cleaved (Figure R21.C), confirming the result obtained with the TGS-1 treated m^{7}G-AAU substrate. Thus, m^{7}G-cap is required for PUCH-mediated cleavage.

4.12. Cleavage assay with cold RNA competitors.

Precursors m^{7}G-CAU, m^{7}G-AAC, and TMG-AAU, synthesized by the Biosynthesis™ company, contained a contaminating band, which was running gel slightly below the size of the cleavage product (Figures R20.D,F and R21.C). This band stayed unchanged during the cleavage reaction and likely represents a 5' truncated oligo. To exclude, however, that this contaminant affects the cleavage reactions, cleavage of the canonical m^{7}G-AAU substrate that did not show such a contaminating band was performed in the presence of 2-fold access of cold m^{7}G-CAU, m^{7}G-AAC, or TMG-AAU RNA. m^{7}G-CAU substrate was used as a control for slow cleavage. None of the reactions were affected (Figure R21.D), demonstrating that the contaminating band does not interfere with PUCH activity.

4.13. Schlafen-domains of TOFU-1, TOFU-2 and SLFL-3 from a mini-PUCH complex

Our collaborators in Sebastian Falk's research group purified a so-called mini-PUCH complex from *E. coli*, containing only the SLFL-domains of TOFU-1 and TOFU-2 and SLFL-3 with truncated C-terminus, both in the active (with wild-type TOFU-2) and catalytically dead (TOFU-2[E216A]) versions. I tested the enzymatic activity of the mini-PUCH complex on the m^{7}G-AAU precursor and m^{7}G-AAC precursor. m^{7}G-AAU precursor was cleaved by mini-PUCH, while m^{7}G-

AAC precursor was not, confirming the selectivity of the full-length IP-PUCH. Mutation in the catalytic center of TOFU-2 in mini-PUCH prevented the m^{7}G-AAU precursor cleavage (Figure R22).

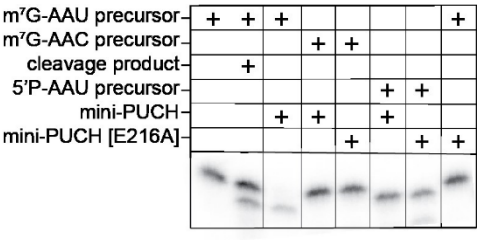

Figure R22. Mini-PUCH has the same specificity as full-length IP-PUCH. *In vitro* cleavage reaction with m^{7}G-AAU, m^{7}G-AAC, and 5'P-AUU substrates and recombinant mini-PUCH complex.

4.14. PUCH leaves a 5'P on the cleaved precursor.

Mature 21U RNAs carry a phosphate group on their 5'ends. Therefore the next experiment was designed to test if a 5'P was formed at the 5' ends of PUCH cleavage products. To do this, cleavage reactions with 3'end radioactively labeled m^{7}G-AAU precursor were performed, cleaved RNA was purified over the column, and ligated with 10-nt long synthetic RNA oligonucleotide, carrying 5'- and 3'-OH groups. Catalytically dead (with E216A mutation) version of mini-PUCH was used as a negative control. The reaction products were resolved on a gel, and even though the ligation efficiency was not very high, the band, corresponding to the ligation product could be clearly seen (Figure R23, marked with the error), indicating that a 5'P is formed upon PUCH cleavage.

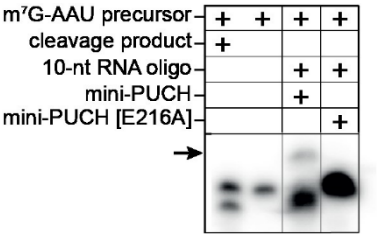

Figure R23. Mini-PUCH has the same specificity as full-length IP-PUCH. RNA obtained from a cleavage reaction (using either wild-type or TOFU-2(E216A)-mutant mini-PUCH) was ligated to a 10-nucleotide-long 5'OH-containing RNA adapter. The ligation product is indicated by an arrow.

4.15. The presence of PETISCO does not disturb the cleavage.

Current understanding of the 21U RNA precursor processing suggests that PETISCO serves as the platform for 21U RNA precursor cleavage. First, it was important to test that PETISCO (lacking PID-1 or TOST-1, because that could not be made) can bind 21U RNA precursors, and it was done by EMSA. M^{7}G-AAU precursor RNA was incubated together with PETISCO proteins (Perez-Borrajero et al. 2021) and resolved on the native acrylamide gel. PETISCO is able to bind RNA precursors containing m^{7}G cap better than precursors with identical sequence, containing 5'OH or 5'P (Figure R24.A). Next, the ability of PUCH to cleave 21U precursor in the presence of

PETISCO was tested. M^7G-AAU precursor was incubated in five molar access of PETISCO proteins, afterwards, cleavage assays in time series of both IP-PUCH (Figure R24.B) and mini-PUCH (Figure R24.C) were performed. There was no difference observed between cleavage reactions containing or lacking PETISCO, suggesting that the presence of PETISCO does not inhibit, nor stimulate the cleavage (Figures R24.D and E).

Next, I tested if PETISCO affects precursors *in vivo.* Worms, carrying a deletion allele of *pid-1(xf35)* were crossed with a *tofu-2(e216a)* strain, and double mutants were isolated. The aim was to test if the accumulation of precursors triggered by *tofu-2(e216a)* may be compromised by the loss of PID-1-PETISCO. Libraries were created for both mature piRNA and piRNA precursors from synchronized young adults. For both, *tofu-2(e216a)* single mutant and *tofu-2(e216a);pid-1(xf35)* double mutant the absence of mature piRNA was observed. As in the experiment before, in the *tofu-2(e216a)* mutant increased levels of precursors were detected, but in *tofu-2(e216a);pid-1(xf35)* double mutants the precursors were back to wild-type levels (Figure 24.F and G). The interpretation of this result could be that PETISCO together with PID-1 is needed to protect unprocessed 21U RNA precursors from degradation.

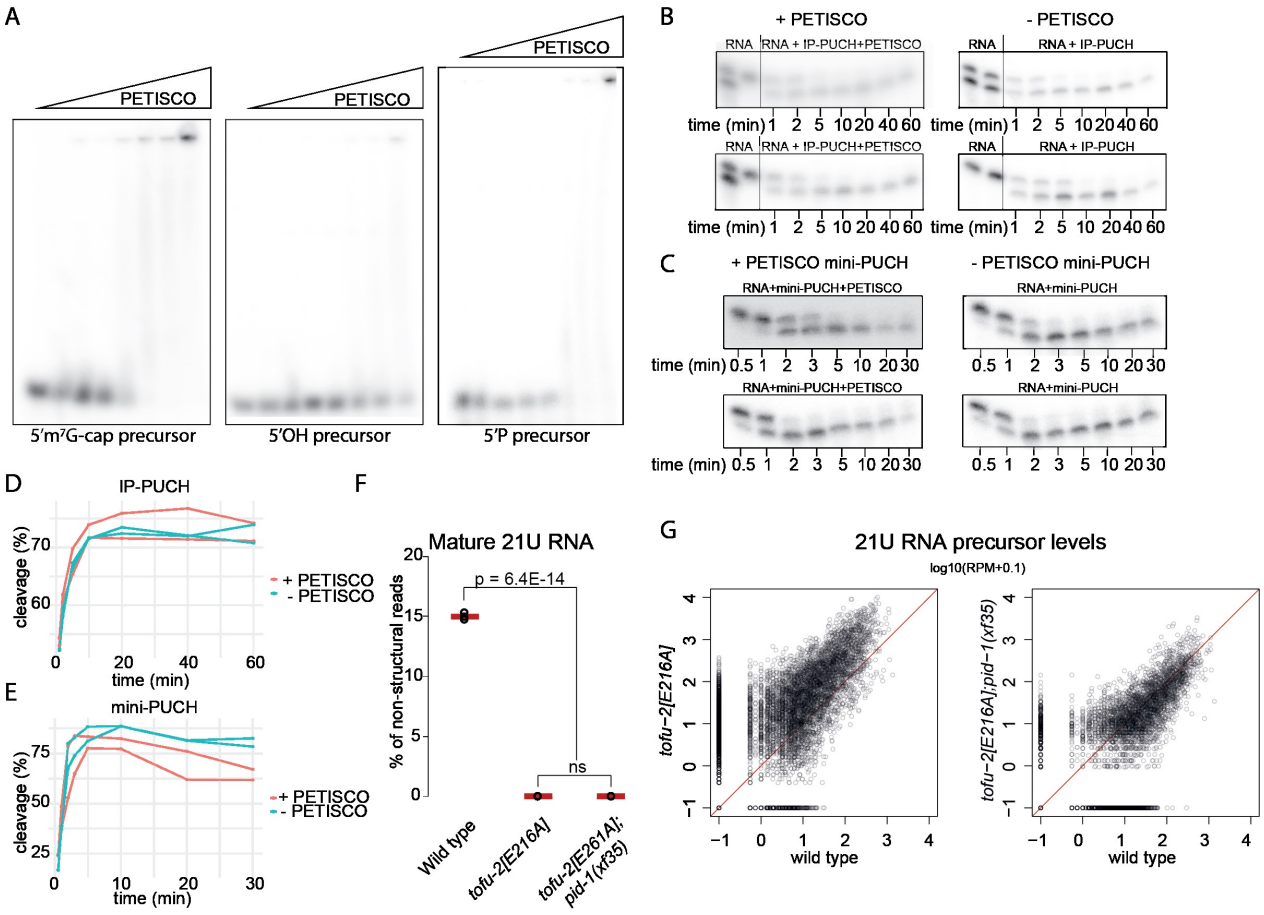

Figure R24. A. Electrophoretic mobility shift assay for PETISCO protein complex and three piRNA precursors m^{7}G-AAU on the left, 5′OH-AAU in the middle, and 5′P-AAU on the right. The concentration of PETISCO gradually increases from left to right. **B-C.** Time series of the cleavage reactions of m^{7}G-AAU substrate with and without PETISCO by IP-PUCH (**B**) and mini-PUCH(**C**). **D-E.** Quantification of the radioactive signal in panels B and C. **F.** Total mature piRNA levels in wild-type, *tofu-2*(*E216A*) mutant and *tofu-2*(*E216A*);*pid-1*(*xf35*) double-mutant young adult hermaphrodites. The experiment is done on three biological replicates. The red lines represent the group means. *P* values were calculated using one-way ANOVA followed by Tukey's HSD test; the indicated *P* value relates to both mutant samples. NS stands for non-significant. **G.** The relative abundance of individual 21U RNA precursors in *tofu-2*(*E216A*) mutant (left) and *tofu-2*(*E216A*);*pid-1*(*xf35*) double-mutant (right) versus wild-type young adult hermaphrodites. The experiment has been done in three biological replicates.

4.16. IP/MS experiments do not reveal the interaction between PETISCO and PUCH.

In TOFU-2 IP-MS experiments, there was no interaction detected between PETISCO and TOFU-2. Often, interactions with enzymes are too transient to be detected in co-precipitations. Therefore, an IP-MS experiment was performed in the strain that expresses catalytically inactive TOFU-2[E216A]:HA, as this mutation could stabilize the interaction between the two complexes. Unfortunately, there were no PETISCO proteins detected among the TOFU-2 interactors.

However, TOFU-1 was still enriched, meaning that the mutation in the catalytic center did not affect the complex formation (also the recombinant PUCH with E216A mutation did not show any signs of de-stabilization (Falk lab)). SLFL-3 and SLFL-4 were also not present among the interactions of TOFU-2[E216A]:HA (Figure R25.A). Most probably this happen due to the close similarity of SLFL-3 and SLFL, and therefore low number of unique peptides, which can be recognized by the algorithm.

Next, I crossed TOFU-6:3MYC with TOFU-2[E216A] expressing strains and performed an IP-MS experiment in which TOFU-6 interactors were compared in the wild-type TOFU-2 and TOFU-2 catalytic mutant backgrounds. There were no PUCH complex components identified in the obtained MS data; however, PID-1 was enriched in the E216A mutant TOFU-2 IP, suggesting that in the absence of TOFU-2 activity, accumulation of precursors may drive a bigger portion of PETISCO to be associated with PID-1, instead of with TOST-1 (Figure R25.B).

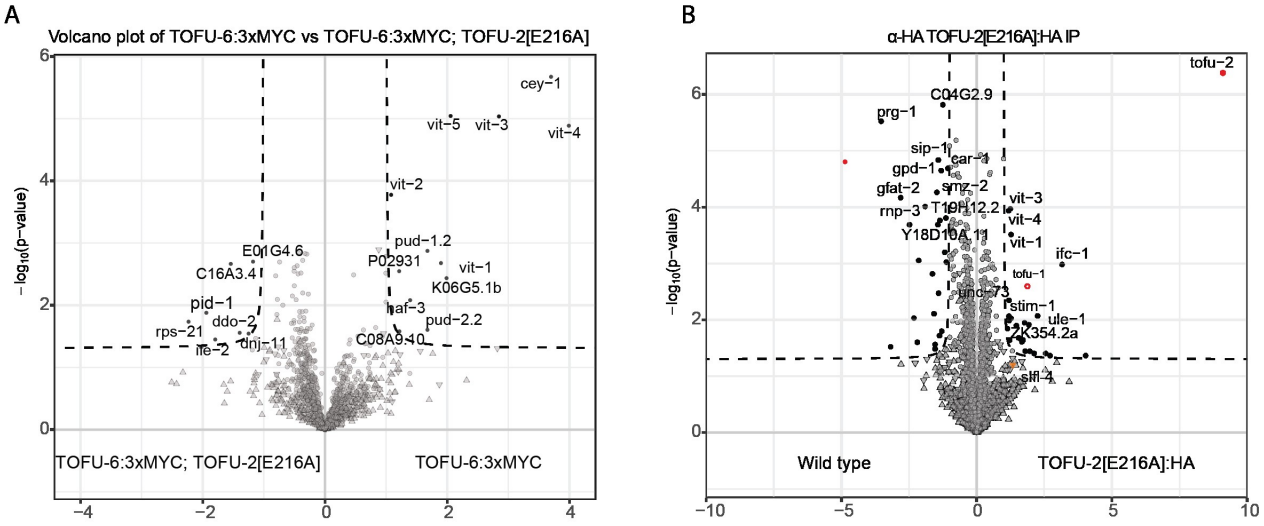

Figure 25. IP-MS experiments could not reveal the interaction between PUCH and PETISCO. Label-free proteomic quantification of TOFU-6:3xMYC and TOFU-6:3MYC in the TOFU-2[E216A] mutant background immunoprecipitates from young adult extracts. Results are obtained from four biological replicates. The *x-axis* shows the median fold enrichment of individual proteins, and the *y-axis* shows $-\log_{10}[P]$. *P* values were calculated using Welch's two-sided *t*-tests. The dashed lines represent enrichment thresholds at $P = 0.05$ and fold change > 2, curvature of enrichment threshold $c = 0.05$. The dots represent enriched (black) or quantified (grey) proteins. **B.** Label-free proteomic quantification of TOFU-2[E216A] and wild-type immunoprecipitates from young adult extracts. Results are obtained from four biological replicates. The *x-axis* shows the median fold enrichment of individual proteins, and the *y-axis* shows $-\log_{10}[P]$. *P* values were calculated using Welch's two-sided *t*-tests. The dashed lines represent enrichment

thresholds at $P = 0.05$ and fold change > 2, curvature of enrichment threshold $c = 0.05$. The dots represent enriched (black) or quantified (grey) proteins.

4.17. PETISCO and PUCH interact via TOFU-6 and TOFU-1 respectively.

As IP-MS experiments could not help to reveal the interaction between PETISCO and PUCH, an *in vitro* approach was chosen. In a series of pull-down experiments performed in the Falk laboratory, MBP pull-down for MBP-tagged TOFU-1 together with proteins of PETISCO-complex identified an interaction between TOFU-1 and PETISCO. Further investigation shows that TOFU-1 interacts with TOFU-6 via a small helix from position 82–113 in TOFU-1 and the eTUDOR domain in TOFU-6. The eTUDOR domain of TOFU-6 and the region 82–113 from TOFU-1 were crystallized together, leading to the finding that the TOFU-1 peptide binds TOFU-6 eTudor domain not at the canonical aromatic cage of the eTUDOR domain, but on the surface of the staphylococcal nuclease-like domain of the eTUDOR domain. This is the first time this region has been described to mediate protein–protein interactions (Podvalnaya et al. 2023). Based on pull-down experiments using mutated proteins, L88 and L92 of TOFU-1, and V266 of TOFU-6 were identified as important residues for the interaction between PETISCO and PUCH (Figure R26.A).

To confirm the role of the established PUCH-PETISCO interaction in 21U RNA precursor cleavage *in vivo*, mutations L88R and L92R were introduced into the *tofu-1* gene and V266E into the *tofu-6* gene in *tofu-6:3xmyc* carrying strain. These two strains were crossed and a double mutant was obtained. To test that the protein levels were not affected by the introduced mutations, tofu-1 was tagged endogenously with 3xmyc tag by CRISPR/CAS9 genome editing and mutation, disrupting the interaction between TOFU-6 and TOFU-1 was reintroduced. Western blot was performed with anti-myc and anti-H3 antibodies to confirm, that levels of TOFU-1 and TOFU-6 were unchanged (Figure R26.B). Next, small RNA sequencing for *tofu-1(L88R;L92R)* and *tofu-6(V266E)* mutants and *tofu-1(L88R;L92R);tofu-6(V266E)* double mutant was performed for both for mature 21U RNA and 21U RNA precursors. In all three strains, we observed a reduction of mature 21U RNA levels compared to the wild type (Figure R26.C). Additionally, in the double mutant, the accumulation of precursors was detected, confirming that the interaction between PETISCO and PUCH is important for 21U RNA precursor processing *in vivo* (Figure R26.D and E).

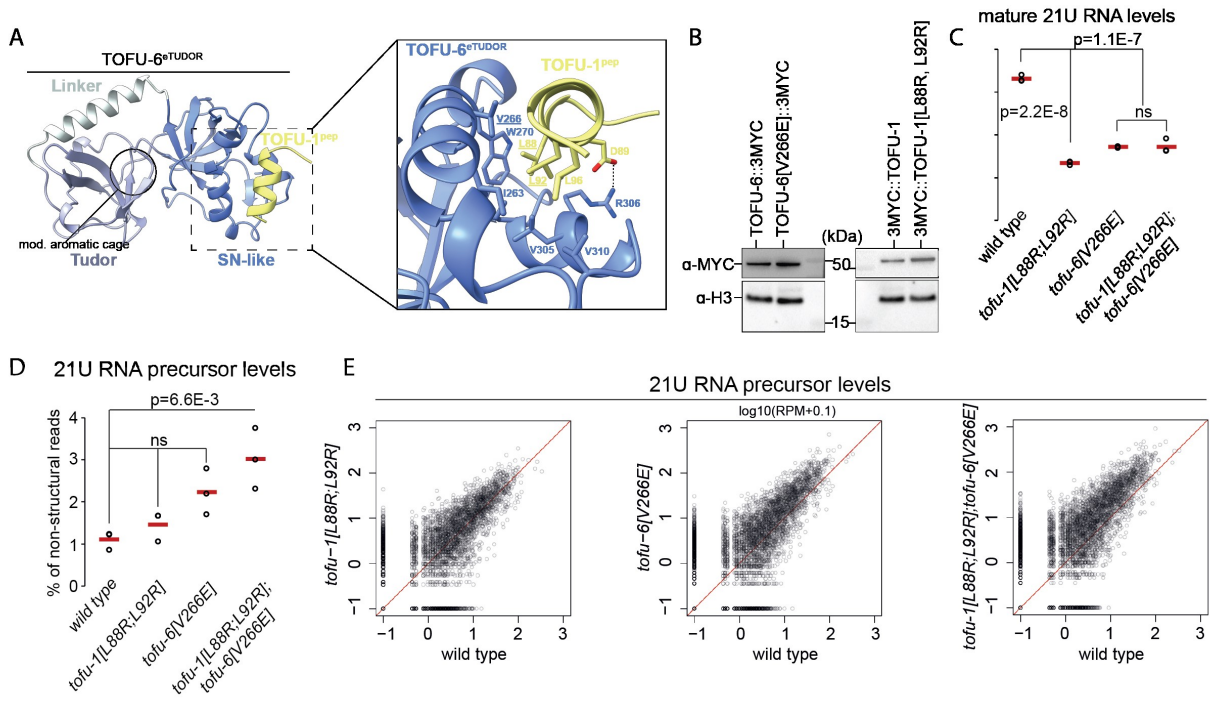

Figure R26. PUCH interacts with PETISCO through TOFU-1 and TOFU-6. A. The crystal structure of the TOFU-6^{eTUDOR}–TOFU-1^{pep} complex is shown as a cartoon. The TOFU-6^{eTUDOR} domain is shown in different shades of blue and TOFU-1^{pep} in yellow. The magnified view shows the interaction interface; involved residues are shown as sticks. **B.** Western blot analysis of the expression levels of TOFU-1 and TOFU-6 for the indicated genotypes using anti-MYC and anti-H3 antibodies, followed by visualization using horseradish-peroxidase-linked secondary antibodies. The numbers indicate the approximate molecular mass (kDa). **C-D.** The total number of mature (**C**) piRNA and precursors piRNA (**D**). The experiment is performed in three biological replicates. Red lines represent the mean of the sample. P-values are using one-way ANOVA followed by Tukey's HSD test. NS stands for non-significant. **E.** The relative abundance of precursors from individual piRNA loci in young adult hermaphrodites of the indicated genotypes. The experiment is performed in three biological replicates.

4.18. SLFL-3 and SLFL-4 have a transmembrane domain that is important for piRNA biogenesis.

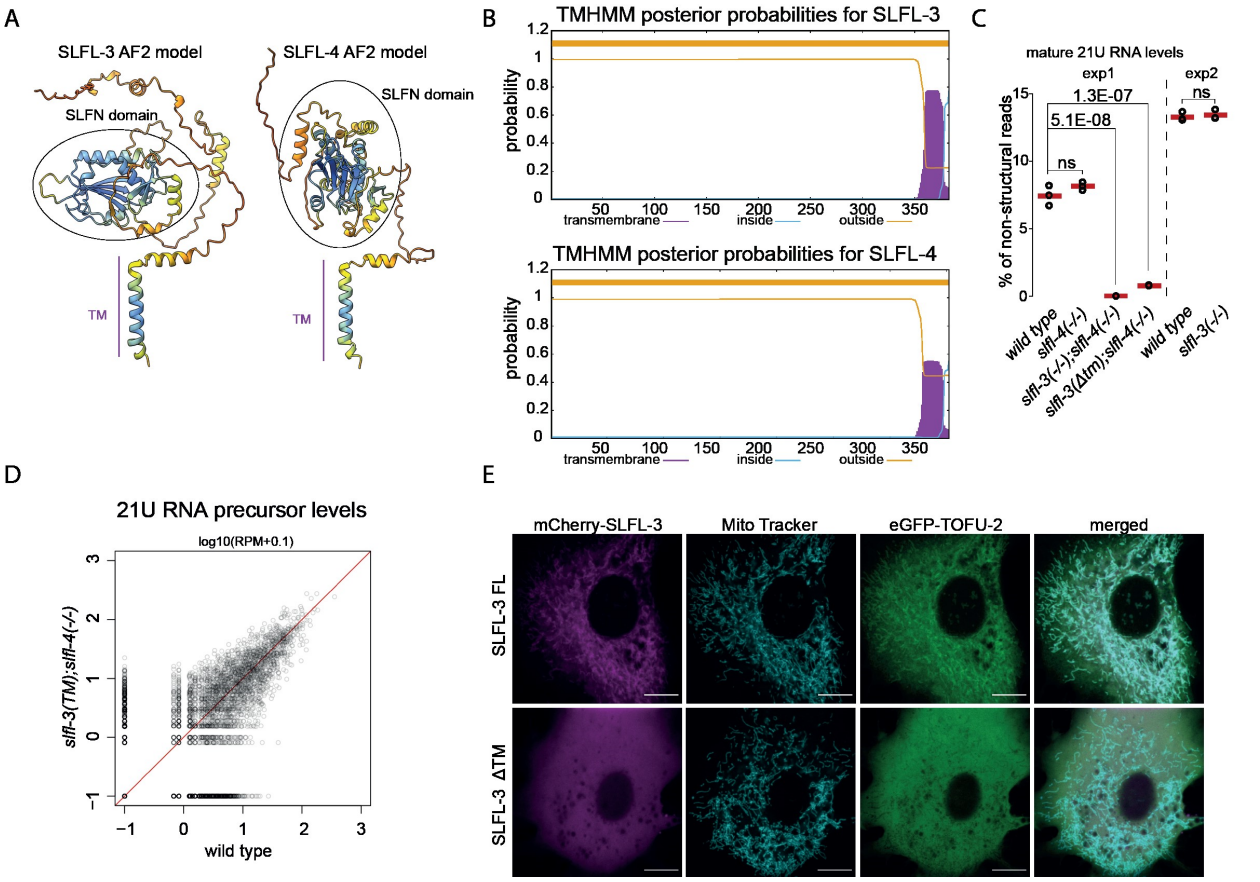

Figure R27. SLFL-3 and SLFL-4 contain transmembrane domains and localize on the mitochondria membrane. A. AlphaFold2 predicted structures of SLFL-3/4 shown as cartoon and colored by pLDDT score, which reports on the model confidence. Dark blue indicates very high, light blue confidence, yellow low, and orange very low model confidence. **B.** Prediction of transmembrane helices in SLFL-3 and SLFL-4 using TMHMM - 2.0.**C.** Total mature piRNA levels in young adult hermaphrodites of the indicated genotypes were done in three biological replicates. The red lines depict group means and *P* values were calculated using a one-way analysis of variance (ANOVA) followed by Tukey's honest significant difference (HSD) test (left) and two-tailed unpaired *t*-tests (right). The plot is based on two independent experiments (exp. 1 and 2). NS stands for non-significant. Underlying data are the same as in figure R7.F **D.** The relative abundance of piRNA precursors from individual loci in *slfl-3(Δtm);slfl-4(-/-)* mutant versus wild-type young adult hermaphrodites, performed in three biological replicates. **E**. Single-plane confocal micrographs of BmN4 cells that were transfected with eGFP–TOFU-2 and full-length mCherry–SLFL-3 (top) or mCherry–SLFL-3(ΔTM) (bottom). TOFU-1 was also transfected but was not tagged with a fluorescent protein. Mitochondria were stained with Mito Tracker. Scale bars, 10 µm.

Alphafold structure prediction for SLFL-3 and SLFL-4 showed the presence of a potential transmembrane helix (Figure R27.A). Both proteins were analyzed with TMHMM 2.0 predictor, and both SLFL-3 and SLFL-4 might contain a transmembrane domain (Figure R27.B). To investigate

if the potential TM helix in SLFL-3 is important for 21U RNA biogenesis we created a CRISPR/CAS9 deletion of the transmembrane domain of SLFL-3 in the SLFL-4 mutant background to prevent SLFL-4 compensating for the loss of TM domain in SLFL-3. Small RNA sequencing of mature 21U RNA and 21U RNA precursors was performed for *slfl-3(Δtm);slfl-4(-/-)*. It showed a 90% reduction in mature 21U RNA counts (Figure R27.C). Interestingly, the level of precursors was not increased (Figure R27.D).

To investigate the sub-cellular localization of the SLFL-3 we transfected BmN4 cells with plasmids encoding TOFU-1, eGFP-TOFU-2, and mCherry-SLFL-3 with or without transmembrane helix (ΔTM). Wild-type PUCH complex displayed a fiber-like localization pattern that is reminiscent of mitochondria. Indeed, the signal co-localized with a mitochondrial marker. TOFU-2 followed the localization of SLFL-3, consistent with them forming a complex. MCherry-SLFL-3 (and TOFU-2) signal was dispersed when the TM helix was deleted, confirming the role of the SLFL-3 TM helix in mitochondrial localization (Figure R27.E).

4.19. Evolutionary conservation of PUCH and PETISCO.

In collaboration with the laboratory of Peter Sakies, we looked at the evolutionary conservation of PUCH. PUCH shares the same conservation pattern as nematode-specific PETISCO proteins (Figure R28.A). Interestingly, the analysis did not identify a direct orthologue of TOFU-2 in *C. inopinata*, which was surprising as the rest of the piRNA pathway proteins were present. Using BLAST and HHpred we identified a 125 amino acids-long predicted protein Sp34_50366100.t1 with high sequence homology to the SLFN-fold of TOFU-2 with conserved catalytic center (Figure R28.B). We ran Alphafold predictions on Sp34_50366100.t and aligned it together with the SLFN domain of TOFU-2. There are some differences in the structural organization of SLFN-domains of TOFU-2 and TOFU-2-like *C. inopinata* proteins, however, the catalytic center and the area around are intact (Figure R28.C). Thus, the TOFU-2 analog in *C. inopinata* may contain an SLFN-fold that closely resembles that of TOFU-2, but does not contain the N-terminal SPRY domain; possibly, this short TOFU-2 protein is a result of the wrong genome annotation.

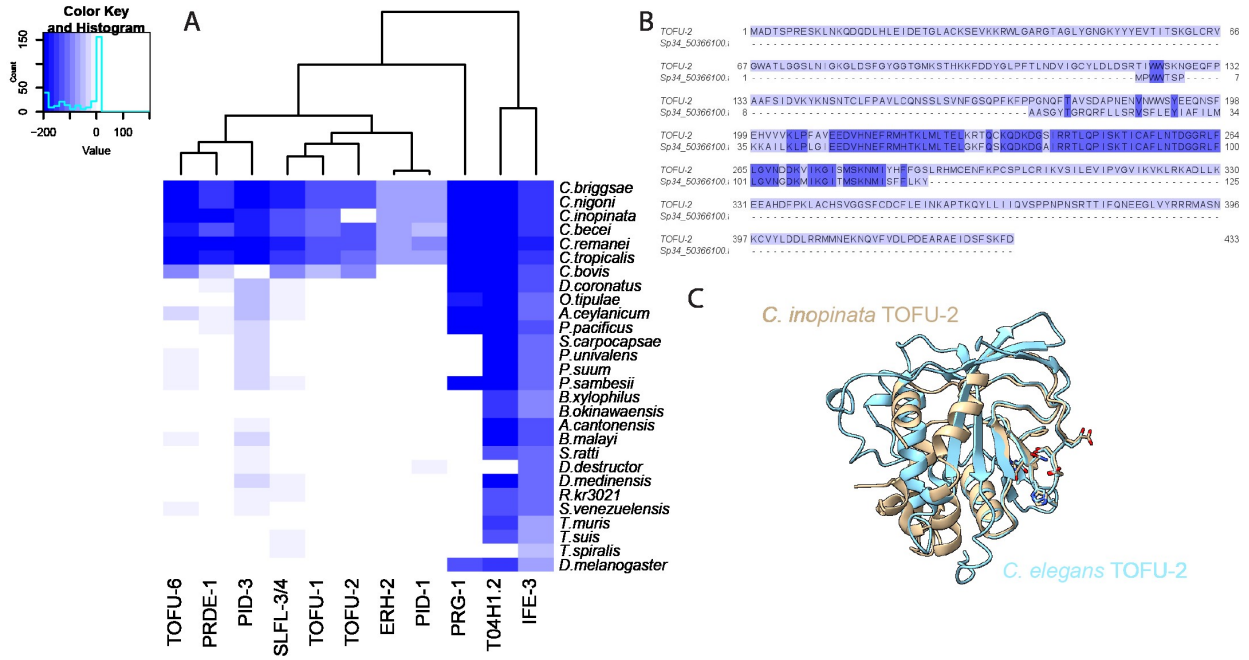

Figure R28. PUCH conservation resembles the conservation of the PETISCO complex. A. Conservation heat map for the PUCH proteins and PETISCO protein complex. The color scheme is blue to white where blue represents high conservation. A heat map is generated by Peter Saskies' lab. **B**. Sequence alignment of *C. elegans* TOFU-2 and TOFU-2-like protein of *C. inopinata*. **C**. Structure alignment of Alphafold predictions of *C. elegans* TOFU-2 and TOFU-2-like protein of *C. inopinata*.

5. Discussion

5.1. Structural and functional analysis of PETISCO.

Successful purification of PETISCO components from bacterial cells has allowed further studies of the PETISCO, which helped to lid the light on the complex organization and in the end, obtain the crystal structure of the PETISCO core.

In short, PETISCO consists of four proteins, IFE-3, TOFU-6, ERH-2, and PID-3, which form an octamer, through the dimerization of ERH-2 and the RRM domain of PID-3. TOFU-6 RRM domain interacts with the RRM domain of PID-3. TOFU-6 attracts IFE-3 to the complex through an eIF4E-interaction stretch of amino acids, as described in the Introduction and the Result sections. Below I am going to discuss the interaction between different PETISCO sub-units and the effect of these interactions on the complex localization and functioning.

5.1.1 Disruption of PID-3:TOFU-6 sub-complex leads to mislocalization of PETISCO complex.

Interestingly, *in vitro* experiments have shown, that dimerization of the PID-3 RRM domain is crucial for the PID-3/TOFU-6 interaction (Perez-Borrajero et al. 2021). Mutation A220E, which prevents the PID-3 dimerization *in vitro*, resulted in the loss of interaction with TOFU-6 (Perez-Borrajero et al. 2021). Introduction of an A220E mutation *in vivo* into a PID-3:GFP carrying strain caused a strong, but incomplete Mel phenotype at 20 °C. 100% of the embryos, obtained from PID-3(E220A):GFP worms kept at 25 °C did not develop, in other words showing a full Mel phenotype (Figure R.8). The A220E mutation did not change the amount of the GFP signal, compared to the wild type PID-3:GFP signal, implying that the Mel phenotype derived from the disruption of the interactions within PETISCO complex, and not due to the decrease of the amount of PID-3 abundance.

Wild-type PID-3:GFP localized into the P-granules, and a cytoplasmic dispersed signal is detected too (Figure R8., Rodrigues, 2019). Most interestingly, in PID-3(A220E):GFP expressing worms, localization of GFP signal into P-granules at 25 °C was not observed anymore, meaning that either PID-3 dimerization of its interaction with TOFU-6 is crucial for the P-granular localization of PID-3 and maybe for the rest of PETISCO as well.

In addition, the work of another group independently investigating the structure of PETISCO (PICS) complex, describing the behavior of a TOFU-6(D26A;W27A):GFP-expressing

transgene in a *tofu-6* mutant background (X. Wang, 2021). According to their *in vitro* experiments, these mutations were preventing TOFU-6/PID-3 interaction on the TOFU-6 side. They observed, that TOFU-6(D26A;W27A):GFP protein was not localized to the P-granules anymore. Localization of the PID-3:GFP transgene into the P-granules in the *tofu-6* deletion background was also shown to be affected, however, it may well be that this results stemmed from a significant decrease in PID-3 levels that was also apparent from the images. To investigate this effect more detailed studies of the endogenous mutant worms are required.

Together, localization results from our group and X.Wang publication suggest that neither PID-3 nor TOFU-6 alone can find their way to the P-granules, making the interaction between PID-3 and TOFU-6, and therefore also the dimerization of PID-3, crucial for the P-granular localization of PETISCO complex. Thus, I assume that PID-3/TOFU-6 together bring the rest of PETISCO complex to the P-granules, however, investigation of the PID-3:GFP localization in the *tofu-6(d26a;w27a)* mutant background is required.

In general, the role of the localization of PETISCO into P-granules needs to be investigated. It is not clear if P-granular localization is relevant for PETISCO functions. I would assume that P-granular localization of PETISCO is required for the loading of PETISCO with RNA molecules, the presence of loaded PRG-1 in P-granules may suggest, that localization of PETISCO to the P-granules is needed for the further transition of processed 21U RNA transcript from PETISCO to PRG-1.

5.1.2. Disruption of PID-3:ERH-2 complex causes a Mel phenotype and a decrease in 21U RNA levels.

Strains, endogenously expressing PID-3(V182A;V186A):GFP and PID-3(V182E;V186E):GFP were created. These mutations disrupted the interaction between ERH-2 and PID-3 *in vitro* (Perez-Borrajero et al. 2021). *In vivo*, we observed both PETISCO phenotypes, Mel and 21U RNA deficiency (Figure R9) for the strain carrying analogous mutations, suggesting that interaction with ERH-2 *in vivo* was also compromised by these mutations. Replacement V182 and V186 to alanine had a milder effect *in vivo*. Small RNA sequencing revealed depletion of 21U RNAs in *pid-3(v182a;v186a):gfp* strain, which was grown at 25 °C. Thus, disruption of the interaction between ERH-2 and PID-3 affects PETISCO complex functions.

In contrast to the Wang et al paper, in which disruption of the interaction between ERH-2 and PID-3 by mutations of the same residues in PID-3 to aspartic acid caused elevation of the PID-3:GFP (PICS) signal in P-granules and enlargement of PID-3 granules, in our study such an effect was not observed. I believe, that the absence of this effect can be explained by the lower expression level of the endogenous PID-3:GFP compared to the transgenic PID-3:GFP used by the other group. Another possible explanation for the difference in the granule size between the two experiments could be the milder effect of the alanine mutation, which I used for imaging. However, in my experiments, the animals were kept and imaged at 25 °C, and at this temperature, a complete Mel phenotype was already observed. Thus, I assume that the main reason why we do not see the enlargement of the PID-3:GFP granules in the absence of ERH-2 is the PID-3:GFP expression level.

Besides the enlargement of PID-3 granules, changes in the localization of ERH-2 in the absence of PID-3 were described (Wang et al. 2021). Instead of the combination of the cytoplasmic and P-granular localization, in the absence of PID-3, ERH-2 was found inside the nucleus. *C. elegans* contains two proteins, belonging to the ERH family, and the second ERH protein ERH-1 localizes to the nucleus. ERH-2 and ERH-1 proteins share high similarity and most probably in the absence of the direct interactor PID-3, ERH-2 is competing with ERH-1 for its interactors. Localization of ERH-2 to the nucleus when it does not bind PID-3 is another piece of evidence for the idea that the PID-3/TOFU-6 hetero-tetramer brings the rest of PETISCO to the P granules.

5.1.3. TOST-1 and PID-1 share the same interaction surface on ERH-2.

Purifications of the full-length PID-1 and TOST-1 proteins were not successful enough to use the protein in the crystallization or NMR experiments, due to the fast degradation of both proteins, most probably because of their disordered nature. To study the interaction between ERH-2 and TOST-1/PID-1 short synthetic peptides were designed. A sequence of these peptides covers the conserved area between TOST-1 and PID-1, the anticipated position where interaction with ERH-2 takes place (Figure R10).

Protein NMR analysis, performed on ^{13}C and ^{15}N double-labeled ERH-2 protein and the PID-1 or TOST-1 peptides resulted in similar chemical shifts of the same ERH-2 residues,

suggesting that both peptides occupy the same binding surface on ERH-2. PID-3 was binding ERH-2 at different positions, opposite of the PID-1/TOST-1 shared surface.

According to ITC measurements, performed on purified ERH-2, mixed with TOST-1 and PID-1 peptides showed that ERH-2 has more affinity towards ERH-2 than PID-1. We still do not know the exact expression pattern of TOST-1 and PID-1, as these proteins have not been yet endogenously tagged with a fluorescent protein. Experiments on transgenic worm lines, expressing GFP-tagged TOST-1 and PID-1, described by the Guang group (Zeng et al. 2019), suggest that PID-1 is preferably expressed in the distal part of the gonad, in the mitotic region; while TOST-1 is expressed closer to the meiotic part of the gonad and in the embryos. It is possible, that upon increased expression of TOST-1 in the later part of the gonad, TOST-1 outcompetes PID-1 from PETISCO complex and takes its spot, and this is how two functions of the complex are separated in space and between different stages of germ cell development. However, more studies with endogenously tagged proteins are required to fully resolve this aspect and to address its relevance.

5.1.3.1. How do TOST-1 and PID-1 control the fate of PETISCO?

ERH-2 is an evolutionarily conserved protein, present in many different organisms, and it seems to be a scaffold for protein complexes, involved in RNA metabolism (Kozlowski 2023). ERH-2 interacts with TOST-1 and PID-1 on the opposite side of interaction with PID-3. Interactions, similar to one with PID-3 are known for ERH protein family, for instance in fusion yeast protein Mmi1 Interacts with ERH in a similar way (Xie et al. 2019). Interaction, similar to the one between ERH-2 and PID-1/TOST-1 was observed between human ERH protein and POLDIP3, Polymerase delta-interacting protein 3 (Kozlowski 2023). POLDIP3 was also identified in my IP-MS experiment, performed with the GFP-IP of the overexpressed ERH-mTurquiose2 from HEK293T cells (data not shown). However, as POLDIP3 is a relatively big protein, compared to TOST-1 and PID-1, I did not follow it up. Still, the exact role of the interaction between PID-1/TOST-1 and PETISCO remains unclear as these two proteins do not have direct homologous. According to the results above and published data on PETISCO, PETISCO bound by TOST-1 or PETISCO bound by PID-1 have distinct biological functions, but how exactly the interactor determines the fate of the PETISCO complex is a mystery. Again, both TOST-1 and PID-1 are very unstable and disordered proteins, and it makes it hard to perform *in vitro* studies with full-length proteins. Nevertheless, I have several

thoughts on how interaction with PID-1 or TOST-1 can change PETISCO function I would like to share.

First, I have been able to tag neither TOST-1 nor PID-1 endogenously with a fluorescent tag yet, mostly due to the very low GC-content in the gene flanking area, and therefore the expression pattern of PID-1 and TOST-1 remains unknown. Zeng 2019 et al. claim, that TOST-1 and PID-1 transgenes tend to localize in the different parts of the gonad: PID-1 in the mitotic area, while TOST-1 is expressed mostly in the meiotic region and embryos (Zeng et al. 2019). It is possible, however unlikely, that it does not matter which protein is binding to ERH-2, it is just important that ERH-2 has a binding partner. If it is true, then substitution of the *tost-1* gene sequence for *pid-1,* with preservation of UTRs of *tost-1* in the *tost-1* endogenous loci can rescue Mel phenotype caused by the absence of TOST-1. I did not get to test this idea yet.

But what exactly do these two proteins do? In my opinion, the most promising idea is that TOST-1 or PID-1 changes the conformation of PETISCO complex, allowing the switch between binding of the different RNA substrates. PETISCO has multiple predicted RNA binding sites, and maybe some of them are not accessible in PID-1 or TOST-1 bound PETISCO. A comparison of the PETISCO-bound RNA species, obtained from TOST-1 and PID-1 IP would be informative to see if there are any differences in RNA binding. Obtaining crystal structures of PETISCO in the presence of TOST-1/PID-1 is another possibility to address this question.

We were not able to identify any RNA binding motifs within TOST-1 and PID-1. Several RNA-binding strategies have been described for the disordered proteins, including π-π stacking between aromatic rings containing amino acids located close together and RNA bases, and simple electrostatic interactions (Zeke et al. 2022). I was not able to identify such described RNA-interacting motifs in TOST-1 or PID-1. Nevertheless, I cannot exclude, that TOST-1 and PID-1 bind RNA directly.

If not direct RNA binding, TOST-1/PID-1 could attract other proteins, possibly loaded with RNA, towards PETISCO. A direct comparison of IP/MS results, performed on TOST-1 and PID-1 might be informative to address this. However, such interactions could be transient, and hard to detect.

According to the localization studies described by Wang et al., and Peretz-Borejero et al. in 2021 (Wang et al. 2021; Perez-Borrajero et al. 2021), TOST-1 and PID-1 do not affect the

localization of PETISCO complex. However, this conclusion needs to be confirmed by experiments using endogenously expressed proteins.

Thus, the exact role of TOST and PID-1 in PETISCO remains unclear. To arrive at a better understanding, a better description of RNA substrates bound to PETISCO will be required including experiments to understand how RNA substrates arrive to bind PETISCO. Protein and RNA localization studies will be needed for this.

In another project, conducted by another PhD student in our group, was found, that the Mel phenotype of PETISCO and TOST-1 may be a result of the depletion of histone mRNA in the PETISCO mutants. It seems very likely that PETISCO is needed to provide stability for the histone mRNA transcripts, which are not poly-adenylated and not trans-spliced, and in this way resemble 21U RNA precursors. However, why TOST-1 is needed for histone mRNA binding by PETISCO is still unclear.

5.2. PUCH is a novel 21U RNA processing nuclease.

In this part of the Discussion, I will describe the newly identified nuclease PUCH, which is processing 21U RNA precursors in *C. elegans*. PUCH is formed by three proteins – TOFU-1, TOFU-2, and SLFL-3 or SLFL-4. SLFL-3 and SLFL-4 are very similar proteins that act redundantly in the PUCH complex. I will talk about the cleavage specificity of the identified complex, reasons why PUCH was not discovered before, the role of PUCH localization, and the possibility of the presence of similar nucleases in other organisms. Besides that, I will talk about PUCH's relations with PETISCO complex and 21U RNA processing pathway in *C. elegans*.

5.2.1. PUCH complex specificity.

21U RNA precursors contain m^7G cap and two nucleotides which are removed by PUCH. Even though we have not detected the 5'end cleavage product of the synthetic precursor, I believe that the processing happens as a single cleavage event, by direct removal of the first two nucleotides and the cap, as after removal of the cap RNA stops being a PUCH substrate. Uncontrolled capped removal could be fatal for the organism, therefore I assume that this should be strictly regulated.

5.2.1.1. Precursor requirements for PUCH-mediated cleavage.

21U RNA precursors do not share sequence similarities, except the positions 1, 2, and 3; for the rest of the precursor, the nucleotide distribution is close to random in every position. Almost 100% of 21U RNA precursors contain U in position three, and this Uracil is crucial for PUCH-mediated cleavage. Synthetic 21U RNA precursor, where this Uracil in position three was mutated to Cytidine was not processed by PUCH (Figure R20.B). Due to the high cost of the capped precursor synthesis, it was not possible to test precursors containing G and A in position three, and we do not have direct evidence that only precursors, which contain U are processed by PUCH. However, as all mature 21U RNAs start from U, I would assume that having U in position three is a strict requirement for a PUCH-mediated cleavage.

The vast majority of 21U RNA precursors have Adenine or Guanine in position one, and less than one present of precursors contains Cytosine. This pattern could be explained by both transcription requirements and PUCH cleavage preferences. We tested the substrate, containing C in position one of the 21U RNA precursor, referred to as m^7G-CAU precursor (Figure R20.E and

R20.F). Surprisingly, PUCH processes the original m^7G-AAU precursor much faster, than the m^7G-CAU precursor.

The biological relevance of the difference in speed of processing between m^7G-AAU and m^7G-CAU precursors still needs to be investigated. As mentioned above, production of the capped RNA sequences is very expensive, and testing of all possible combinations of nucleotides was not achievable. Thus, for further studies of the PUCH preferences, randomly synthesized capped RNA oligonucleotides could be obtained, cleaved, and sequenced to conclude PUCH preferences.

Besides the precursor sequence requirements, we found that only m^7G-cap-carrying precursors can be processed by PUCH. RNA, containing 5'P or TMG cap stayed intact after the incubation with PUCH (Figure R20.B). This cap specificity probably evolved to prevent the processing of regular mRNA in *C. elegans*.

Such strong substrate restrictions are required possibly to prevent unwanted PUCH activity. The current understanding of 21U RNA precursor biogenesis suggests that PETISCO complex serves as a platform for precursor cleavage. As mentioned above, PETISCO complex combines two functions, and most likely, PETISCO binds other kinds of RNA. One of the phenotypes caused by the absence of IFE-3, a cap-binding protein in PETISCO, is Mel, suggesting that this protein is also required for the non-21U RNA-related PETISCO function. IFE-3 selectively binds m^7G (Keiper et al. 2000), thus other PETISCO-bound RNA may carry m^7G cap as well. Therefore, a strict requirement for U in position three might be necessary to protect other kinds of RNA from unwanted nuclease activity of PUCH. Indeed, histone transcripts do not have a U at position 3 (data belonging to another project).

5.2.1.2. Other substrates, processed by PUCH.

Besides canonical 21U RNA, we found that PUCH processes type II 21U RNA as well. In contrast to usual type I 21U RNAs, which are transcribed from short clustered individual genes, 21U RNAs type II are formed as a result of the aborted transcription of the gene sequences by RNA Pol II. Some of those transcripts contain uracil in position three as well, and those transcripts are processed by PUCH (Podvalnaya et al. 2023). The function of those transcripts remains unknown. In theory, some of the 21U RNA type II might be involved in the regulation of transcription of the genes they derived from, but most of them are probably just an accidental substrate for PUCH, as 21U RNA type II are much less abundant than other kinds of 21U RNA.

Another interesting point is whether the delivery of 21U RNA type II precursors happens via the same mechanism as for the 21U RNA precursors type I.

Yet another, fully untested idea for additional PUCH substrates is the following: RNA Modification m^7G can occur not only as a cap structure but also internally in the RNA of RNA (Li-Sheng Zhang, 2022). Interestingly, one of the most common locations for m^7G modifications is in the position 46 variable loop of tRNA (Tomikawa 2018). m^7G46 is introduced by METTL1/WRD4 machinery, and the impairment of m^7G46 modification of tRNA causes microcephalic primordial dwarfism (Shaheen et al. 2015). Misregulation of m^7G affects the differentiation and self-renewal of embryonic stem cells (Lin, S.; 2018; Deng, Y, 2020) and promotes the development of different types of cancer (Orellana et al. 2021). Interestingly, long mammalian Schlafen proteins are proven to cleave tRNA. There is no evidence, that m^7G modification is anyhow affecting Schlafen-mediated cleavage

In the experiment, performed to identify possible modifications on the 21U RNA precursors, an increase of the m^7G modifications was observed in the *tofu-2(e216a)* mutant, both in total RNA and IP RNA samples. It is not clear if this difference could be explained by 21U RNA precursors only. Despite the differences between internal m^7G modifications and cap structure, which is connected to the 5' end of RNA via the PPP bridge, it is possible, that PUCH could process substrates, containing internal m^7G-NNU sequences.

5.2.2. SLFL-3 and SLFL-4 act redundantly.

SLFL-3 and SLFL-4 have 90% sequence homology and, likely, they act completely redundantly. 21U RNA precursors are cleaved *in vitro* both by PUCH containing SLFL-3 and SLFL-4, deletion of only SLFL-3 or SLFL-4 does not cause a major effect on 21U RNA levels (Figure R17). The cross of *slfl-3(-/-)* deletion causes the temporal activation of the 21U RNA sensor which lasted for several generations (Figure R17), which was reversed either by compensation of SLFL-4 or by activation of the RNAe effect.

Until now we were not able to tag SLFL-3 and SLFL-4 with fluorescent tags in *vivo*, therefore we cannot compare if the localization, expression pattern, and abundance of these two protein is identical. Genes, encoding SLFL-3 and SLFL-4 are located on the different chromosomes and their expression patterns could be different. I assume, that SLFL-3 protein is most expressed in this couple. SLFL-3 and not SLFL-4 was identified during the RNAi and genetic screens. The worm

strain, carrying the deletion of the *slfl-3* gene, crossed to the 21U RNA sensor, and caused expression of mCherry for some generations, while the strain, carrying the deletion of *slfl-4* did not show activation of the sensor (data not shown). However, it is possible that in wild-type conditions, these two proteins are expressed during different developmental stages or have different preferences for different substrates, but it all still has to be researched. Differences in the expression between *slfl-3* and *slfl-4* genes could be identified by qPCR. In addition, analysis of the *slfl-4* expression levels within several generations of *slfl-3* deletion could give an answer if the 21U RNA sensor becomes inactive due to the increasing expression of *slfl-4* or due to the activation of the RNAe effect.

5.2.3. The transmembrane domain of SLFL-3 and SLFL-4 proteins leads to mitochondrial localization.

TMHMM2.0 analyzer predicted the presence of a transmembrane helix in SLFL-3 and SLFL-4. Interestingly, the newer version of this software, DeepTMHMM, did not confirm this prediction. I have not been able to add a fluorescent tag coding sequence to *slfl-3* and *slfl-4* endogenously yet; thus, the localization of SLFL-3 and SLFL-4 proteins in worms still needs to be investigated. In the meantime, to obtain preliminary data on the localization of SLFL-3 and SLFL-4 proteins, a study conducted on BmN4 cells, showed that PUCH, containing wild-type SLFL-3 was colocalizing with MitoTracker, while SLFL-3 lacking the transmembrane helix lost localization in mitochondria. TOFU-2 followed the localization of SLFL-3 in both full-length and truncated versions.

Even though there is no direct information on SLFL-3 and SLFL-4 localization obtained in worms yet, it seems very probable that in *C. elegans* SLFL-3 localized indeed to the mitochondria. Interestingly, *Drosophila* and mammalian piRNA processing nuclease Zucchini (Zuc) had been shown to localize to mitochondria as well (Saito et al. 2010). Zucchini's nuclease activity is not Schlafen-based and belongs to the phospholipase D family (Pane et al. 2007). In contrast to PUCH, Zuc has only a mild preference towards generating cleavage products containing 5'U nucleotide, and only 5'U-containing sequences are then utilized by PIWI (Nishimasu et al. 2012). PUCH and Zuc have different evolutional origins, but still, both piRNA biogenesis factors appear to localize on the mitochondria, suggesting the general importance of mitochondrial localization for piRNA biogenesis factors.

5.2.4. Mitochondrial localization is crucial for piRNA biogenesis.

Deletion of *slfl-3* or *slfl-4* alone did not affect 21U RNA levels, suggesting that these two proteins compensate for the loss of each other. To avoid the compensation for the loss of the TM domain in SLFL-3 by SLFL-4, the TM-deletion in *slfl-3* gene was introduced to the *slfl-4(-/-)* mutant. Small RNA sequencing of the *slfl-3(Δtm);slfl-4(-/-)* strain revealed that the transmembrane domain is important for the 21U RNA pathway in worms: The total amount of 21U RNAs in the TM mutant was decreased to about 10% from the amount of mature 21U RNAs in the wild-type worms. I need to admit, that the tagged version of SLFL-3 or SLFL-4 was not yet obtained, and the statement on the stability of SLFL-3(ΔTM) protein cannot be made yet. *In vitro*, the loss of the TM domain neither destabilized the SLFL-3 protein nor prevented its purification. I assume that the reduction in 21U RNA level is happening because of the loss of mitochondrial localization, rather than the reduction in the protein level. However, this still needs to be shown.

Interestingly, the level of precursors was not elevated, in contrast to the *tofu-2(e216a)* or *slfl-3(-/-);slfl-4(-/-)* mutants. Possibly, despite the wrong localization of SLFL-3 protein in the absence of the TM domain, 21U RNA precursor still was happening, preventing the accumulation of the precursors. Possibly precursor processing is not affected, but loading to PRG-1 of processed 21U RNA transcripts is impaired when PUCH is not localizing to mitochondria. Either way, mitochondrial localization seems to be an important location for piRNA biogenesis factors. I cannot exclude the existence of a feedback loop that increases transcription levels of 21U RNA precursors in the absence of mature 21U RNA, but I do not think that it is very likely.

There may be various reasons why 21U RNA processing is happening at the mitochondrial surface. Mitochondria are the energy stations of the cells, and the temperature in mitochondria is higher compared to the surroundings. Analyzing the precursor reads, originating from the *slfl-3(Δtm);slfl-4(-/-)* mutant worms showed an increased amount of precursors containing "suboptimal sequence", with a C in positions one or two, compared to the worms carrying *tofu-2(e216a)* mutation and wild type worms, suggesting that the processing of the precursors, containing "less popular" bases in positions 1 and 2 was affected more strongly than the processing of optimal precursors (data not shown). Thus, it is possible that the reaction, happening close to mitochondria, is simply faster, and allows for more efficient processing of sub-optimal precursors when it happens close to mitochondria. Another possibility is that the surface

of mitochondria serves as a meeting point for additional 21U RNA biogenesis factors, allowing easy transfer of the 21U RNA transcripts from precursors to mature 21U RNA between different 21U RNA pathway players. Clearly, the exact role of this localization needs to be further investigated.

5.2.5. Why was the PUCH nuclease not discovered before?

21U RNA, a class of small RNAs in *C. elegans*, was identified in 2006 (Ruby et al. 2006). 21U RNA pathway is well described, and PUCH was the last major unknown factor in 21U RNA biogenesis. In contrast, the piRNA processing enzyme Zucchini in *Drosophila melanogaster* was described as a putative nuclease in 2007 (Pane et al. 2007). Thus, the question is why did it take so long for *C. elegans*? In my opinion, there are three reasons for that.

5.2.6.1. Substrate limitations.

First, but not the most obvious reason is that PUCH-mediated cleavage is very specific. To achieve cleavage *in vitro*, PUCH needs to have a capped substrate, containing U in position three. Therefore, obtaining a suitable substrate is a hard task. Standard *in vitro* transcription does not allow obtaining a suitable amount of RNA with U in position three, due to the minimal promoter sequence. Suboptimal *in vitro* transcription from modified SP6 promotor can be achieved, but the production of the short transcripts is already a reason to apply troubleshooting. In other words, obtaining a clean and clear substrate by *in vitro* transcription is close to impossible. The addition of the cap to the synthetic RNA by chemical synthesis became available very recently (Abe et al. 2022; Muttach et al. 2017), and ordering capped substrate is expensive, and commercially performed by two companies to my knowledge.

5.2.6.2. The function of the Schlafen domains was assumed wrongly.

The second reason is, that although Schlafen-domain-containing proteins have been known for about 30 years (Frickey and Lupas 2004; Lupas and Martin 2002), their nuclease function was discovered only in 2018. Until then, Schlafen domain-containing proteins were mistakenly assumed to have an ATPase activity, based on homology with AAA-domain-containing proteins (Frickey and Lupas 2004; Lupas and Martin 2002). Low amino acid sequence homology between different Schlafen proteins prevented the initial recognition of the SLFN-folds in TOFU-2 and TOFU-1 based on the sequence alone. Thus, the resolved crystal structure of Schlafen13

(Yang et al. 2018) was a critical key to the identification of SLFN-folds in TOFU-1 and TOFU-2 by structure-based analyzers HHpred or more recent structure predictors like AlphaFold or SwissModel. Thus, even though the initial description of the 21U RNA phenotype of *tofu-1* and *tofu-2* mutants matched with them acting at the 5' cleavage step, the lack of good structural information prevented their identification as nucleases.

5.2.6.3. SLFL-3 and SLFL-4 proteins are too similar to each other.

The last reason, why PUCH nuclease was not identified for so long, is a very high sequence similarity between SLFL-3 and SLFL-4 and redundancy in their functions. SLFL-3 and SLFL-4 are 90% similar on the amino acid sequence level, but their nucleotide sequence happened to be different enough for SLFL-4 to escape the silencing by the same RNAi molecule as SLFL-3 and vice versa. Thus, only SLFL-3 was identified in the RNAi screen (Goh et al. 2014). In addition to the screening performed by the Hannon group, SLFL-3 appeared in a mutagenic screen, performed by the Ketting group, in which also *pid-1* was identified, with a missense mutation G16E in the *slfl-3* gene. Still, the effect of a single mutation in SLFL-3 was not strong enough in both screens to attract the research interest. Most of the genetic screens are designed to identify individual genes, and as mutations in duplicated genes are a rare situation, it is more complicated to identify genes, that act redundantly.

Identification of proteins, having very close paralogs, by mass spectrometry is also not a simple task. We were fortunate to identify SLFL-3 among TOFU-2 interactors during the IP/MS experiment, as most of the MS quantifying pipelines, including one used for TOFU-2 IP MS are based on Occam's Razer principle, meaning that the peptides, which cannot be assigned to unique proteins, are excluded from the analysis. In the case of SLFL-3 and SLFL-4 the number of unique peptides is very low, and it is most probably the reason why SLFL-3 or SLFL-4 were not detected in TOFU-2[E216A]:HA version. However, even the identification of SLFL-3 among TOFU-2 interactors was not convincing enough to include SLFL-3/4 in the initial cleavage reaction attempts.

5.2.7. Relationship between PUCH and PETISCO

No interaction between PUCH and PETISCO could be detected by IP-MS approaches (Figure R25). Possibly this is due to the interaction between PUCH and PETISCO complexes being

too transient to be detected by such approaches. Analysis of interactions between recombinant proteins revealed the interaction between the TOFU-1 protein from PUCH and TOFU-6 from PETISCO (Figure R26). Disruption of this interaction by creating point mutations in the endogenous loci encoding TOFU-6 and TOFU-1 proteins resulted in a reduction of 21U RNA levels and moderate accumulation of 21U RNA precursors (Figure R26.C and R26.D), suggesting that interaction between the two complexes is important for the 21U RNA biogenesis. PUCH complex components and PETISCO subunits show the same conservation pattern within nematodes, with an exception for ERH-2 and IFE-3, which are more broadly conserved proteins, suggesting that PUCH and PETISCO evolve together to be a part of the 21U RNA processing pathway.

5.2.7.1. PETISCO protects the precursors from degradation and does not interfere with PUCH-mediated processing.

Our current understanding of the PUCH-mediated processing of 21U RNA precursors suggests that within PETISCO IFE-3 binds precursors via the m^7G cap. Indeed, EMSA performed with PETISCO complex showed that PETISCO binds capped RNA better, than non-capped, but it is still capable of binding RNA, containing 5'P. Then, the TOFU-6-TOFU-1 interaction brings PUCH to the PETISCO complex, and during the processing, the precursor may or may not remain bound to PETISCO. This model suggests that PETISCO should at least not prevent the processing of the precursors by PUCH *in vitro*, and possibly may stimulate PUCH activity. Kinetics of the cleavage reactions, performed in the presence of PETISCO showed that the presence of PETISCO had no effect on PUCH cleavage, indicating that PETISCO is neutral with respect to PUCH activity (Figure R24). It should be noted, though, that PID-1 was not included in these reactions, and I cannot exclude that the presence of PID-1 may in fact stimulate PUCH cleavage, and TOST-1, respectively, could prevent interaction between PUCH and PETISCO by blocking the interaction interface.

But why then is PETISCO binding the precursors *in vivo*? 21U RNA precursors are short, non-spliced, m^7G-capped transcripts, and they do not have a poly-A tail. This makes them an easy target for RNA degradation pathways. Hence, in order to be used, they will need to be stabilized, and this could be the function of PETISCO in the 21U RNA pathway.

An interesting point is, how exactly precursors are recognized by PETISCO. Even excluding the mRNAs, most of which have a TMG cap, there are a lot of other transcripts, for instance, such as long non-coding RNA, which are also produced by RNA Pol II and potentially could have an m^7G

cap, however, most of them are trans-spliced as well. There should be a balance between protection from unwanted degradation and allowance of degradation of unwanted transcripts. Another RNA species, that had been identified to be bound by PETISCO is histone mRNA, which is also m^7G-capped and does not have Poly-A tale (Cordeiro Rodrigues et al. 2019) and unpublished data from another project in the lab). Long non-coding RNAs, for instance, have a long poly-A tail. It is possible, that Poly-A tale helps the transcripts to avoid PETISCO, and serves as another selection function in addition to the m^7G cap. Possibly, Poly-A binding machinery catches such transcripts before PETISCO manages to bind them and transport them to the cytoplasm. However, the exact mechanism of how PETISCO chooses the transcripts to bind still needs clarification. PID-1 has a potential N-terminus nuclear localization signal, and the nuclear export signal as well (Albuquerque et al. 2014), and therefore it is possible that PID-1 shuttles between nucleus and cytoplasm, collecting the RNA transcripts and bringing them to PETISCO.

5.2.7.2. PETISCO might promote the loading of the processed precursors to PRG-1.

After processing by PUCH, 5'-processed 21U RNA precursors, possibly could still be bound by PETISCO, as they are vulnerable to RNA degradation machinery. The MID-domain of PID-3 could play a role in this, as MID-domains in Piwi proteins bind the 5'P end of the piRNAs (MacRae et al. 2008). However, whether the MID-domain of PID-3 indeed has such biochemical activity has not been tested, and the conservation of the residues, necessary for 5'P binding is doubtful. I speculate that PETISCO should interact with PRG-1 to transfer the precursors to it. PRG-1 has never appeared in our IP/MS experiments for PETISCO complex proteins, but it is possible, that the interaction between PETISCO and PRG-1 is transient, like the interaction between PUCH and PETISCO. I think, that the binding is happening either via PID-3, because it is assumed to bind the precursors, or, possibly via the eTUDOR domain of TOFU-6, as TUDOR domains have been shown to interact with PIWI proteins (Siomi et al, 2010). Even though the eTUDOR-domain of TOFU-6 does not have a conserved aromatic cage, which could be binding methylated arginines of PRG-1, it is possible, that TOFU-6 binds unloaded PRG-1, which has not been methylated yet. It has been shown, that in *Drosophila*, arginine methylation of Aubergine happens only after loading with piRNA, and if the same is true for PRG-1, then the absence of the conserved aromatic cage in TOFU-6 is not critical for the interaction with PRG-1 (Huang et al. 2021).

5.2.8. Function of the TOFU-2 SPRY domain.

Besides the SLFN-like fold, TOFU-2 contains the SPRY domain, which provides another link between the 21U RNA formation and innate immune response. SPRY domain was identified in many different proteins, including proteins, responsible for RNA processing (D'Cruz et al. 2013). The main function of the SPRY domain is to be involved in protein-protein interactions, however, not all of the SPRY-domain interactors are known yet. Interestingly, the SPRY domain is found in half of TRIM proteins, which regulate the interferon signaling cascades in reply to the activation of the innate immune response. In addition, The SPRY domain of TRIM5α protein has been shown to bind HIV-1 capsid and prevent viral replication through an unknown mechanism (D'Cruz et al. 2013). Mammalian Schlafen proteins have been shown to prevent viral replication as well (see the introduction), and even though these two mechanisms might be not related, the participation of the SPRY domain in the innate immunity response brings another evolutionary link between TOFU-2 protein and Schlafen proteins. It is possible, that the SPRY domain of TOFU-2 is involved in some protein-protein interactions and immune responses, but it needs to be investigated further.

5.2.9. Other Schlafen-containing protein in *C. elegans*.

TOFU-1, TOFU-2. SLFL-3 and SLFL-4 are not only SLFN-like domain-containing proteins in *C. elegans.* Protein T04H1.2, named SLFL-5, is expressed in gonads and is predicted to be an orthologue of human GTPB2 protein. SLFL-5 is much more conserved than other SLFN proteins (Figure R28) and it did not appear in TOFU-2 IP-MS data. It is not known yet if this protein is interacting with other SLFL proteins, or if this protein is involved in the 21U RNA pathway. Knockout of SLFL-5 did not lead to any visible phenotype, however knockout of GTB2 in mice causes the synthetic neurodegenerative phenotype.

5.2.10. Outside the worms.

Identification of the nuclease PUCH, formed by three different SLFN-fold containing proteins, shows a new kind of very specific nucleases. Previously, only proteins containing a complete Schlafen domain were described. SLFN11 has been shown to process tRNA with a long variable loop, SLFN12 and SLFN14 process different types of rRNA (Garvie et al. 2021; Pisareva et al. 2015), *in vitro* studies of SLFN13 showed the acceptor stem of tRNA, and rRNAs were also

affected (Yang et al. 2018). SLFN5 has an intact active center, but its substrate has not been yet identified (Jo and Pommier 2022). All those findings suggest that SLFN domains can be shaped to perform a very specific job. In addition, there is much evidence that SLFN nucleases undergo positive selection towards specialization of SLFN nuclease activity (Bustos et al. 2009; Stabell et al. 2016).

Partial SLFN-domain-containing proteins, similar to the SLFN-folds identified in PUCH complex, are encoded in the mammalian genome as well. Their function has not yet been studied, but it is possible, that they form active complexes with other SLFNs in a PUCH-like manner. Mostly, SLFN functions are related to the function of the innate immune system, and PUCH is involved in piRNA formation, showing the link between the innate immune system and immunity against invading nucleic acids. As the SLFN-like domain of *C. elegans* T04H1.2 is conserved in human GTPB2, it is clear, that there might be many proteins with uncharacterized small SLFN-like domains.

Besides *C. elegans* and mammals, partial SLFN-folds are found in many different organisms, including bacteria and archaea. Identifications of the partial SLFN-folds require analysis with structure prediction tools, as their sequence similarities to the canonical Schlafen proteins are pretty low (2023)). It is possible, that PUCH is only the first discovered multi-domain SLFN nuclease.

It is easy to imagine, that misregulation of SLFN proteins can lead to the formation of unwanted nucleases, or lack of protective nuclease activity. Mutations in the Om cluster in mice, overlapping with the Schlafen-proteins containing part of the genome, cause DDK syndrome. Mice, containing a deletion of the OM cluster, can give viable progeny when mating with mutant animals while mating with wild-type mice causes lethality of the offspring. Truncated SLFN proteins formed by deletion of the Om cluster could in theory form some detrimental nucleases that would affect the crucial kinds of the RNA leading to sterility or newly inherited wild-type nucleases that were not regulated properly.

The Schlafen-like domain is also present in the family of pox viruses, where is it attached to poxin, a viral cyclic guanosine monophosphate adenosine monophosphate (cGAMP) nuclease, involved in inactivating the anti-viral cGAS–STING pathway. It has been shown, that absence of *v-slfn* does not affect the ability of the viruses to infect cells, and the role of the v-slfn remains

unknown (Hernaez et al. 2020). It is possible, that *v-slfn* forms a complex with one of the essential SLFN proteins, preventing the anti-viral activity of SLFN, but this still needs to be studied.

Interaction between different SLFN-fold containing proteins and the ability of SLFNL1 and other incomplete SLFN-like proteins to form working enzymes, or affect the functions of canonical SLFN-proteins still needs to be investigated. Understanding how exactly the Schlafen domains recognize the substrate may allow protein engineering to design artificial specific nucleases, which will aim for a certain class of RNA.

Publication bibliography

Ablasser, Andrea; Chen, Zhijian J. (2019): cGAS in action: Expanding roles in immunity and inflammation. In *Science (New York, N.Y.)* 363 (6431). DOI: 10.1126/science.aat8657.

Akkouche, Abdou; Mugat, Bruno; Barckmann, Bridlin; Varela-Chavez, Carolina; Li, Blaise; Raffel, Raoul et al. (2017): Piwi Is Required during Drosophila Embryogenesis to License Dual-Strand piRNA Clusters for Transposon Repression in Adult Ovaries. In *Molecular cell* 66 (3), 411-419.e4. DOI: 10.1016/j.molcel.2017.03.017.

Albuquerque, Bruno F. M. de; Luteijn, Maartje J.; Cordeiro Rodrigues, Ricardo J.; van Bergeijk, Petra; Waaijers, Selma; Kaaij, Lucas J. T. et al. (2014): PID-1 is a novel factor that operates during 21U-RNA biogenesis in Caenorhabditis elegans. In *Genes & Development* 28 (7), pp. 683–688. DOI: 10.1101/gad.238220.114.

Alexopoulou et. al. (2001): Recognition of double-stranded RNA and activation of NF-kB by Toll-like receptor 3. In *Nature* 413.

Al-Marsoummi, Sarmad; Vomhof-DeKrey, Emilie E.; Basson, Marc D. (2021): Schlafens: Emerging Proteins in Cancer Cell Biology. In *Cells* 10 (9). DOI: 10.3390/cells10092238.

Aravin, Alexei; Gaidatzis, Dimos; Pfeffer, Sébastien; Lagos-Quintana, Mariana; Landgraf, Pablo; Iovino, Nicola et al. (2006): A novel class of small RNAs bind to MILI protein in mouse testes. In *Nature* 442 (7099), pp. 203–207. DOI: 10.1038/nature04916.

Aravin, Alexei A.; Hannon, Gregory J.; Brennecke, Julius (2007): The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. In *Science (New York, N.Y.)* 318 (5851), pp. 761–764. DOI: 10.1126/science.1146484.

Ariumi, Yasuo; Trono, Didier (2006): Ataxia-telangiectasia-mutated (ATM) protein can enhance human immunodeficiency virus type 1 replication by stimulating Rev function. In *Journal of virology* 80 (5), pp. 2445–2452. DOI: 10.1128/JVI.80.5.2445-2452.2006.

Ashe, Alyson; Sapetschnig, Alexandra; Weick, Eva-Maria; Mitchell, Jacinth; Bagijn, Marloes P.; Cording, Amy C. et al. (2012): piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans. In *Cell* 150 (1), pp. 88–99. DOI: 10.1016/j.cell.2012.06.018.

Bagijn, Marloes P.; Goldstein, Leonard D.; Sapetschnig, Alexandra; Weick, Eva-Maria; Bouasker, Samir; Lehrbach, Nicolas J. et al. (2012): Function, targets, and evolution of Caenorhabditis elegans piRNAs. In *Science (New York, N.Y.)* 337 (6094), pp. 574–578. DOI: 10.1126/science.1220952.

Bartel, David P. (2018): Metazoan MicroRNAs. In *Cell* 173 (1), pp. 20–51. DOI: 10.1016/j.cell.2018.03.006.

Batista, Pedro J.; Ruby, J. Graham; Claycomb, Julie M.; Chiang, Rosaria; Fahlgren, Noah; Kasschau, Kristin D. et al. (2008): PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. In *Molecular cell* 31 (1), pp. 67–78. DOI: 10.1016/j.molcel.2008.06.002.

Bazzini et al. (2012): Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. In *Science (New York, N.Y.)* 336 (6078), pp. 229–233. DOI: 10.1126/science.1216533.

Behm-Ansmant, I.; Rehwinkel, J.; Izaurralde, E. (2006): MicroRNAs silence gene expression by repressing protein expression and/or by promoting mRNA decay. In *Cold Spring Harbor symposia on quantitative biology* 71, pp. 523–530. DOI: 10.1101/sqb.2006.71.013.

Bell, Timothy A.; La Casa-Esperón, Elena de; Doherty, Heather E.; Ideraabdullah, Folami; Kim, Kuikwon; Wang, Yunfei et al. (2006): The paternal gene of the DDK syndrome maps to the Schlafen gene cluster on mouse chromosome 11. In *Genetics* 172 (1), pp. 411–423. DOI: 10.1534/genetics.105.047118.

Beltran, Toni; Barroso, Consuelo; Birkle, Timothy Y.; Stevens, Lewis; Schwartz, Hillel T.; Sternberg, Paul W. et al. (2019): Comparative Epigenomics Reveals that RNA Polymerase II Pausing and Chromatin Domain Organization Control Nematode piRNA Biogenesis. In *Developmental cell* 48 (6), 793-810.e6. DOI: 10.1016/j.devcel.2018.12.026.

Berger, Michael; Krebs, Philippe; Crozat, Karine; Li, Xiaohong; Croker, Ben A.; Siggs, Owen M. et al. (2010): An Slfn2 mutation causes lymphoid and myeloid immunodeficiency due to loss of immune cell quiescence. In *Nature immunology* 11 (4), pp. 335–343. DOI: 10.1038/ni.1847.

Billi, Allison C.; Alessi, Amelia F.; Khivansara, Vishal; Han, Ting; Freeberg, Mallory; Mitani, Shohei; Kim, John K. (2012): The Caenorhabditis elegans HEN1 ortholog, HENN-1, methylates and stabilizes select subclasses of germline small RNAs. In *PLoS genetics* 8 (4), e1002617. DOI: 10.1371/journal.pgen.1002617.

Blackford, Andrew N.; Jackson, Stephen P. (2017): ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. In *Molecular cell* 66 (6), pp. 801–817. DOI: 10.1016/j.molcel.2017.05.015.

Bobadilla Ugarte, Pilar; Barendse, Patrick; Swarts, Daan C. (2023): Argonaute proteins confer immunity in all domains of life. In *Current opinion in microbiology* 74, p. 102313. DOI: 10.1016/j.mib.2023.102313.

Bohmert et al. (1998): AGO1 defines a novel locus of Arabidopsis controlling leaf development. In *The EMBO Journal* 17 (1), pp. 170–180.

Bonilla, Francisco A.; Oettgen, Hans C. (2010): Adaptive immunity. In *The Journal of allergy and clinical immunology* 125 (2 Suppl 2), S33-40. DOI: 10.1016/j.jaci.2009.09.017.

Brady, Gareth; Boggan, Louise; Bowie, Andrew; O'Neill, Luke A. J. (2005): Schlafen-1 causes a cell cycle arrest by inhibiting induction of cyclin D1. In *The Journal of biological chemistry* 280 (35), pp. 30723–30734. DOI: 10.1074/jbc.M500435200.

Braun, Joerg E.; Huntzinger, Eric; Izaurralde, Elisa (2012): A molecular link between miRISCs and deadenylases provides new insight into the mechanism of gene silencing by microRNAs. In *Cold Spring Harbor perspectives in biology* 4 (12). DOI: 10.1101/cshperspect.a012328.

Buckley, Bethany A.; Burkhart, Kirk B.; Gu, Sam Guoping; Spracklin, George; Kershner, Aaron; Fritz, Heidi et al. (2012): A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. In *Nature* 489 (7416), pp. 447–451. DOI: 10.1038/nature11352.

Burkhart, Kirk B.; Guang, Shouhong; Buckley, Bethany A.; Wong, Lily; Bochner, Aaron F.; Kennedy, Scott (2011): A pre-mRNA-associating factor links endogenous siRNAs to chromatin regulation. In *PLoS genetics* 7 (8), e1002249. DOI: 10.1371/journal.pgen.1002249.

Bustos, Olivia; Naik, Saijal; Ayers, Gayle; Casola, Claudio; Perez-Lamigueiro, Maria A.; Chippindale, Paul T. et al. (2009): Evolution of the Schlafen genes, a gene family associated with embryonic lethality, meiotic drive, immune processes and orthopoxvirus virulence. In *Gene* 447 (1), pp. 1–11. DOI: 10.1016/j.gene.2009.07.006.

CA. Janeway (1992): The immune system evolved to discriminate infectious nonself from noninfectious self. In *Immunology Today* Vol. 13 No. 1.

Cecere, Germano; Zheng, Grace X. Y.; Mansisidor, Andres R.; Klymko, Katherine E.; Grishok, Alla (2012): Promoters recognized by forkhead proteins exist for individual 21U-RNAs. In *Molecular cell* 47 (5), pp. 734–745. DOI: 10.1016/j.molcel.2012.06.021.

Cerwenka and Lanier (2001): Natural killer cells, viruses and cancer. In *NATURE REVIEWS, IMMUNOLOGY* 1.

Chakrabarti, Arindam; Jha, Babal Kant; Silverman, Robert H. (2011): New insights into the role of RNase L in innate immunity. In *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research* 31 (1), pp. 49–57. DOI: 10.1089/jir.2010.0120.

Chaturvedi, Lakshmishankar; Sun, Kelian; Walsh, Mary F.; Kuhn, Leslie A.; Basson, Marc D. (2014): The P-loop region of Schlafen 3 acts within the cytosol to induce differentiation of human Caco-2 intestinal epithelial cells. In *Biochimica et biophysica acta* 1843 (12), pp. 3029–3037. DOI: 10.1016/j.bbamcr.2014.09.017.

Chen, Jiaxing; Kuhn, Leslie A. (2019): Deciphering the three-domain architecture in schlafens and the structures and roles of human schlafen12 and serpinB12 in transcriptional regulation. In *Journal of molecular graphics & modelling* 90, pp. 59–76. DOI: 10.1016/j.jmgm.2019.04.003.

Chen, Yung-Chia Ariel; Stuwe, Evelyn; Luo, Yicheng; Ninova, Maria; Le Thomas, Adrien; Rozhavskaya, Ekaterina et al. (2016): Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. In *Molecular cell* 63 (1), pp. 97–109. DOI: 10.1016/j.molcel.2016.05.010.

Cimprich, Karlene A.; Cortez, David (2008): ATR: an essential regulator of genome integrity. In *Nature reviews. Molecular cell biology* 9 (8), pp. 616–627. DOI: 10.1038/nrm2450.

Colins et. al (1989): The Tc3 Familyo f Transposable Genetic Elementsi n Caenorhabditis elegans. In *Genetics* 121 (1), pp. 47–55.

Cordeiro Rodrigues, Ricardo J.; Jesus Domingues, António Miguel de; Hellmann, Svenja; Dietz, Sabrina; Albuquerque, Bruno F. M. de; Renz, Christian et al. (2019): PETISCO is a novel protein complex required for 21U RNA biogenesis and embryonic viability. In *Genes & Development* 33 (13-14), pp. 857–870. DOI: 10.1101/gad.322446.118.

Cosby et al. (2019): Host–transposon interactions: conflict, cooperation, and cooption. In *Genes & Development* 119 (33), pp. 1098–1116.

Cost et. al (2002): Human L1 element target-primed reverse transcripion in vitro. In *EMBO Journal* 21, pp. 5899–5910.

Cox et al. (1998): A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. In *Genes & Development* 12, pp. 3715–3727.

Czech, Benjamin; Malone, Colin D.; Zhou, Rui; Stark, Alexander; Schlingeheyde, Catherine; Dus, Monica et al. (2008): An endogenous small interfering RNA pathway in Drosophila. In *Nature* 453 (7196), pp. 798–802. DOI: 10.1038/nature07007.

D. N. Cox; A. Chao and H. Lin: PIWI, a nuclear stem cell factor.

Dexheimer, Philipp J.; Cochella, Luisa (2020): MicroRNAs: From Mechanism to Organism. In *Frontiers in cell and developmental biology* 8, p. 409. DOI: 10.3389/fcell.2020.00409.

Diebold et. al. (2004): Innate Antiviral Responses by Means of TLR7-Mediated Recognition of Single-Stranded RNA. In *Science (New York, N.Y.)* 303 (5663), pp. 1529–1531.

Drinnenberg et al. (2009): RNAi in Budding Yeast. In *Science (New York, N.Y.)* 326 (5952), pp. 544–550.

Dubin, Manu J.; Mittelsten Scheid, Ortrun; Becker, Claude (2018): Transposons: a blessing curse. In *Current opinion in plant biology* 42, pp. 23–29. DOI: 10.1016/j.pbi.2018.01.003.

Dupressoir et al (2005): Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. In *PNAS* 102 (3), pp. 725–730.

Eaglesham, James B.; Pan, Youdong; Kupper, Thomas S.; Kranzusch, Philip J. (2019): Viral and metazoan poxins are cGAMP-specific nucleases that restrict cGAS-STING signalling. In *Nature* 566 (7743), pp. 259–263. DOI: 10.1038/s41586-019-0928-6.

Eickbush, Thomas H. (2007): R2 and Related Site-Specific Non-Long Terminal Repeat Retrotransposons. In Nancy L. Craig, Robert Craigie, Martin Gellert, Alan M. Lambowitz (Eds.): Mobile DNA II, vol. 10: Wiley, pp. 813–835.

ElMaghraby, Mostafa F.; Andersen, Peter Refsing; Pühringer, Florian; Hohmann, Ulrich; Meixner, Katharina; Lendl, Thomas et al. (2019): A Heterochromatin-Specific RNA Export Pathway Facilitates piRNA Production. In *Cell* 178 (4), 964-979.e20. DOI: 10.1016/j.cell.2019.07.007.

Emmons et. al (1983): Evidence for a Transposon in Caenorhabditis elegans. In *Cell* 32, pp. 55–65.

Estécio, Marcos R. H.; Gharibyan, Vazganush; Shen, Lanlan; Ibrahim, Ashraf E. K.; Doshi, Ketan; He, Rong et al. (2007): LINE-1 hypomethylation in cancer is highly variable and inversely correlated with microsatellite instability. In *PloS one* 2 (5), e399. DOI: 10.1371/journal.pone.0000399.

Feng et. al (1996): Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition. In *Cell* 87, pp. 905–916.

Ferrari, Roberto; Grandi, Nicole; Tramontano, Enzo; Dieci, Giorgio (2021): Retrotransposons as Drivers of Mammalian Brain Evolution. In *Life (Basel, Switzerland)* 11 (5). DOI: 10.3390/life11050376.

Fire et al. (1998): Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. In *Nature* 391, pp. 806–811.

Frank et al (2022): Evolution and antiviral activity of a human protein of retroviral origin. In *Science (New York, N.Y.)* 378 (6618), pp. 422–428.

Gardner, Eric E.; Lok, Benjamin H.; Schneeberger, Valentina E.; Desmeules, Patrice; Miles, Linde A.; Arnold, Paige K. et al. (2017): Chemosensitive Relapse in Small Cell Lung Cancer Proceeds through an EZH2-SLFN11 Axis. In *Cancer cell* 31 (2), pp. 286–299. DOI: 10.1016/j.ccell.2017.01.006.

Garvie, Colin W.; Wu, Xiaoyun; Papanastasiou, Malvina; Lee, Sooncheol; Fuller, James; Schnitzler, Gavin R. et al. (2021): Structure of PDE3A-SLFN12 complex reveals requirements for activation of SLFN12 RNase. In *Nature communications* 12 (1), p. 4375. DOI: 10.1038/s41467-021-24495-w.

Gasser, Stephan; Raulet, David H. (2006): Activation and self-tolerance of natural killer cells. In *Immunological reviews* 214, pp. 130–142. DOI: 10.1111/j.1600-065X.2006.00460.x.

Geserick, Peter; Kaiser, Frank; Klemm, Uwe; Kaufmann, Stefan H. E.; Zerrahn, Jens (2004): Modulation of T cell development and activation by novel members of the Schlafen (slfn) gene family harbouring an

RNA helicase-like motif. In *International immunology* 16 (10), pp. 1535–1548. DOI: 10.1093/intimm/dxh155.

Goh, Wee-Siong Sho; Seah, Jun Wen Eugene; Harrison, Emily J.; Chen, Caifu; Hammell, Christopher M.; Hannon, Gregory J. (2014): A genome-wide RNAi screen identifies factors required for distinct stages of C. elegans piRNA biogenesis. In *Genes & Development* 28 (7), pp. 797–807. DOI: 10.1101/gad.235622.113.

Gonzalez, Segundo; González-Rodríguez, Ana Pilar; Suárez-Álvarez, Beatriz; López-Soto, Alejandro; Huergo-Zapico, Leticia; Lopez-Larrea, Carlos (2011): Conceptual aspects of self and nonself discrimination. In *Self/nonself* 2 (1), pp. 19–25. DOI: 10.4161/self.2.1.15094.

Goriaux, Coline; Desset, Sophie; Renaud, Yoan; Vaury, Chantal; Brasset, Emilie (2014): Transcriptional properties and splicing of the flamenco piRNA cluster. In *EMBO reports* 15 (4), pp. 411–418. DOI: 10.1002/embr.201337898.

Grabundzija, Ivana; Messing, Simon A.; Thomas, Jainy; Cosby, Rachel L.; Bilic, Ilija; Miskey, Csaba et al. (2016): A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. In *Nature communications* 7, p. 10716. DOI: 10.1038/ncomms10716.

Gregory J. Hannon (2002): RNA interference. In *Nature* 418, pp. 244–251.

Gu, Sam Guoping; Pak, Julia; Guang, Shouhong; Maniar, Jay M.; Kennedy, Scott; Fire, Andrew (2012): Amplification of siRNA in Caenorhabditis elegans generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. In *Nature genetics* 44 (2), pp. 157–164. DOI: 10.1038/ng.1039.

Gu, Weifeng; Shirayama, Masaki; Conte, Darryl; Vasale, Jessica; Batista, Pedro J.; Claycomb, Julie M. et al. (2009): Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. In *Molecular cell* 36 (2), pp. 231–244. DOI: 10.1016/j.molcel.2009.09.020.

Gubser, Caroline; Goodbody, Rory; Ecker, Andrea; Brady, Gareth; O'Neill, Luke A. J.; Jacobs, Nathalie; Smith, Geoffrey L. (2007): Camelpox virus encodes a schlafen-like protein that affects orthopoxvirus virulence. In *The Journal of general virology* 88 (Pt 6), pp. 1667–1676. DOI: 10.1099/vir.0.82748-0.

Gunawardane et al. (2007): A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5' End Formation in Drosophila. In *Science (New York, N.Y.)* 315 (5818), pp. 1587–1590.

Guo, Ge; Wang, Yang; Hu, Xiao-Mei; Li, Zhuo-Ran; Tan, Juan; Qiao, Wen-Tao (2021): Human Schlafen 11 exploits codon preference discrimination to attenuate viral protein synthesis of prototype foamy virus (PFV). In *Virology* 555, pp. 78–88. DOI: 10.1016/j.virol.2020.12.015.

Haber (2000): Partners and pathways repairing a double-strand break. In *Trends in genetics : TIG* 16 (6), pp. 259–264.

Hammond et al. (2000): An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. In *Nature* 404, pp. 293–296.

Hancks, Dustin C.; Kazazian, Haig H. (2012): Active human retrotransposons: variation and disease. In *Current opinion in genetics & development* 22 (3), pp. 191–203. DOI: 10.1016/j.gde.2012.02.006.

He, Tao; Zhang, Meiying; Zheng, Ruipan; Zheng, Shufang; Linghu, Enqiang; Herman, James G.; Guo, Mingzhou (2017): Methylation of SLFN11 is a marker of poor prognosis and cisplatin resistance in colorectal cancer. In *Epigenomics* 9 (6), pp. 849–862. DOI: 10.2217/epi-2017-0019.

Hegge, Jorrit W.; Swarts, Daan C.; van der Oost, John (2018): Prokaryotic Argonaute proteins: novel genome-editing tools? In *Nature reviews. Microbiology* 16 (1), pp. 5–11. DOI: 10.1038/nrmicro.2017.73.

Hemmi et. al (2000): A Toll-like receptor recognizes bacterial DNA. In *Nature* 408.

Hernaez et al. (2020): Viral cGAMP nuclease reveals the essential role of DNA sensing in protection against acute lethal virus infection. In *Science Advances* 6 (38).

Hornung, Veit; Ellegast, Jana; Kim, Sarah; Brzózka, Krzysztof; Jung, Andreas; Kato, Hiroki et al. (2006): 5'-Triphosphate RNA is the ligand for RIG-I. In *Science (New York, N.Y.)* 314 (5801), pp. 994–997. DOI: 10.1126/science.1132505.

Horwich, Michael D.; Li, Chengjian; Matranga, Christian; Vagin, Vasily; Farley, Gwen; Wang, Peng; Zamore, Phillip D. (2007): The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. In *Current biology : CB* 17 (14), pp. 1265–1272. DOI: 10.1016/j.cub.2007.06.030.

Hou, Pengjiao; Hao, Wei; Qin, Bo; Li, Mengyun; Zhao, Rong; Cui, Sheng (2023): Structural and biochemical characterization of Schlafen11 N-terminal domain. In *Nucleic Acids Research* 51 (13), pp. 7053–7070. DOI: 10.1093/nar/gkad509.

Huang, Cheng Ran Lisa; Burns, Kathleen H.; Boeke, Jef D. (2012): Active transposition in genomes. In *Annual review of genetics* 46, pp. 651–675. DOI: 10.1146/annurev-genet-110711-155616.

Hur, Junho K.; Olovnikov, Ivan; Aravin, Alexei A. (2014): Prokaryotic Argonautes defend genomes against invasive DNA. In *Trends in biochemical sciences* 39 (6), pp. 257–259. DOI: 10.1016/j.tibs.2014.04.006.

Ichiyanagi K (2013): Epigenetic regulation of transcription and possible functions of mammalian short interspersedelements, SINEs. In *Genes Genet. Syst* 88 (88), pp. 19–29.

Ilyna and Koopin (1992): Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. In *Nucleic Acids Research* 20 (13), pp. 3279–3285.

Ipsaro, Jonathan J.; Haase, Astrid D.; Knott, Simon R.; Joshua-Tor, Leemor; Hannon, Gregory J. (2012): The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. In *Nature* 491 (7423), pp. 279–283. DOI: 10.1038/nature11502.

Ishikawa, Hiroki; Barber, Glen N. (2008): STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. In *Nature* 455 (7213), pp. 674–678. DOI: 10.1038/nature07317.

Ishikawa, Hiroki; Ma, Zhe; Barber, Glen N. (2009): STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. In *Nature* 461 (7265), pp. 788–792. DOI: 10.1038/nature08476.

Iskow, Rebecca C.; McCabe, Michael T.; Mills, Ryan E.; Torene, Spencer; Pittard, W. Stephen; Neuwald, Andrew F. et al. (2010): Natural mutagenesis of human genomes by endogenous retrotransposons. In *Cell* 141 (7), pp. 1253–1261. DOI: 10.1016/j.cell.2010.05.020.

Iwasaki, Shintaro; Kobayashi, Maki; Yoda, Mayuko; Sakaguchi, Yuriko; Katsuma, Susumu; Suzuki, Tsutomu; Tomari, Yukihide (2010): Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. In *Molecular cell* 39 (2), pp. 292–299. DOI: 10.1016/j.molcel.2010.05.015.

Janeway, Charles A.; Medzhitov, Ruslan (2002): Innate immune recognition. In *Annual review of immunology* 20, pp. 197–216. DOI: 10.1146/annurev.immunol.20.083001.084359.

Jinek, Martin; Doudna, Jennifer A. (2009): A three-dimensional view of the molecular machinery of RNA interference. In *Nature* 457 (7228), pp. 405–412. DOI: 10.1038/nature07755.

Jo, Ukhyun; Pommier, Yves (2022): Structural, molecular, and functional insights into Schlafen proteins. In *Experimental & molecular medicine* 54 (6), pp. 730–738. DOI: 10.1038/s12276-022-00794-0.

Kalmykova, Alla I.; Klenov, Mikhail S.; Gvozdev, Vladimir A. (2005): Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. In *Nucleic Acids Research* 33 (6), pp. 2052–2059. DOI: 10.1093/nar/gki323.

Kapitonov, Vladimir V.; Jurka, Jerzy (2007): Helitrons on a roll: eukaryotic rolling-circle transposons. In *Trends in genetics : TIG* 23 (10), pp. 521–529. DOI: 10.1016/j.tig.2007.08.004.

Karre et. al. (1986): Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy. In *Nature* vol. 319.

Kato, Hiroki; Sato, Shintaro; Yoneyama, Mitsutoshi; Yamamoto, Masahiro; Uematsu, Satoshi; Matsui, Kosuke et al. (2005): Cell type-specific involvement of RIG-I in antiviral response. In *Immunity* 23 (1), pp. 19–28. DOI: 10.1016/j.immuni.2005.04.010.

Katsoulidis, Efstratios; Carayol, Nathalie; Woodard, Jennifer; Konieczna, Iwona; Majchrzak-Kita, Beata; Jordan, Alison et al. (2009): Role of Schlafen 2 (SLFN2) in the generation of interferon alpha-induced growth inhibitory responses. In *The Journal of biological chemistry* 284 (37), pp. 25051–25064. DOI: 10.1074/jbc.M109.030445.

Kawai, Taro; Akira, Shizuo (2010): The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. In *Nature immunology* 11 (5), pp. 373–384. DOI: 10.1038/ni.1863.

Kazazian, H. H. (1998): Mobile elements and desease. In *Current opinion in genetics & development* 8, pp. 343–350.

Keiper, B. D.; Lamphear, B. J.; Deshpande, A. M.; Jankowska-Anyszka, M.; Aamodt, E. J.; Blumenthal, T.; Rhoads, R. E. (2000): Functional characterization of five eIF4E isoforms in Caenorhabditis elegans. In *The Journal of biological chemistry* 275 (14), pp. 10590–10596. DOI: 10.1074/jbc.275.14.10590.

Kennerdell and Carthew (1998): Use of dsRNA-Mediated Genetic Interference to Demonstrate that frizzled and frizzled 2 Act in the Wingless Pathway. In *Cell* 95, pp. 1017–1026.

ketting et al. (1999): mut-7 of C. elegans, Required for Transposon Silencing and RNA Interference, Is a Homolog of Werner Syndrome Helicase and RNaseD. In *Cell* 99, pp. 133–141.

René F. Ketting, Luisa Cochella. Chapter Three - Concepts and functions of small RNA pathways in C. elegans. *Current Topics in Developmental Biology*, 144, 2021, 45-89, ISSN 0070-2153.

Khazina, Elena; Truffault, Vincent; Büttner, Regina; Schmidt, Steffen; Coles, Murray; Weichenrieder, Oliver (2011): Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. In *Nature structural & molecular biology* 18 (9), pp. 1006–1014. DOI: 10.1038/nsmb.2097.

Kidwell M. (2002): Transposable elements and the evolution of genome size in eukaryotes. In *Genetica* 115, pp. 46–63.

Kim, Eui Tae; Weitzman, Matthew D. (2022): Schlafens Can Put Viruses to Sleep. In *Viruses* 14 (2). DOI: 10.3390/v14020442.

Kolosha and Martin (1997): In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. In *Proc. Natl. Acad. Sci.* 94, pp. 10155–10160.

Koonin, Eugene V.; Makarova, Kira S.; Zhang, Feng (2017): Diversity, classification and evolution of CRISPR-Cas systems. In *Current opinion in microbiology* 37, pp. 67–78. DOI: 10.1016/j.mib.2017.05.008.

Koopal, Balwina; Mutte, Sumanth K.; Swarts, Daan C. (2023): A long look at short prokaryotic Argonautes. In *Trends in cell biology* 33 (7), pp. 605–618. DOI: 10.1016/j.tcb.2022.10.005.

Koopal, Balwina; Potocnik, Ana; Mutte, Sumanth K.; Aparicio-Maldonado, Cristian; Lindhoud, Simon; Vervoort, Jacques J. M. et al. (2022): Short prokaryotic Argonaute systems trigger cell death upon detection of invading DNA. In *Cell* 185 (9), 1471-1486.e19. DOI: 10.1016/j.cell.2022.03.012.

Krieg et. al (1995): CpG motifs in bacterial DNA trigger direct B-cell activation. In *Nature* 374.

Kuramochi-Miyagawa, Satomi; Watanabe, Toshiaki; Gotoh, Kengo; Totoki, Yasushi; Toyoda, Atsushi; Ikawa, Masahito et al. (2008): DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. In *Genes & Development* 22 (7), pp. 908–917. DOI: 10.1101/gad.1640708.

Lahmy, Sylvie; Pontier, Dominique; Bies-Etheve, Natacha; Laudié, Michèle; Feng, Suhua; Jobet, Edouard et al. (2016): Evidence for ARGONAUTE4-DNA interactions in RNA-directed DNA methylation in plants. In *Genes & Development* 30 (23), pp. 2565–2570. DOI: 10.1101/gad.289553.116.

Laski et al (1986): Tissue Specificity of Drosophila P Element Transposition Is Regulated at the Level of mRNA Splicing. In *Cell* 44, pp. 7–19.

Le Thomas, Adrien; Stuwe, Evelyn; Li, Sisi; Du, Jiamu; Marinov, Georgi; Rozhkov, Nikolay et al. (2014): Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. In *Genes & Development* 28 (15), pp. 1667–1680. DOI: 10.1101/gad.245514.114.

Lee, Heng-Chi; Gu, Weifeng; Shirayama, Masaki; Youngman, Elaine; Conte, Darryl; Mello, Craig C. (2012): C. elegans piRNAs mediate the genome-wide surveillance of germline transcripts. In *Cell* 150 (1), pp. 78–87. DOI: 10.1016/j.cell.2012.06.016.

Lee, Yoontae; Ahn, Chiyoung; Han, Jinju; Choi, Hyounjeong; Kim, Jaekwang; Yim, Jeongbin et al. (2003): The nuclear RNase III Drosha initiates microRNA processing. In *Nature* 425 (6956), pp. 415–419. DOI: 10.1038/nature01957.

Lee et al. (2004): MicroRNA genes are transcribed by RNA polymerase II. In *The EMBO Journal* 23, pp. 4051–4060.

Li, Manqing; Kao, Elaine; Gao, Xia; Sandig, Hilary; Limmer, Kirsten; Pavon-Eternod, Mariana et al. (2012): Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. In *Nature* 491 (7422), pp. 125–128. DOI: 10.1038/nature11433.

Li, Manqing; Kao, Elaine; Malone, Dane; Gao, Xia; Wang, Jean Y. J.; David, Michael (2018): DNA damage-induced cell death relies on SLFN11-dependent cleavage of distinct type II tRNAs. In *Nature structural & molecular biology* 25 (11), pp. 1047–1058. DOI: 10.1038/s41594-018-0142-5.

Lin, Ching-Jung; Hu, Fuqu; Dubruille, Raphaelle; Vedanayagam, Jeffrey; Wen, Jiayu; Smibert, Peter et al. (2018): The hpRNA/RNAi Pathway Is Essential to Resolve Intragenomic Conflict in the Drosophila Male Germline. In *Developmental cell* 46 (3), 316-326.e5. DOI: 10.1016/j.devcel.2018.07.004.

Lisitskaya, Lidiya; Aravin, Alexei A.; Kulbachinskiy, Andrey (2018): DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. In *Nature communications* 9 (1), p. 5165. DOI: 10.1038/s41467-018-07449-7.

Liu, Dong; Wu, Hao; Wang, Chenguang; Li, Yanjun; Tian, Huabin; Siraj, Sami et al. (2019): STING directly activates autophagy to tune the innate immune response. In *Cell death and differentiation* 26 (9), pp. 1735–1749. DOI: 10.1038/s41418-018-0251-z.

Liu et al. (2004): Argonaute2 Is the Catalytic Engine of Mammalian RNAi. In *Science (New York, N.Y.)* 305 (5689).

López-Larrea, Carlos; Suárez-Alvarez, Beatriz; López-Soto, Alejandro; López-Vázquez, Antonio; Gonzalez, Segundo (2008): The NKG2D receptor: sensing stressed cells. In *Trends in molecular medicine* 14 (4), pp. 179–189. DOI: 10.1016/j.molmed.2008.02.004.

Lupas and Martin (2002): PII: S0959-440X(02)00388-3. In *Current Opinion in Structural biology* 12 (6), pp. 746–753.

Maartje J Luteijn; Petra van Bergeijk; Lucas J T Kaaij; Miguel Vasconcelos Almeida; Elke F Roovers; Eugene Berezikov et al. (2012): Extremely stable Piwi-induced gene silencing in Caenorhabditis elegans. In *The EMBO Journal* 31, pp. 3422–3430.

Makarova, Kira S.; Wolf, Yuri I.; van der Oost, John; Koonin, Eugene V. (2009): Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. In *Biology direct* 4, p. 29. DOI: 10.1186/1745-6150-4-29.

Malone, Colin D.; Brennecke, Julius; Dus, Monica; Stark, Alexander; McCombie, W. Richard; Sachidanandam, Ravi; Hannon, Gregory J. (2009): Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. In *Cell* 137 (3), pp. 522–535. DOI: 10.1016/j.cell.2009.03.040.

Maniar, Jay M.; Fire, Andrew Z. (2011): EGO-1, a C. elegans RdRP, modulates gene expression via production of mRNA-templated short antisense RNAs. In *Current biology : CB* 21 (6), pp. 449–459. DOI: 10.1016/j.cub.2011.02.019.

Mao, Hui; Zhu, Chengming; Zong, Dandan; Weng, Chenchun; Yang, Xiangwei; Huang, Hui et al. (2015): The Nrde Pathway Mediates Small-RNA-Directed Histone H3 Lysine 27 Trimethylation in Caenorhabditis elegans. In *Current biology : CB* 25 (18), pp. 2398–2403. DOI: 10.1016/j.cub.2015.07.051.

Maroney, Patricia A.; Yu, Yang; Fisher, Jesse; Nilsen, Timothy W. (2006): Evidence that microRNAs are associated with translating messenger RNAs in human cells. In *Nature structural & molecular biology* 13 (12), pp. 1102–1107. DOI: 10.1038/nsmb1174.

Martin, S. L.; Bushman, F. D. (2001): Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. In *Molecular and cellular biology* 21 (2), pp. 467–475. DOI: 10.1128/MCB.21.2.467-475.2001.

Martin, Sandra L. (2006): The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. In *Journal of biomedicine & biotechnology* 2006 (1), p. 45621. DOI: 10.1155/JBB/2006/45621.

Mathias, S. L.; Scott, A. F.; Kazazian, H. H.; Boeke, J. D.; Gabriel, A. (1991): Reverse transcriptase encoded by a human transposable element. In *Science (New York, N.Y.)* 254 (5039), pp. 1808–1810. DOI: 10.1126/science.1722352.

McCullers, Tabitha J.; Steiniger, Mindy (2017): Transposable elements in Drosophila. In *Mobile genetic elements* 7 (3), pp. 1–18. DOI: 10.1080/2159256X.2017.1318201.

McLsaght et al. (2003): Extensive gene gain associated with adaptive evolution of poxviruses. In *PNAS* 100 (26), pp. 15655–15660.

Meister, Gunter (2007): miRNAs get an early start on translational silencing. In *Cell* 131 (1), pp. 25–28. DOI: 10.1016/j.cell.2007.09.021.

Metzner, Felix J.; Huber, Elisabeth; Hopfner, Karl-Peter; Lammens, Katja (2022a): Structural and biochemical characterization of human Schlafen 5. In *Nucleic Acids Research* 50 (2), pp. 1147–1161. DOI: 10.1093/nar/gkab1278.

Metzner, Felix J.; Wenzl, Simon J.; Kugler, Michael; Krebs, Stefan; Hopfner, Karl-Peter; Lammens, Katja (2022b): Mechanistic understanding of human SLFN11. In *Nature communications* 13 (1), p. 5464. DOI: 10.1038/s41467-022-33123-0.

Mills, Ryan E.; Bennett, E. Andrew; Iskow, Rebecca C.; Devine, Scott E. (2007): Which transposable elements are active in the human genome? In *Trends in genetics : TIG* 23 (4), pp. 183–191. DOI: 10.1016/j.tig.2007.02.006.

Mohn, Fabio; Sienski, Grzegorz; Handler, Dominik; Brennecke, Julius (2014): The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. In *Cell* 157 (6), pp. 1364–1379. DOI: 10.1016/j.cell.2014.04.031.

Mohn et al. (2015): piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. In *Science (New York, N.Y.)* 348 (6236), pp. 812–817.

Montgomery, Mary K.; Xu, SiQun; Fire, Andrew (1998): RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans. In *Proceedings of the National Academy of Sciences of the United States of America* 95 (26), pp. 15502–15507.

Munafò, Marzia; Manelli, Vera; Falconio, Federica A.; Sawle, Ashley; Kneuss, Emma; Eastwood, Evelyn L. et al. (2019): Daedalus and Gasz recruit Armitage to mitochondria, bringing piRNA precursors to the biogenesis machinery. In *Genes & Development* 33 (13-14), pp. 844–856. DOI: 10.1101/gad.325662.119.

Muñoz-López (2010): DNA Transposons: Nature and Applications in Genomics. In *Current Genomics* 11, pp. 115–128.

Murai, Junko; Thomas, Anish; Miettinen, Markku; Pommier, Yves (2019a): Schlafen 11 (SLFN11), a restriction factor for replicative stress induced by DNA-targeting anti-cancer therapies. In *Pharmacology & therapeutics* 201, pp. 94–102. DOI: 10.1016/j.pharmthera.2019.05.009.

Murai, Junko; Thomas, Anish; Miettinen, Markku; Pommier, Yves (2019b): Schlafen 11 (SLFN11), a restriction factor for replicative stress induced by DNA-targeting anti-cancer therapies. In *Pharmacology & therapeutics* 201, pp. 94–102. DOI: 10.1016/j.pharmthera.2019.05.009.

Murai, Junko; Zhang, Hongliang; Pongor, Lorinc; Tang, Sai-Wen; Jo, Ukhyun; Moribe, Fumiya et al. (2020): Chromatin Remodeling and Immediate Early Gene Activation by SLFN11 in Response to Replication Stress. In *Cell reports* 30 (12), 4137-4151.e6. DOI: 10.1016/j.celrep.2020.02.117.

Nakanishi, Kotaro; Weinberg, David E.; Bartel, David P.; Patel, Dinshaw J. (2012): Structure of yeast Argonaute with guide RNA. In *Nature* 486 (7403), pp. 368–374. DOI: 10.1038/nature11211.

Neumann, Brent; Zhao, Liang; Murphy, Kathleen; Gonda, Thomas J. (2008): Subcellular localization of the Schlafen protein family. In *Biochemical and biophysical research communications* 370 (1), pp. 62–66. DOI: 10.1016/j.bbrc.2008.03.032.

Nguyen, Tuan Anh; Jo, Myung Hyun; Choi, Yeon-Gil; Park, Joha; Kwon, S. Chul; Hohng, Sungchul et al. (2015): Functional Anatomy of the Human Microprocessor. In *Cell* 161 (6), pp. 1374–1387. DOI: 10.1016/j.cell.2015.05.010.

Ohtani, Hitoshi; Iwasaki, Yuka W. (2021): Rewiring of chromatin state and gene expression by transposable elements. In *Development, growth & differentiation* 63 (4-5), pp. 262–273. DOI: 10.1111/dgd.12735.

Okamura, Katsutomo; Balla, Sudha; Martin, Raquel; Liu, Na; Lai, Eric C. (2008): Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster. In *Nature structural & molecular biology* 15 (6), pp. 581–590. DOI: 10.1038/nsmb.1438.

Pace, John K.; Feschotte, Cédric (2007): The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. In *Genome research* 17 (4), pp. 422–432. DOI: 10.1101/gr.5826307.

Pak and Fire (2007): Distinct Populations of Primary and Secondary Effectors During RNAi in C. elegans. In *Science (New York, N.Y.)* 315 (5809).

Peters, Lasse; Meister, Gunter (2007): Argonaute proteins: mediators of RNA silencing. In *Molecular cell* 26 (5), pp. 611–623. DOI: 10.1016/j.molcel.2007.05.001.

Phillips, Carolyn M.; Montgomery, Taiowa A.; Breen, Peter C.; Ruvkun, Gary (2012): MUT-16 promotes formation of perinuclear mutator foci required for RNA silencing in the C. elegans germline. In *Genes & Development* 26 (13), pp. 1433–1444. DOI: 10.1101/gad.193904.112.

PII: S0955-0674(02)00338-1.

Pisareva, Vera P.; Muslimov, Ilham A.; Tcherepanov, Andrew; Pisarev, Andrey V. (2015): Characterization of Novel Ribosome-Associated Endoribonuclease SLFN14 from Rabbit Reticulocytes. In *Biochemistry* 54 (21), pp. 3286–3301. DOI: 10.1021/acs.biochem.5b00302.

Pritykin et al. (2017): Integrative analysis unveils new functions for the Drosophila Cutoff protein in noncoding RNA biogenesis and gene regulation. In *RNA* 23, pp. 1097–1109.

Puck, Alexander; Aigner, Regina; Modak, Madhura; Cejka, Petra; Blaas, Dieter; Stöckl, Johannes (2015): Expression and regulation of Schlafen (SLFN) family members in primary human monocytes, monocyte-derived dendritic cells and T cells. In *Results in immunology* 5, pp. 23–32. DOI: 10.1016/j.rinim.2015.10.001.

Rath, Devashish; Amlinger, Lina; Rath, Archana; Lundgren, Magnus (2015): The CRISPR-Cas immune system: biology, mechanisms and applications. In *Biochimie* 117, pp. 119–128. DOI: 10.1016/j.biochi.2015.03.025.

Reinhold, William C.; Thomas, Anish; Pommier, Yves (2017): DNA-Targeted Precision Medicine; Have we Been Caught Sleeping? In *Trends in cancer* 3 (1), pp. 2–6. DOI: 10.1016/j.trecan.2016.11.002.

Roach et. al. (2005): The evolution of vertebrate Toll-like receptors. In *PNAS* vol. 102 (no. 27), pp. 9577–9582.

Ruby, J. Graham; Jan, Calvin; Player, Christopher; Axtell, Michael J.; Lee, William; Nusbaum, Chad et al. (2006): Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. In *Cell* 127 (6), pp. 1193–1207. DOI: 10.1016/j.cell.2006.10.040.

Ruby, J. Graham; Jan, Calvin H.; Bartel, David P. (2007): Intronic microRNA precursors that bypass Drosha processing. In *Nature* 448 (7149), pp. 83–86. DOI: 10.1038/nature05983.

Ryazansky, Sergei; Kulbachinskiy, Andrey; Aravin, Alexei A. (2018): The Expanded Universe of Prokaryotic Argonaute Proteins. In *mBio* 9 (6). DOI: 10.1128/mBio.01935-18.

SanMiguel, P.; Tikhonov, A.; Jin, Y. K.; Motchoulskaia, N.; Zakharov, D.; Melake-Berhan, A. et al. (1996): Nested retrotransposons in the intergenic regions of the maize genome. In *Science (New York, N.Y.)* 274 (5288), pp. 765–768. DOI: 10.1126/science.274.5288.765.

Schirle and MacRae (2012): The Crystal Structure of Human Argonaute2. In *Science (New York, N.Y.)* 336 (6084).

Schlee, Martin; Hartmann, Gunther (2016): Discriminating self from non-self in nucleic acid sensing. In *Nature reviews. Immunology* 16 (9), pp. 566–580. DOI: 10.1038/nri.2016.78.

Schlee, Martin; Roth, Andreas; Hornung, Veit; Hagmann, Cristina Amparo; Wimmenauer, Vera; Barchet, Winfried 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. In *Immunity* 31 (1), pp. 25–34. DOI: 10.1016/j.immuni.2009.05.008.

Schwarz et al. (1998): Schlafen, a New Family of Growth Regulatory Genes that Affect Thymocyte Development. In *Immunity* 9, pp. 657–668.

Schwarz et al. (2203): Asymmetry in the Assembly of the RNAi Enzyme Complex. In *Cell* 115, pp. 199–2008.

Seong, Rak-Kyun; Seo, Seong-Wook; Kim, Ji-Ae; Fletcher, Sarah J.; Morgan, Neil V.; Kumar, Mukesh et al. (2017): Schlafen 14 (SLFN14) is a novel antiviral factor involved in the control of viral replication. In *Immunobiology* 222 (11), pp. 979–988. DOI: 10.1016/j.imbio.2017.07.002.

Setten, Ryan L.; Rossi, John J.; Han, Si-Ping (2019): The current state and future directions of RNAi-based therapeutics. In *Nature reviews. Drug discovery* 18 (6), pp. 421–446. DOI: 10.1038/s41573-019-0017-4.

Shirayama, Masaki; Seth, Meetu; Lee, Heng-Chi; Gu, Weifeng; Ishidate, Takao; Conte, Darryl; Mello, Craig C. (2012): piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. In *Cell* 150 (1), pp. 65–77. DOI: 10.1016/j.cell.2012.06.015.

Sienski, Grzegorz; Dönertas, Derya; Brennecke, Julius (2012): Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. In *Cell* 151 (5), pp. 964–980. DOI: 10.1016/j.cell.2012.10.040.

Smardon et al. (2000): EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. In *Current Biology* 10, pp. 169–178.

Solyom, Szilvia; Ewing, Adam D.; Rahrmann, Eric P.; Doucet, Tara; Nelson, Heather H.; Burns, Michael B. et al. (2012): Extensive somatic L1 retrotransposition in colorectal tumors. In *Genome research* 22 (12), pp. 2328–2338. DOI: 10.1101/gr.145235.112.

Sotero-Caio, Cibele G.; Platt, Roy N.; Suh, Alexander; Ray, David A. (2017): Evolution and Diversity of Transposable Elements in Vertebrate Genomes. In *Genome biology and evolution* 9 (1), pp. 161–177. DOI: 10.1093/gbe/evw264.

Sultana, Tania; Zamborlini, Alessia; Cristofari, Gael; Lesage, Pascale (2017): Integration site selection by retroviruses and transposable elements in eukaryotes. In *Nature reviews. Genetics* 18 (5), pp. 292–308. DOI: 10.1038/nrg.2017.7.

Swarts, Daan C.; Makarova, Kira; Wang, Yanli; Nakanishi, Kotaro; Ketting, René F.; Koonin, Eugene V. et al. (2014): The evolutionary journey of Argonaute proteins. In *Nature structural & molecular biology* 21 (9), pp. 743–753. DOI: 10.1038/nsmb.2879.

Szittya, G., Burgyán, J. (2013). RNA Interference-Mediated Intrinsic Antiviral Immunity in Plants. In: Cullen, B. (eds) Intrinsic Immunity. Current Topics in Microbiology and Immunology, vol 371. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37765-5_6

Taylor, Martin S.; LaCava, John; Mita, Paolo; Molloy, Kelly R.; Huang, Cheng Ran Lisa; Li, Donghui et al. (2013): Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. In Cell 155 (5), pp. 1034–1048. DOI: 10.1016/j.cell.2013.10.021.

Thomas, Jainy; Pritham, Ellen J. (2015): Helitrons, the Eukaryotic Rolling-circle Transposable Elements. In Microbiology spectrum 3 (4). DOI: 10.1128/microbiolspec.MDNA3-0049-2014.

Valdez, Federico; Salvador, Julienne; Palermo, Pedro M.; Mohl, Jonathon E.; Hanley, Kathryn A.; Watts, Douglas; Llano, Manuel (2019): Schlafen 11 Restricts Flavivirus Replication. In Journal of virology 93 (15). DOI: 10.1128/JVI.00104-19.

Van der Krol et al. (1990): Flavonoid Genes in Petunia: Addition of a Limited Number of Gene Copies May Lead to a Suppression of Gene Expression. In The Plant Cell 2, pp. 291–299.

Vaucheret, Hervé; Vazquez, Franck; Crété, Patrice; Bartel, David P. (2004): The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. In Genes & Development 18 (10), pp. 1187–1197. DOI: 10.1101/gad.1201404.

Wang, Guilin; Reinke, Valerie (2008): A C. elegans Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. In Current biology : CB 18 (12), pp. 861–867. DOI: 10.1016/j.cub.2008.05.009.

Wang, Hui; Ma, Zaijun; Niu, Kongyan; Xiao, Yi; Wu, Xiaofen; Pan, Chenyu et al. (2016): Antagonistic roles of Nibbler and Hen1 in modulating piRNA 3' ends in Drosophila. In Development (Cambridge, England) 143 (3), pp. 530–539. DOI: 10.1242/dev.128116.

Wang, Xian-Bing; Jovel, Juan; Udomporn, Petchthai; Wang, Ying; Wu, Qingfa; Li, Wan-Xiang et al. (2011): The 21-nucleotide, but not 22-nucleotide, viral secondary small interfering RNAs direct potent antiviral defense by two cooperative argonautes in Arabidopsis thaliana. In The Plant Cell 23 (4), pp. 1625–1638. DOI: 10.1105/tpc.110.082305.

Weiner (2002): SINEs and LINEs: the art of biting the hand that feeds you. In Current Opinion in Cell Biology 14, pp. 343–350.

Weng, Chenchun; Kosalka, Joanna; Berkyurek, Ahmet C.; Stempor, Przemyslaw; Feng, Xuezhu; Mao, Hui et al. (2019): The USTC co-opts an ancient machinery to drive piRNA transcription in C. elegans. In Genes & Development 33 (1-2), pp. 90–102. DOI: 10.1101/gad.319293.118.

Wu, Jian-jun; Li, Wenwei; Shao, Yaming; Avey, Denis; Fu, Bishi; Gillen, Joseph et al. (2015): Inhibition of cGAS DNA Sensing by a Herpesvirus Virion Protein. In Cell host & microbe 18 (3), pp. 333–344. DOI: 10.1016/j.chom.2015.07.015.

Yang, Jin-Yu; Deng, Xiang-Yu; Li, Yi-Sheng; Ma, Xian-Cai; Feng, Jian-Xiong; Yu, Bing et al. (2018): Structure of Schlafen13 reveals a new class of tRNA/rRNA- targeting RNase engaged in translational control. In Nature communications 9 (1), p. 1165. DOI: 10.1038/s41467-018-03544-x.

Yashiro, Ryu; Murota, Yukiko; Nishida, Kazumichi M.; Yamashiro, Haruna; Fujii, Kaede; Ogai, Asuka et al. (2018): Piwi Nuclear Localization and Its Regulatory Mechanism in Drosophila Ovarian Somatic Cells. In Cell reports 23 (12), pp. 3647–3657. DOI: 10.1016/j.celrep.2018.05.051.

Yigit, Erbay; Batista, Pedro J.; Bei, Yanxia; Pang, Ka Ming; Chen, Chun-Chieh G.; Tolia, Niraj H. et al. (2006): Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. In Cell 127 (4), pp. 747–757. DOI: 10.1016/j.cell.2006.09.033.

Yu, Tianxiong; Koppetsch, Birgit S.; Pagliarani, Sara; Johnston, Stephen; Silverstein, Noah J.; Luban, Jeremy et al. (2019): The piRNA Response to Retroviral Invasion of the Koala Genome. In Cell 179 (3), 632-643.e12. DOI: 10.1016/j.cell.2019.09.002.

Yuan, Yu-Ren; Pei, Yi; Ma, Jin-Biao; Kuryavyi, Vitaly; Zhadina, Maria; Meister, Gunter et al. (2005): Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. In Molecular cell 19 (3), pp. 405–419. DOI: 10.1016/j.molcel.2005.07.011.

Zeng, Chenming; Weng, Chenchun; Wang, Xiaoyang; Yan, Yong-Hong; Li, Wen-Jun; Xu, Demin et al. (2019): Functional Proteomics Identifies a PICS Complex Required for piRNA Maturation and Chromosome Segregation. In Cell reports 27 (12), 3561-3572.e3. DOI: 10.1016/j.celrep.2019.05.076.

Zhang, Bingnan; Ramkumar, Kavya; Cardnell, Robert John; Gay, Carl Michael; Stewart, C. Allison; Wang, Wei-Lien et al. (2021): A wake-up call for cancer DNA damage: the role of Schlafen 11 (SLFN11) across multiple cancers. In British journal of cancer 125 (10), pp. 1333–1340. DOI: 10.1038/s41416-021-01476-w.

Zhang, Xu; Shi, Heping; Wu, Jiaxi; Zhang, Xuewu; Sun, Lijun; Chen, Chuo; Chen, Zhijian J. (2013): Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. In Molecular cell 51 (2), pp. 226–235. DOI: 10.1016/j.molcel.2013.05.022.

Zhao, Liang; Neumann, Brent; Murphy, Kathleen; Silke, John; Gonda, Thomas J. (2008): Lack of reproducible growth inhibition by Schlafen1 and Schlafen2 in vitro. In Blood cells, molecules & diseases 41 (2), pp. 188–193. DOI: 10.1016/j.bcmd.2008.03.006.

Zhou, Rui; Rana, Tariq M. (2013): RNA-based mechanisms regulating host-virus interactions. In Immunological reviews 253 (1), pp. 97–111. DOI: 10.1111/imr.12053.