From Department of Cell and Molecular Biology
Karolinska Institutet, Stockholm, Sweden

THE ROLE OF PAIRING BEYOND THE SEED IN MICRORNA TARGETING

David Kosek

Karolinska
Institutet

Stockholm 2024

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2024
© David Kosek, 2024
ISBN 978-91-8017-308-7

The role of pairing beyond the seed in microRNA targeting

Thesis for Doctoral Degree (Ph.D.)

By

David Kosek

The thesis will be defended in public at Jacob Berzelius lecture hall, Berzelius väg 3, 171 65 Solna, March 22 9:00

Principal Supervisor:
Associate Professor Emma Andersson
Karolinska Institutet
Department of Cell and Molecular Biology

Co-supervisor(s):
Professor Katja Petzold
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Uppsala University
Department of Medical Biochemistry and Microbiology

Professor Gonçalo Castelo-Branco
Karolinska Institutet
Department of Medical Biochemistry and Biophysics

Opponent:
Professor François Major
Université de Montréal
Institute for Research in Immunology and Cancer and Department of Computer Science and Operations Research

Examination Board:
Associate Professor Weng-Onn Lui
Karolinska Institutet
Department of Oncology-Pathology

Associate Professor Marc Friedländer
Stockholm University
Department of Molecular Biosciences

Associate Professor Ning Xu Landén
Karolinska Institutet
Department of Medicine

Abstract

MicroRNAs (miRNAs) are short non-coding RNAs which act as post-transcriptional regulators of gene expression. MiRNAs complex with Argonaute proteins to form an RNA-guided silencing complex (RISC). The guide RNA recognizes specific mRNAs via base-pairing to complementary target sites and typically induces repression of the gene product. In animals, target recognition is primarily mediated by the seed region, comprising the first eight nucleotides of the miRNA, but pairing outside the seed (3'-pairing) is sometimes required to achieve substantial repression. The determinants for effective 3'-pairing are not fully understood, limiting the accuracy of computational target site predictions.

To determine the precise base-pairing interactions in miRNA-target complexes, we developed RNA-RNA binding by SHAPE (RABS) (Paper I). We used this technique together with affinity measurements and reporter assays to obtain detailed information about the biochemistry of miRNA-target complexes with a variety of secondary structures, using the conserved *miR-34a* as a model (Paper II). Our results suggest that Argonaure modulates the affinity of the miRNA for its target sites in two directions, strengthening weak RNA:RNA binders and weakening strong ones, but the affinity is only weakly correlated with the amount of repression in cells.

We further explored the combined impact on site efficiency of 3'-pairing and secondary structures in the mRNA, present prior to miRNA binding or formed within the miRNA-target complex (Paper III). Using structural probing of miRNA-target interactions by RABS and reporter assays to measure target repression in cells, we showed that 3'-pairing can compensate for decreased seed binding due to self-pairing in the mRNA. This enables downregulation of sites which would be non-functional if only seed pairing was available.

Finally, we modified nucleotides in *miR-34a* to disrupt pairing beyond the seed, enabling high-throughput screening of target sites for effective 3'-pairing in cells and subsequent identification of favourable and unfavourable structural features (Paper IV). We found that *miR-34a* is differentially sensitive to GU wobble pairs depending on their position in the 3'-pairing helix. It also prefers unpaired nucleotides on the miRNA side over the target site, in contrast with what has previously been observed for other miRNAs. This adds to a growing body of evidence that 3'-pairing preferences vary between different miRNAs.

List of scientific papers

I. Banijamali E, Baronti L, Becker W, Sajkowska-Kozielewicz JJ, Huang T, Palka C, **Kosek D**, Sweetapple L, Müller J, Stone MD, Andersson ER, Petzold K. (2023). RNA:RNA interaction in ternary complexes resolved by chemical probing. RNA 29: 317-329.

II. Sweetapple L, **Kosek DM**, Banijamali E, Becker W, Müller J, Karadiakos C, Baronti L, Guzzetti I, Schritt D, Chen A, Andersson ER*, Petzold K*. (2024). Biophysics of microRNA-34a targeting and its influence on down-regulation. bioRxiv pre-print. doi:10.1101/2024.02.14.580117 (*co-corresponding authors)

III. **Kosek DM**, Banijamali E, Becker W, Petzold K*, Andersson ER*. (2023). Efficient 3'-pairing renders microRNA targeting less sensitive to mRNA seed accessibility. Nucleic Acids Research 51: 11162–11177. (*co-corresponding authors)

IV. **Kosek DM**, Petzold K, Andersson ER. Mapping effective microRNA pairing beyond the seed using abasic mutations. Manuscript.

Contents

1 Introduction ... 1
  1.1 The biogenesis and function of miRNAs ... 1
    1.1.1 Sense and antisense in gene regulation ... 1
    1.1.2 The biogenesis of miRNAs ... 2
    1.1.3 The structure of Argonaute proteins ... 4
    1.1.4 Classification of miRNAs ... 5
    1.1.5 Modes and mechanisms of miRNA regulation ... 7
  1.2 The evolution and physiological roles of miRNAs ... 9
    1.2.1 The origin of RNA interference ... 9
    1.2.2 Evolution and diversification of miRNAs ... 11
    1.2.3 The biological functions of miRNAs ... 13
  1.3 Target selection by metazoan miRNAs ... 16
    1.3.1 Canonical and noncanonical binding sites ... 16
    1.3.2 The molecular mechanism of miRNA binding ... 18
    1.3.3 Pairing beyond the seed ... 20
    1.3.4 Other determinants of site selection and efficiency ... 22
2 Research Aims ... 27
3 Results and Discussion ... 29
  3.1 Paper I ... 29
  3.2 Paper II ... 30
  3.3 Paper III ... 30
  3.4 Paper IV ... 33
4 Concluding Remarks and Future Perspectives ... 37
5 Acknowledgements ... 39
6 References ... 41

List of abbreviations

AGO	Argonaute
EMSA	Electromobility shift assay
FRET	Förster resonance energy transfer
miRNA	microRNA
NMR	Nuclear magnetic resonance
piRNA	PIWI-interacting RNA
RABS	RNA-RNA-binding by SHAPE
RBP	RNA-binding protein
RISC	RNA-induced silencing complex
RNAi	RNA interference
RNA-seq	RNA sequencing
SHAPE	Selective 2′-hydroxyl acylation analyzed by primer extension
siRNA	Small interfering RNA
UTR	Untranslated region

1 Introduction

1.1 The biogenesis and function of miRNAs

1.1.1 Sense and antisense in gene regulation

In 1961, François Jacob and Jacques Monod wrote that "the fundamental problem of chemical physiology and of embryology is to understand why tissue cells do not all express, all the time, all the potentialities inherent in their genome" (Jacob & Monod 1961). For reasons ranging from the high energetic cost of protein synthesis to the need to coordinate physiological processes on time scales from minutes to months, cells regulate every step of gene expression. Nucleic acid sequence complementarity is a fundamental mode of biomolecular interaction across all domains of life, and in the same paper Jacob and Monod speculated that the principle could be used by gene regulatory elements, which at the time had only recently been discovered, to recognize other genetic elements.

In 1993, it was discovered that the *lin-4* gene in the nematode *Caenorhabditis elegans*, a regulator of postembryonic development, does not encode a protein but rather two RNA molecules approximately 22 and 61 nucleotides long (Lee *et al.* 1993, Wightman *et al.* 1993). The longer of the two RNAs was predicted to fold into a hairpin structure and was hypothesized to be a precursor to the shorter one. This short RNA had antisense complementarity to the sites in the 3' untranslated region (3'-UTR) of *lin-14*, which were located in a region of the mRNA previously linked to repression of *lin-14* by *lin-4* (Wightman *et al.* 1991). The interaction between the two RNA molecules caused a temporal reduction in expression of the LIN-14 protein. The proposed model was that the short RNA paired to the complementary sites in the mRNA and induced repression of the gene product by some unknown mechanism.

For some years after this discovery, no trace was found of such a regulatory mechanism in other groups of organisms. This changed with the discovery of *let-7*, another developmental regulator in *C. elegans* (Reinhart *et al.* 2000). While *lin-4* appeared to be exclusive to nematodes, homologs of the *let-7* gene and expressed RNA were soon discovered in human, fly and numerous other bilaterian animals (Pasquinelli *et al.* 2000). These regulatory RNAs were initially named "small temporal RNAs" due to their apparent role in coordinating developmental timing,

but following the discovery of numerous similar RNAs in flies, worms and humans (Lagos-Quintana *et al.* 2001, Lau *et al.* 2001, Lee & Ambros 2001), many of which were not expressed at any specific developmental stage, the name was changed to microRNAs (commonly abbreviated as miRNAs).

1.1.2 The biogenesis of miRNAs

MiRNAs are endogenous noncoding RNA molecules encoded by their own distinct genes (Shabalina & Koonin 2008). They are transcribed as part of longer precursor RNAs featuring a hairpin structure, which are known as pri-miRNAs (Lee *et al.* 2002) (Fig 1.). Although miRNA expression levels are regulated by multiple mechanisms, they typically correlate well with rates of pri-miRNA transcription (Reichholf *et al.* 2019). At least one region of the pri-miRNA folds into a hairpin, which acts as a substrate for the nuclear Microprocessor enzyme complex, composed of two subunits of the RNA-binding protein DGCR8 (called Pasha in flies and nematodes) and one subunit of the endonuclease Drosha (Nguyen *et al.* 2015). Drosha contains two RNase III domains that cut one strand each of the pri-miRNA hairpin, generating stem loops of around 60 nucleotides with a characteristic two nucleotide overhang at the 3' end (Lee *et al.* 2003) (Fig. 1). These structures, termed pre-miRNAs, are exported from the nucleus to the cytoplasm by a specialized exportin (exportin-5) (Yi *et al.* 2003).

Pre-miRNAs are cleaved in the cytoplasm by another endonuclease called Dicer, which also possesses two RNase III domains (Bernstein *et al.* 2001, Grishok *et al.* 2001, Hutvágner *et al.* 2001). Dicer removes the loop of the hairpin structure to create a duplex consisting of the miRNA and its so-called passenger strand (often referred to as "miRNA*") (Fig 1.). As with Drosha, cleavage of the hairpin stem by Dicer results in unpaired terminal nucleotides at the 3' end (Lee *et al.* 2003, Zhang *et al.* 2004). The dangling ends on both sides of the duplex produced by the combined action of the two endonucleases are characteristic of mature miRNAs and are one of the features used to identify candidate miRNAs (Ruby *et al.* 2006).

In the final step of the miRNA biogenesis pathway, the guide RNA is loaded into an Argonaute protein, forming an RNA-induced silencing complex (RISC) (Fig. 1). Argonautes are the effectors of miRNA-mediated gene repression once the guide RNA has bound to a target mRNA. The miRNA duplex is loaded into Argonaute with help from the chaperones HSC70 and HSP90, which ease the binding of the guide

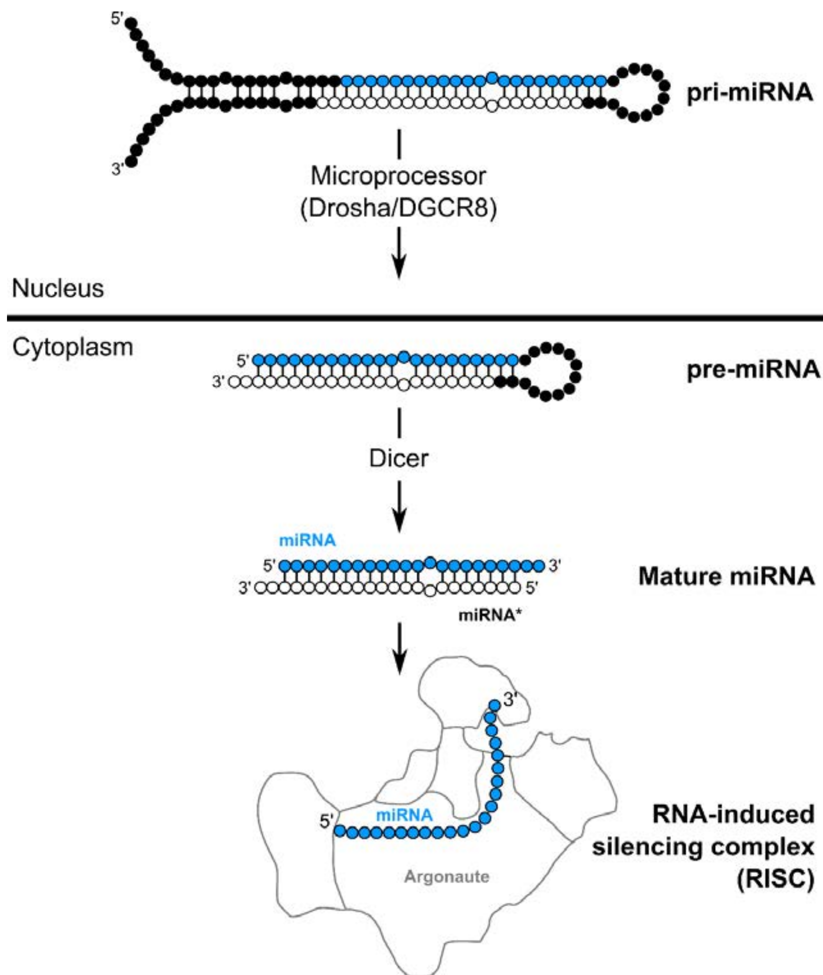

Figure 1. The miRNA biogenesis pathway in animals. Transcripts from miRNA genes are processed by the Microprocessor enzyme complex and exported from the nucleus. In the cytoplasm, the hairpin loop is removed by Dicer and one of the two RNA strands is incorporated into an Argonaute protein to form the RNA-induced silencing complex.

RNA by making Argonaute adopt a high energy open conformation, in an ATP-consuming process (Iwasaki *et al.* 2010). While either strand of the duplex can in principle become the guide strand, most miRNAs are strongly biased towards one of the strands (Chiang *et al.* 2010). Strand choice is determined by a pocket in Argonaute which binds the 5' monophosphate of the guide RNA. Argonaute prefers a uridine or adenosine residue at this position (Frank *et al.* 2010, Suzuki *et al.* 2015), and favours the strand with less thermodynamically stable binding at the 5' end (Khvorova *et al.* 2003, Schwarz *et al.* 2003).

The loading of the miRNA into Argonaute appears to be the rate-limiting step of the biogenesis pathway in many contexts (Diederichs & Haber 2007, Reichholf *et al.* 2019), possibly reflecting the requirement of chaperones and energy from ATP for efficient loading. Thus, the degree of Argonaute binding may be a better indicator of the regulatory capacity of the miRNA than the total concentration of the guide RNA in the cell (Flores *et al*. 2014). After loading of the guide RNA, RISC can exert its effects on gene expression. Most miRNAs bound to an Argonaute protein are stable for a long time, with half-lives of days (Bail *et al.* 2010, Gantier *et al.* 2011).

1.1.3 The structure of Argonaute proteins

The Argonaute proteins received their name from a mutant phenotype in *Arabidopsis thaliana* which gave the plant's leaves a tentacle-like appearance, likened by Bohmert *et al*. (1998) to the small octopus *Argonauta argo*. Shortly thereafter they were shown to be the catalytic component of the eukaryotic RNA interference pathway, cleaving RNA molecules base-paired to an Argonaute-bound guide RNA (Song *et al*. 2004, Liu *et al.* 2004). They were also shown to complex with miRNAs to direct silencing of endogenous transcripts, without endonuclease activity (Hutvágner & Zamore 2002). Argonaute proteins are found in prokaryotes as well as eukaryotes, although only about a third of archaeal genomes and a tenth of bacterial genomes encode a family member (Swarts *et al*. 2014a).

Argonaute proteins belong to the PIWI superfamily, defined by the presence of the P element-induced wimpy testis (PIWI) domain (Swarts *et al*. 2014a). All eukaryotic Argonautes also contain an N-terminal (N) domain, a PIWI-Argonaute-Zwille (PAZ) domain and a middle (MID) domain, as well as two linkers (L1 and L2). The proteins have a bilobed architecture, with the PIWI and MID domains in one lobe and N and PAZ in the other. The L1 linker connects the N and PAZ domains, while L2 connects PAZ and MID. While eukaryotic Argonaute is traditionally said to possess four domains and two linkers, it has been argued based on data from crystal structures that the linkers are structured enough to be considered domains in their own right (Nakanishi 2022).

The MID domain contains a nucleotide-binding pocket, where a number of conserved amino acids interact with the 5' terminal phosphate group of the miRNA (Ma *et al.* 2005, Parker *at al.* 2005, Boland *et al.* 2010, Frank *et al.* 2010). As noted

above, the binding preferences of this pocket are important for miRNA strand choice. The MID domain also promotes target binding by stacking the nucleotides at the 5' end of the guide RNA in a helical conformation (Schirle *et al.* 2014). The PIWI domain contains the active site of catalytically competent Argonaute paralogs (Song *et al.* 2004, Parker *et al.* 2004, Yuan *et al.* 2005). The PAZ domain anchors the 3' end of the guide RNA through interactions with the backbone of residues 20-21 (Ma *et al.* 2004, Lingel *et al.* 2004, Wang *et al.* 2008, Sheng *et al.* 2014). While inessential for binding, this interaction protects the guide RNA from degradation (Hur *et al.* 2013). Finally, the N domain is involved in target cleavage (Faehnle *et al.* 2013, Hauptmann *et al.* 2013) and the dissociation of cleaved strands (Faehnle *et al.* 2013, Kwak & Tomari 2012). The majority of the contacts between the guide RNA and Argonaute are hydrogen bonds and salt linkages to the sugar-phosphate backbone of the RNA, which are not affected by the identity of the nucleobases (Schirle *et al.* 2014). This enables RISC to accommodate guide RNAs in a sequence-independent way.

Humans and other vertebrates have four Argonaute paralogs, numbered AGO1 to AGO4. Of these, only AGO2 shows robust catalytic activity (Liu *et al.* 2004, Meister *et al.* 2004), although AGO3 is also capable of endonucleolytic cleavage under certain conditions (Park *et al.* 2017). Human cells do not appear to contain any specialized sorting system for miRNAs to be loaded into specific Argonaute paralogs and the process has been proposed to be random (Hafner *et al.* 2010, Dueck *et al.* 2012, Wang *et al.* 2012). Indeed, sequencing of Argonaute-bound small RNAs show that most guide RNAs are bound to some extent to all paralogs (Azuma-Mukai *et al.* 2008). Nevertheless, sometimes strong biases for certain miRNAs over others are observed for different paralogs (see examples in Nakanishi 2022).

1.1.4 Classification of miRNAs

The primary determinant of target recognition for animal miRNAs are nucleotides 2 to 8 of the guide RNA (numbered from the 5' end), known as the seed region. This is the part of the guide that is preorganized for binding by the Argonaute MID domain in RISC. Pairing to the seed nucleotides alone is often sufficient for effective target repression (Doench & Sharp 2004, Brennecke *et al.* 2005, Lim *et al.* 2005). MiRNAs are divided into families based on the identity of the nucleotides in the seed, with members of the same family targeting overlapping sets of genes.

Although members of a family typically share an evolutionary history, arising for instance from duplications of an ancestral gene, this is not necessarily the case. Two separate lineages of miRNAs may converge on the same seed sequence. Conversely, a single mutation in the seed can shift a miRNA to a new family with an entirely different set of potential binding sites.

MiRNA families are numbered in their order of discovery (Ambros *et al.* 2003). For the most part, the same name has been assigned to orthologs in different species, so that, for example, eleven of twelve human orthologs of the *C. elegans let-7* are also named *let-7* (Bartel 2018). Letter suffixes (*a*, *b*, *c*, ...) are used to distinguish paralogs within a species encoding similar mature miRNAs. MiRNAs produced from the 5' strand of the pre-miRNA hairpin are given the suffix *5p*, while the opposite strand is denoted *3p*. A few of the earliest miRNAs to be discovered, such as *lin-4*, *let-7* and *lsy-6*, are still referred to by their original names derived from mutant phenotypes. Because cells contain many types of small RNAs, both functional molecules and fragments of degraded longer RNAs, accurate identification of miRNAs in sequencing datasets is challenging and the number of mature miRNAs listed in different databases vary considerably. According to Fromm *et al.* (2015), there are 519 confidently annotated miRNA genes in humans.

More than one mature miRNA can be produced from the same gene. As noted above, either of the two strands in the duplex produced by Dicer from a pre-miRNA may be loaded into Argonaute, each with its own set of target genes. Although one strand usually predominates, miRNA* molecules can sometimes regulate targets with physiologically relevant consequences (Okamura *et al.* 2008, Kuchenbauer *et al.* 2011). Furthermore, the ratio of miRNA/miRNA* incorporation in Argonaute can change over the course of evolution, in a process termed "arm switching" (Okamura *et al.* 2008). Strand preference also varies between tissues and developmental time points, indicating the presence of determinants other than thermodynamic stability and 5' nucleotide preference (Ro *et al.* 2007, Chiang *et al.* 2010).

The same gene can also produce multiple miRNA isoforms (sometimes named isomiRs). This is caused by heterogenous cleavage by Drosha or Dicer, resulting in mature miRNAs with variable 5' or 3' ends. 5' isomiRs, where the sequence variation alters the seed region, are less common than 3' isomiRs (Chiang *et al.* 2010, Berezikov *et al.* 2011), but the dominant 5' isoform of a given miRNA can change during evolution through so-called "seed shifting" (Wheeler *et al.* 2009). The

impact of the 3' end of the miRNA on target selection is smaller, but some 3' isomiRs are developmentally regulated and have physiological effects (Fernandez-Valverde *et al.* 2010).

1.1.5 Modes and mechanisms of miRNA regulation

There are two distinct modes of gene silencing mediated by RISC. If there is extensive base-pairing between the guide RNA and its target site and the guide is bound to a catalytically active Argonaute paralog, the mRNA can be endonucleolytically cleaved, resulting in its degradation as the cell recognizes the shortened RNA as abnormal (Hutvágner & Zamore 2002, Liu *et al.* 2004, Meister *et al.* 2004, Yekta *et al.* 2004). This appears to be the common mode of miRNA regulation in plants (Jones-Rhoades *et al.* 2006) and possibly sea anemones (an outgroup to bilaterian animals) (Moran *et al.* 2014). By contrast, in human cells miRNA-directed slicing has only been observed for a small number of targets (e.g. Yekta *et al.* 2004). Instead, miRNA binding causes degradation of the mRNA by pathways other than direct cleavage, and sometimes repression of translation without an associated decrease in mRNA levels.

The relative contributions of mRNA degradation and translational repression are disputed and may be context-dependent. Experiments where miRNAs are overexpressed or knocked out in cultures of post-embryonic cells indicate that mRNA decay predominates (Baek *et al.* 2008, Hendrickson *et al.* 2009, Guo *et al.* 2010). One study estimates that mRNA decay accounts for 66% to >90% of steady-state downregulation of genes by miRNAs (Eichhorn *et al.* 2014). Nevertheless, other studies demonstrate translational repression without substantial mRNA decay (e.g. Jin *et al.* 2017, Mayya *et al.* 2021). Translational repression sometimes precedes mRNA degradation (Béthune *et al.* 2012, Djuranovic *et al.* 2012), but this does not necessarily imply a mechanistic link and might instead reflect kinetic differences.

To induce degradation of mRNAs, miRNAs make use of enzymes involved in the 5'-to-3' mRNA decay pathway (reviewed in a miRNA context by Jonas & Izaurralde 2015). The first step of miRNA decay in this pathway is the removal of the poly(A) tail that is added to the 3'-end of eukaryotic mRNAs following transcription. This deadenylation reaction is initiated by the PAN2-PAN3 complex and continued by the CCR4-NOT complex. Next, the m^7G cap at the 5' end of the mRNA is removed by decapping protein 2 (DCP2). Numerous other proteins stimulate the decapping

reaction, including DEAD box protein 6 (DDX6), which also interacts with the CCR4-NOT complex. The deadenylation and decapping steps are followed by degradation of the mRNA by 5'-to-3' exoribonuclease 1 (XRN1).

Subunits of the different catalytic complexes directly interact throughout the process, ensuring tight coupling between deadenylation and decoupling and swift recruitment of the exonuclease to decapped mRNAs, thus preventing the accumulation of decay pathway intermediates in cells. The coupling is disrupted in some contexts, including oocytes, early embryos and possibly neuronal cells (Jonas & Izaurralde 2015). This results in the accumulation of translationally repressed mRNAs with 5' caps but no poly(A) tails. These mRNAs could in principle be translated again following cytoplasmic polyadenylation, enabling potentially reversible miRNA silencing (Schratt *et al*. 2006, Muddashetty *et al*. 2011, Subtelny *et al*. 2014).

The trinucleotide repeat-containing 6 (TNRC6) family of proteins are a key intermediate between Argonautes and the mRNA decay machinery, as shown by the finding that depletion of these proteins inhibits miRNA-mediated silencing (Rehwinkel *et al*. 2005, Behm-Ansmant *et al*. 2006). The number of family members varies between animal groups and their structures diverge (Jonas & Izaurralde 2015). In vertebrates there are three paralogs: TNRC6A, TNRC6B and TNRC6C. All three have a similar domain organization, with an amino-terminal Argonaute-binding domain and a carboxy-terminal silencing domain. In humans, TNRC6 can bind up to three Argonaute proteins, while each Argonaute can bind one TNRC6 protein (Elkayam *et al.* 2017).

TNRC6 proteins directly interact with the adaptor protein PAN3, recruiting this protein and its catalytic partner PAN2 to the site of the bound RISC. It also interacts with a subunit of the CCR4-NOT complex. Lastly, TNRC6 interacts with cytoplasmic poly(A)-binding protein (PABP), which in turns interacts with translation factors to stimulate initiation of protein synthesis. Argonaute, TNRC6, the decapping proteins and XRN1 localize to cytoplasmic domains known as P-bodies, where proteins involved in mRNA decay and translational repression accumulate.

MiRNA binding can sometimes lead to repression of translation without corresponding mRNA decay. The mechanisms involved are less understood than the ones underlying mRNA degradation, but several hypotheses have been proposed. One possibility involves the TNRC6-mediated interaction between

RISC and PABP. PABP bound to poly(A) tails stabilizes the association of the translational initiation complex eIF4F with the 5' cap of mRNAs. This interaction is thought to lead to the formation of a closed-loop structure facilitating translation and ribosome recycling (Sonenberg & Hinnebusch 2009). RISC promotes dissociation of PABP, inhibiting translation initiation by disrupting the closed-loop structure (Moretti *et al.* 2012). Notably, the motif responsible for binding PABP is conserved in TNRC6 proteins (Jonas & Izaurralde 2015). Another proposed mechanism involves CCR4-NOT, which also interacts with RISC via TNRC6. As mentioned earlier, this complex interacts with DDX6, which can promote miRNA-mediated downregulation even when deadenylation is blocked (Chen *et al.* 2014, Mathys *et al.* 2014, Rouya *et al.* 2014).

It is important to note that these different silencing mechanisms are not mutually exclusive and may take place simultaneously in the cell, possibly with different kinetics. The relative contributions of each mechanism could further depend on the cell type, the developmental stage of the tissue and possibly the characteristics of individual target mRNAs. Post-translational modifications of interacting proteins can also affect their binding properties and affinities; for example, phosphorylation of human AGO2 increases its interaction with GW182 proteins (Horman *et al.* 2013). Interestingly, while miRNAs repress translation, active translation of an mRNA can also stimulate miRNA binding as the movement of ribosomes disrupts local secondary structures, making target sites more accessible for the miRNA (Ruijtenberg *et al.* 2020). It is likely that the steady-state silencing of each individual target gene mediated by a miRNA is modulated by many such feedback mechanisms.

1.2 The evolution and physiological roles of miRNAs

1.2.1 The origin of RNA interference

MiRNAs have been identified in several eukaryotic lineages, including animals, land plants, green algae, brown algae, filamentous fungi and slime molds (Bartel 2018). The miRNA pathway is part of a network of regulatory and immunological systems called RNA interference (RNAi). This network also comprises small interfering RNAs (siRNAs), which silence expression of foreign genes, and PIWI-interacting RNAs (piRNAs), which are predominantly found in animal germline cells, where they repress the activity of transposable elements (Baulcombe 2004, Saito *et al.* 2006,

Vagin *et al.* 2006). The miRNA and siRNA silencing pathways both involve a guide RNA forming a complex with an Argonaute protein to target specific RNA molecules through base complementarity. The two systems differ in that siRNAs typically require fully complementary binding sites and induce Argonaute-catalyzed cleavage of the target strand, neither of which is generally true of miRNAs. Another important difference is that miRNAs are encoded by distinct genes of endogenous origin, as opposed to being transcribed from genomically integrated transposons or generated directly from exogenous RNA molecules.

All RNAi pathways ultimately depend on three key proteins: AGO-PIWI-like protein, Dicer-like protein (typically consisting of RNase III and helicase domains) and RNA-dependent RNA polymerase (RdRP) (Shabalina & Koonin 2008). These core proteins have undergone numerous duplications over the course of eukaryotic evolution and the resulting paralogs underlie the diverse functions of RNAi in the present day. Their distribution among extant organisms suggests that all three were present in the last eukaryotic common ancestor (LECA), which is therefore likely to have featured some form of RNA-mediated silencing. Moreover, the distribution of Argonaute and PIWI proteins indicate that this ancient duplication precedes the emergence of current eukaryotic supergroups, meaning that both the siRNA and piRNA pathways may have been present in some form in LECA (Shabalina & Koonin 2008).

All three of the core RNAi proteins have prokaryotic homologs. Argonaute has distant relatives among both bacteria and archaea, with the eukaryotic form seemingly closest to its archaeal counterpart (Aravind *et al.* 2000). Meanwhile, Dicer displays a curious mix of archaeal and bacterial roots. The helicase domain is related to the archaeal Superfamily II helicases (Hef proteins) (Aravind *et al.* 1999), whereas the RNAase III nuclease domain is closest to bacterial homologs involved in rRNA and mRNA processing (MacRae & Doudna 2007). The fusion of these two domains, not seen in prokaryotes but present in all eukaryotic supergroups, may have been an important early step in the emergence of RNAi as a functional system.

Modern prokaryotic Argonaute proteins play a role in immune responses, using short DNA or RNA guide strands to recognize foreign DNA elements (Olovnikov *et al.* 2013, Swarts *et al.* 2014b). The switch to RNA guides and predominantly RNA targets in eukaryotes may have been a response to the much higher prevalence of RNA viruses in the eukaryotic virome (Koonin *et al.* 2015, Koonin 2017). This

would have set the stage for the evolution of miRNAs, with the cytoplasmic Argonaute-based defense system being coopted for the regulation of endogenous gene expression. The involvement of prokaryotic Argonaute proteins in immunity, the mechanistic differences between miRNA systems in different eukaryotic lineages (such as the divergent approaches to target selection observed in animals and plants) and the likely presence of RdRP (which is involved in siRNA but not miRNA pathways) early in evolution all support cell defense as the ancestral function of RNAi.

1.2.2 Evolution and diversification of miRNAs

It is unknown if miRNAs evolved independently in separate lineages or if some form of RNAi-based gene regulation was already present prior to the diversification of eukaryotes. The miRNA biogenesis pathway differs between animals and plants, with the latter lacking the Microprocessor enzyme complex, composed of Drosha and Pasha/DGRC8 (see section 1.1.2). This complex was thought to have emerged at the advent of the animal kingdom (Grimson *et al*. 2008), but it was subsequently identified in unicellular sister lineages to animals, indicating that the key components of the miRNA biogenesis pathway were present before multicellularity (Bråte *et al*. 2018). The miRNA pathway appears to have been lost in some non-bilaterian lineages (Grimson *et al.* 2008, Moroz *et al.* 2014). Only one extant animal seed family (*miR-99/100*) predates the bilaterian lineage, as a closely related miRNA is present in sea anemones (Grimson *et al.* 2008).

Novel miRNA genes can arise by multiple mechanisms (Berezikov 2011). MiRNA gene families, which share the same seed sequence, often represent paralogous sequences arising from gene duplication (Hertel *et al*. 2006). These can be divided into local duplication, where the duplicated genes typically remain in the same transcript, and non-local duplication, where the new miRNA is found in a different location, often on another chromosome. Most non-local miRNA duplications in vertebrates are associated with genome duplication events that have taken place during vertebrate evolution (Heimberg *et al*. 2008, Gu *et al*. 2009). Duplication events may be followed by mutations in the guide RNA sequence, leading to the emergence of new miRNAs with altered targeting properties.

The introns of protein-coding genes frequently contain genes encoding miRNAs (Rodriguez *et al*. 2004). Almost half of human miRNA genes are found in introns in the same orientation as the host gene (Campo-Paysaa *et al*. 2011). A plausible

explanation for this observation is that intronic sequences are transcribed but rarely produce functional molecules, enabling new active RNAs to arise without the necessity of evolving a new promoter. Notably, evolutionarily more recent, species-specific miRNAs are more commonly found in introns than older miRNAs (Campo-Paysaa *et al.* 2011). Other miRNAs appear to originate from transposable elements (Piriyapongsa *et al.* 2007). MiRNAs derived from transposons are generally less conserved and more lineage-specific compared to miRNAs with other evolutionary backgrounds, consistent with transposable elements as a continuous source of new miRNA genes (Piriyapongsa *et al.* 2007, Yuan *et al.* 2011).

The human genome is estimated to encode hundreds of thousands of pre-miRNA-like hairpins, which are a potential source of novel miRNAs (Bentwich *et al.* 2005). Functional miRNAs must be processed by Drosha and Dicer after transcription. Accurate cleavage depends on the precise secondary structure of the hairpin, which in *Drosophila* is only observed for a small number of gene products (Lu *et al.* 2008). However, there is evidence for transitional forms on the evolutionary path towards new miRNAs, in the form of small RNA reads displaying partial signatures of Drosha and Dicer processing, among numerous predicted hairpins with matching reads in sequencing data (Berezikov *et al.* 2011).

Any new miRNA faces the problem of avoiding undesirable gene repression via interactions with target sites that match the seed region by chance. In the transcriptional control model of miRNA evolution proposed by Chen & Rajewsky (2007), new miRNAs are initially expressed at low levels to limit the deleterious effects of accidental targeting. Over time, unfavourable target interactions are purged by natural selection, after which the expression of the miRNA can increase. Consistent with this model, there is a positive correlation between expression levels and evolutionary age of miRNAs (Berezikov *et al.* 2006, Lu *et al.* 2008). Furthermore, detectable miRNA-mediated repression requires relatively high concentrations of the miRNA in the cell (Bosson *et al.* 2014, Denzler *et al.* 2016), suggesting that some lowly expressed annotated miRNAs may not yet have acquired biologically significant target interactions.

1.2.3 The biological functions of miRNAs

After their discovery, it soon became clear that miRNAs are abundant and important gene regulators, but the more fundamental reasons for the evolution and persistence of this particular form of regulation remain a matter of debate. Perhaps the most striking observation about miRNAs is the small degree of repression they typically induce. Most target genes are downregulated by less than 50% at the protein level even when the miRNA is overexpressed, which should be within the functional concentration range for many of them.

MiRNAs were first identified in studies of developmental timing and loss of essential components of the system causes severe defects in embryogenesis in all animal species studied so far. Removal of two *C. elegans* Argonaute paralogs exclusive to the miRNA pathway results in non-viable embryos (Vasquez-Rifo *et al.* 2012). In *Drosophila*, loss of *ago1* leads to non-viable embryos with nervous system defects (Kataoka *et al.* 2001). Loss of Dicer in zebrafish causes defects in organogenesis and brain development (Giraldez *et al.* 2005). Notably, the neuronal defects could to a significant extent be rescued by injecting *miR-430* duplex in the embryos. Loss of *Dicer1* in mice similarly leads to non-viable embryos (Bernstein *et al.* 2003). Although some observed effects may be due to disruption of miRNA-independent functions of these proteins, it seems clear that miRNAs are vital for embryonic development in animals.

Assigning functions to specific miRNAs is challenging, partly because a single miRNA can target a large number of genes and partly because members of the same seed family can have redundant functions, so that phenotypic effects are observed only after the removal of all members (see e.g. Alvarez-Saavedra & Horvitz 2010). Furthermore, miRNAs are frequently expressed with high temporal specificity and may have relevant physiological effects only in certain cell types, meaning that even non-redundant functions can be challenging to pinpoint experimentally (Alberti & Cochella 2017). Nonetheless, specific miRNA-target interactions can be of vital importance: for example, in *C. elegans*, regulation of the *nhl-2* gene by the *mir-35* family is required for viability (McJunkin & Ambros 2017).

The role of miRNAs must be considered in the broader context of post-transcriptional gene regulation. The eukaryotic separation of mRNA and protein synthesis in time and space enables regulation at multiple stages: the steady-state levels of a protein depends on the level of transcription, nuclear processing and transport, cytoplasmic localization, translation initiation, and mRNA and

protein degradation. Since functionally related genes in eukaryotes are often not chromosomally adjacent, co-regulation of proteins which form a complex or function in the same pathway must be achieved post-transcriptionally. Accordingly, there are examples of coordinated export and cytoplasmic grouping of mRNA subpopulations through combinatorial action by multiple regulatory factors (Keene 2007). Transcription is in turn controlled by regulatory feedback loops capable of generating diverse temporal expression profiles (Alon 2007). MiRNAs act within such modules and networks to control the stability and translation of individual transcripts, which can serve a number of purposes depending on the design of the system.

Inui *et al.* (2010) considers the possible roles of miRNAs in signal transduction. This is a promising area for miRNA regulation since some pathways are dose sensitive enough that even small fluctuations in the concentration of a signal component can have significant downstream effects. The authors note that different cellular miRNA populations may help explain the variety of outputs that can be achieved by the same extracellular signal, and that the additional layer of regulation added by comparatively fast-evolving miRNAs could enable the frequent repurposing of same signaling cascades seen throughout evolution. For example, miRNAs can raise or lower a cell's response threshold for a signal, by targeting signaling mediators or inhibitors respectively. They can also mediate crosstalk between separate pathways (e.g. if signal A induces miRNA expression to regulate signal B), or coordinate responses by acting simultaneously on two pathways.

If a miRNA targets an activator and a repressor in the same pathway, it can help reduce sensitivity to fluctuations and achieve optimal relative levels of components. In this way, miRNAs can increase the robustness of a system, defined as the ability to generate the same phenotype in the face of genetic and environmental variation. The observed global downregulation of miRNAs in tumour samples compared to healthy tissues (Lu *et al.* 2005) indicates that buffering of signal elements may be an important function. More generally, a miRNA-mediated decrease in translational output combined with an increase in transcription (that is, a reduction of protein output per mRNA) could in principle lower variability (noise) in the expression of a gene (Bartel & Chen 2004). It should be noted though that the addition of a regulatory component to a system also increases intrinsic noise since the expression of the regulator is itself subject to random variation (Schmiedel *et al.* 2015).

An illustrative case study of robustness is provided by the role of *miR-9a* in *D. melanogaster* development. Flies lacking this miRNA are viable but vary more than wild-type flies in the number of sense organs (Li *et al*. 2006). These organs develop from a sensory organ precursor cell, partially controlled by the transcription factor Senseless. By targeting this transcription factor, *miR-9a* raises the activation threshold which must be overcome to create such a precursor cell. When the miRNA is removed, the random variability of precursor cell appearance increases, resulting in greater phenotypic variability in the population. It has been proposed that this kind of reinforcement of phenotypes specified by other factors is a primary function of miRNAs (Hornstein & Shomron 2006). Farh *et al.* (2005) observed that miRNAs which are induced at different time points during mammalian development have predicted target genes that are often preferentially expressed in the prior developmental stage, suggesting an important role for miRNAs in shaping the transition to a new gene expression program.

It has been hypothesized that by decreasing stochastic variability in development, miRNAs may have significantly contributed to the evolution of complex animal body plans (Sempere *et al*. 2006, Peterson *et al*. 2009). This hypothesis is supported by the finding that the number of miRNA families expanded greatly in early metazoan evolution (Hertel *et al*. 2006, Prochnik *et al*. 2007). It has been proposed that the addition of new miRNAs to regulatory networks could help explain why neither genome size nor the number of protein-coding genes correlates with organismal complexity, two observations known as the C-value and G-value paradoxes respectively (Taft *et al*. 2007, Berezikov 2011). Others have challenged this hypothesis, arguing that there is no clear relationship between miRNA count and body complexity across the animal kingdom (Moran *et al*. 2017). A recent study of cephalopod evolution revealed multiple novel miRNAs with conserved target sites expressed in neuronal tissues, comparable with the expansion of the miRNA repertoire seen in vertebrates, suggesting a link between miRNAs and the evolution of complex animal brains (Zolotarov *et al*. 2022).

1.3 Target selection by metazoan miRNAs

1.3.1 Canonical and noncanonical binding sites

To understand the biological roles of miRNA-mediated gene regulation, and to be able to predict functionally important miRNA-target interactions, it is necessary to discover the determinants for target site selection and efficiency. The interactions between a miRNA and other RNA molecules are altered by Argonaute proteins and consequently cannot be predicted by simply applying the normal rules governing base-pairing and dissociation between RNAs. Furthermore, the regulatory capacity of a miRNA binding site is determined not only by the structure and affinity of the miRNA-target complex but also by other factors such as competition with internal mRNA secondary structure and the position of the site within the transcript.

As the functional roles of specific miRNA nucleotides after incorporation into RISC depend on their position in the sequence, the residues are numbered from the 5' end starting with the prefix g for guide (g1, g2, ...). Target nucleotides are correspondingly numbered from the 3' end after the guide nucleotides to which they are paired, starting with the prefix t for target, so that residue t2 is paired with g2 and so on.

Target recognition in animals relies strongly on the seed region, comprising at minimum nucleotides 2-7 of the miRNA. Studies of the global impact of miRNA overexpression on the cellular transcriptome and proteome identified four types of seed-binding sites as mediators of the greater part of miRNA-induced gene repression (Grimson *et al.* 2007, Baek *et al.* 2008). These sites, referred to as the canonical seed-binding sites, feature full complementarity to g2-g7 (Fig. 2). They are divided into 6mer (g2-g7 match), 7mer-A1 (g2-g7 match with an adenosine at t1), 7mer-m8 (g2-g8 match) and 8mer (g2-g8 match with an adenosine at t1) sites, in increased order of average efficiency. The first nucleotide of the guide RNA is buried within Argonaute (Ma *et al.* 2005, Parker *et al.* 2005) and consequently unable to form base-pairing interactions with the target. Instead, the base at t1 interacts with Argonaute, which has a pocket that specifically recognizes adenosine (Schirle *et al.* 2015), resulting in a preference for this base at the 3'-end of binding sites. Two types of offset 6mer sites are sometimes included among the canonical sites; the 6mer-A1 (g2-g6 match with an adenosine at t1) and the

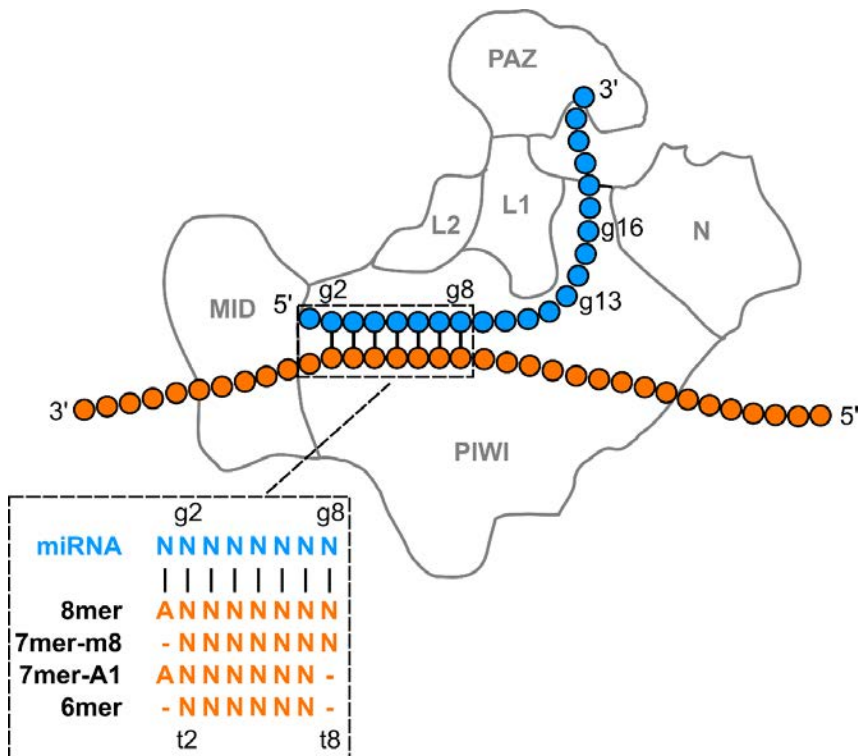

Figure 2. Schematic structure of a miRNA (blue) in complex with an Argonaute protein (grey), bound to a target site in an mRNA (orange). Canonical seed-binding sites are shown in the box (N represents an arbitrary nucleotide).

6mer-m8 (g3-g8 match) (McGeary *et al.* 2019). Outside the seed, the miRNA sequence is frequently divided into the central (g9-g12), supplementary (g13-g16) and tail (g17 and beyond) regions (Wee *et al.* 2012), although there is less agreement regarding the precise definition of these regions (Fig. 2).

Not all binding sites require full pairing to the seed. One class of noncanonical binding sites are the 3'-compensatory sites, where mismatches or nucleotide bulges in the seed are tolerated because of extensive pairing to the remainder of the miRNA (Vella *et al.* 2004, Brennecke *et al.* 2005). High-throughput crosslinking experiments show that RISC frequently binds to sites that lack six contiguous seed pairs, with such noncanonical sites comprising about half of the crosslinking interactions (Chi *et al.* 2009, Hafner *et al.* 2010, Loeb *et al.* 2012, Helwak *et al.* 2013, Grosswendt *et al.* 2014). However, these noncanonical interactions do not mediate detectable mRNA repression in miRNA overexpression and knockdown

experiments (Agarwal *et al.* 2015), suggesting that RISC does not remain bound long enough to the majority of these sites to initiate degradation of the transcript.

Most effective miRNA target sites reside in the 3'-UTRs of transcripts (Farh *et al.* 2005, Lewis *et al.* 2005, Lim *et al.* 2005). Grimson *et al.* (2007) found that potential miRNA binding sites in 5'-UTRs conferred no detectable downregulation, while the efficiency of sites in protein-coding sequences was marginal compared to sites in 3'-UTRs. Nevertheless, there are examples of sites in protein-coding sequences mediating substantial downregulation (Hausser *et al.* 2013). A recent study suggests that they may require more extensive binding than sites in 3'-UTRs (Sapkota *et al.* 2023).

1.3.2 The molecular mechanism of miRNA binding

The primacy of seed binding in miRNA target site selection is supported by structural as well as functional evidence. A crystal structure of human AGO2 with a bound guide RNA shows the 5' end of the guide anchored in the MID domain of the protein, with the seed residues preorganized in a helical conformation to enable target binding (Schirle *et al.* 2014). Residues g12-g20 extend from the centre of the protein to the PAZ domain, which anchors the 3' end of the guide RNA (Fig. 2). The nucleotides from g14 to g18 are held inside a channel between the PAZ and N domains. Interactions with amino acids in this channel disrupt base stacking from g14 to g18 in guide RNAs not paired to a target RNA.

Comparison of crystal structures of RISC with and without a bound target RNA provides structural insights into its binding mechanism (Schirle *et al.* 2014). When the guide RNA is not bound to a target, it is kinked at the end of the seed region (Nakanishi *et al.* 2012, Schirle *et al.* 2012, Faehnle *et al.* 2013, Nakanishi *et al.* 2013). This is stabilized by helix-7 of the L2 domain, which inserts a hydrophobic amino acid between g6 and g7. When the seed binds a target RNA (11 nucleotides in the crystal structure), RISC undergoes a conformational change where helix-7 shifts its position to interact with the minor groove of the newly formed guide-target helix. This avoids steric clashes with target nucleotides t6 and t7 and relaxes the kink in the guide, allowing g6 and g7 to adopt the helical conformation required for target binding. The helix-7 interaction with the minor groove further stabilizes the open confirmation. Mismatches and GU wobbles distort the shape of the minor groove, which destabilizes the open conformation and instead promotes

expulsion of the target from the complex. This explains why GU wobbles are not well tolerated in the seed despite their thermodynamic stability (Wee *et al.* 2012).

A pair of single-molecule FRET studies shed further light on how RISC initiates pairing to a target (Salomon *et al.* 2015, Chandradoss *et al.* 2015). AGO2 diffuses along RNA molecules, scanning for matches to g2-g4, which as described above are preorganized in a helical conformation (Chandradoss *et al.* 2015). If full seed pairing is available, this transient interaction is stabilized, with binding to the remainder of seed slowing dissociation and thus giving RISC time to recruit the molecular machineries of translation repression and mRNA decay. Once RISC has established a full seed-binding interaction, it dissociates more slowly than what would be expected for a seven nucleotide RNA helix by itself (Salomon *et al.* 2015). Argonaute also appears to increase the sensitivity of the complex to the length of the seed, as RISC remained paired to a 7mer site more than 100 times longer than a 6mer, whereas a naked seven base-pair RNA duplex melts only around 10 times slower than a six base-pair duplex (Chandradoss *et al.* 2015).

Regions of mRNAs that lack antisense complementarity to the guide RNA are able to weakly interact with RISC (Ameres *et al.* 2007). A later study employing a single-molecule approach demonstrated that even after removal of all miRNA binding sites from an mRNA, around 10% of RISC binding is retained (Kobayashi & Singer 2022). These findings suggest that RISC can interact with mRNAs in a non-specific, guide-independent manner. Based on this, Nakanishi (2022) proposed a model for miRNA binding which involves interactions between RISC and mRNAs mediated by the protein subunit rather than the guide RNA. The exteriors of Argonaute proteins feature several positively charged patches, which could interact with the negatively charged mRNA sugar-phosphate backbone through sequence-independent ionic interactions. As these patches are scattered across the Argonaute surface, multiple simultaneous interactions with the same or different mRNAs may be possible. Moreover, Argonaute paralogs have different electrostatic potential maps, potentially affecting their affinities for particular mRNAs. These sequence-independent interactions would have to be short-lived, as an excessively strong association would delay mRNA release and slow down the target scanning process, but they would be stabilized if the miRNA seed could initiate pairing to a complementary site. In this case, the Argonaute-target interaction would compete with the guide-target interaction, creating a small energy barrier that must be overcome to achieve proper pairing. If the nature of these putative Argonaute-mRNA interactions were distinct for the four human

paralogs, it could help explain observations like the different recognition of flanking regions of binding sites by AGO2 and AGO3 (Park *et al.* 2017). For the present, there is no direct evidence of interactions between mRNAs and the exterior surfaces of Argonautes.

1.3.3 Pairing beyond the seed

While binding to the seed alone is sometimes sufficient for target repression, the degree of downregulation is variable and not all seed matches confer detectable repression (Doench & Sharp 2004, Brennecke *et al.* 2005, Didiano & Hobert 2006). For some sites, pairing to nucleotides outside the seed, referred to here as 3'-pairing, is necessary to achieve full target repression. The frequency and functional importance of such pairing remains disputed. Cross-linking experiments in cells indicate that 3'-pairing is common (Moore *et al.* 2015, Broughton *et al.* 2016). On the other hand, a study of microRNA target site evolution estimated that only 4.9% ± 1.1% of all preferentially conserved miRNA sites have preferentially conserved 3'-pairing (Friedman *et al.* 2009). While a 3'-pairing site obviously does not need to be conserved to interact with a miRNA in the cell, this suggests that there is limited selective pressure for the maintenance of such sites across the targetome over time.

Grimson *et al.* (2007) found that the supplementary region (g12-g17 and especially g13-g16) has the largest impact on repression other than the seed itself. The relative importance of this region is also supported by structural evidence. As described in the previous section, RISC undergoes a conformational change during target binding, where helix-7 of the L2 domain shifts position to allow full seed pairing to form (Schirle *et al.* 2014). The movement of helix-7 causes positional shifts in the PAZ domain, widening the N-PAZ channel through which the miRNA 3' end is threaded and revealing the supplementary chamber. This causes residues g11-g16 to adopt a helical conformation, with the Watson-Crick surfaces of g13-g16 turned outside in a similar manner as the seed region in guide-only structures.

Intriguingly, a later crystal structure of RISC in complex with a fully complementary sixteen-nucleotide target RNA showed that the miRNA avoids pairing to the central region, which instead forms an unpaired bridge between the seed and supplementary regions (Sheu-Gruttadauria *et al.* 2019a). In this structure, the g9 base is stacked against the terminal base-pair of the 8mer seed to cap the helix,

while g10 is disordered and the electron density for g11 is difficult to interpret. Although g12 is exposed and able to pair to the target, this is prevented by steric clashes between the L2 domain and the sugar-phosphate backbone of t12.

This discovery is notable as pairing to the central region is required for efficient Argonaute-catalyzed slicing of target RNAs (Haley & Zamore 2004, Ameres *et al.* 2007), where RISC cleaves the phosphodiester bond between t10 and t11 (Elbashir *et al.* 2001). On the other hand, central pairing can decrease the binding affinity of RISC for a target site (Schirle *et al.* 2014, Becker *et al.* 2017). This suggests that for biological miRNA target sites, which in most animals are not generally downregulated by slicing, central pairing is unimportant or even disfavoured. Complementarity beyond the supplementary region does not generally appear to be required for target cleavage by *Drosophila* or human AGO2 (Wee *et al.* 2012, Becker *et al.* 2017). It should be noted that the crystal structure of RISC paired to a target RNA across the seed and supplementary regions used a catalytically inactive AGO2 mutant to avoid slicing of the target during crystallization (Sheu-Gruttadauria *et al.* 2019a).

Bartel (2018) proposes a stepwise model for miRNA target binding, in which 3'-pairing does not directly propagate from the seed helix but instead skips past the central region (g9-g12) to form a second helix in the supplementary region (g13-g16). With the seed helix fixed inside its binding channel, rotation of the second helix around its axis permits pairing to the central residues, enabling potential slicing of the target. Following the formation of the seed and supplementary helices, pairing may also propagate to the tail region. A similar model was proposed by Yan *et al.* (2018), in which the 3'-pairing helix nucleates at g12-g14 and propagates to g15-g17, after which conformational changes within RISC enables pairing to the central residues and potentially cleavage of the target between t10 and t11.

A recent high-throughput study of *in vitro* guide-target affinity suggests that different miRNAs have distinct preferences for pairing beyond the seed (McGeary *et al.* 2022). The authors employed a library of 3'-compensatory binding sites (Vella *et al.* 2004, Brennecke *et al.* 2005), featuring seed-binding sites with a mismatch at any position and random-sequence 3'-pairing sites, and measured the enrichment of sequences bound to RISC compared to the full set of sequences in the library. They found unexpected variability in the impact of the location of 3'-pairing relative to the seed. This is called the 3'-pairing offset,

defined as the difference in the number of unpaired nucleotides in the target and guide RNAs separating the seed and 3'-pairing helices. A positive offset indicates a bulge on the target side, a negative offset a bulge on the miRNA side.

While earlier work suggested an optimal arrangement of four contiguous canonical Watson-Crick pairs between the supplementary region and the target nucleotides directly opposite (zero offset) (Grimson *et al*. 2007), McGeary *et al*. (2022) found that human *let-7a* (a member of the conserved *let-7* family) has two binding modes with different offset preferences. If pairing forms to g11, affinity is higher for short positive offsets, whereas if pairing to g11 is unavailable the optimal offset is zero. The impact of 3'-compensatory pairing varied between miRNAs: for *let-7a* and *miR-155*, mismatched seed sites with 3'-compensatory pairing reached or even exceeded the affinity of a canonical 8mer site (without pairing outside the seed), whereas for *miR-1* the 3'-compensatory sites rarely reached the affinity of a canonical 6mer site. Unlike *let-7a*, neither *miR-1* nor *miR-155* significantly changed their preferred offset depending on available pairing. For all miRNAs, the most optimal sites included at least two pairs between g13 and g16, but not necessarily pairing to the entire supplementary region. These results suggest that Argonaute does not "standardize" 3'-pairing preferences across miRNAs as strongly as seed pairing.

Another *in vitro* binding study showed that while RISC first binds seed-matched targets roughly equally, over time the complexes redistribute themselves among the sites based on the strength of interactions in the supplementary region, so that the total dwell time on sites with supplementary pairing is larger even when such sites are greatly outnumbered by seed-only sites (Xiao & MacRae 2020). It is not clear to what extent this applies *in vivo*, where the measured effects of supplementary pairing are generally modest compared to *in vitro* differences in affinity. One possibility is that the repressive machinery recruited by RISC functions so fast upon seed binding that the full effect of 3'-pairing does not have time to exert all its effect, but the similar binding rates observed for 8mer and 6mer sites (Chandradoss *et al*. 2015, Salomon *et al*. 2015), which mediate different levels of repression, argues against this interpretation.

1.3.4 Other determinants of site selection and efficiency

Many potential miRNA target sites, as defined by the presence of a canonical seed-binding site, do not confer detectable repression. From microarray data,

Grimson *et al.* (2007) inferred that a minority of 3'-UTRs with a single target site for a miRNA are downregulated by that miRNA (19%, 23% and 43% for 7-mer-A1, 7mer-m8 and 8mer sites respectively). As the minimum defining features of a site is a sequence motif of only six to eight nucleotides, sites would be expected to frequently occur by chance. In addition to base-pairing interactions between the guide and target RNAs, numerous factors have been demonstrated to affect site selection. Some of these are incorporated into site prediction algorithms (e.g. Agarwal *et al.* 2015).

Cooperativity between sites

It was recognized early on in the exploration of miRNA site efficiency that multiple binding sites in the same 3'-UTR cause more repression than a single site (Doench & Sharp 2004, Brennecke *et al.* 2005, Lai *et al.* 2005). In a microarray-based transcriptomic study examining the effects of 11 different miRNAs in human cells, the greater average repression conferred by two sites compared to one were consistent with independent contributions from the sites (Grimson *et al.* 2007). The exception was that closely spaced binding sites resulted in more repression than expected from this model, suggesting some form of cooperative action. The observed effect was independent of the identity of the miRNAs, meaning that it is not simply a matter of increased probability that a particular miRNA reassociates with a nearby site following dissociation from its initial binding site. The study found that the optimal spacing between sites for cooperative binding was 8 to ~40 nucleotides. This was supported by a different study which demonstrated that the conservation patterns of seed-binding sites are characterized by spacing of ~10-130 nt and showed that 13-35 nucleotides are optimal for cooperative miRNA action in a cell-based reporter assay (Saetrom *et al.* 2007).

The basis of cooperative miRNA binding was subsequently revealed to be the capacity of TNRC6 (see section 1.5) to bind more than one Argonaute protein (Briskin *et al.* 2020). In the proposed model, multiple miRNAs bind the mRNA, after which TNRC6 binds to the Argonaute proteins of two RISCs bound to closely spaced sites. Although the dissociation rate of each individual RISC is unaffected by the interaction, TNRC6 bound to another target-paired RISC keeps the dissociated RISC tethered, maintaining a high local concentration of miRNAs and enabling rapid rebinding. Association with RISCs may similarly keep TNRC6 in place. The slower dissociation of multiple RISCs bound to TNRC6 relative to a single RISC gives TNRC6 more time to induce repression of the target gene. This

model explains why cooperative action is possible between different miRNAs, as only the spacing between binding sites is important, not their sequence. Although the study focused on two adjacent binding sites, the ability of human TNRC6 to bind up to three Argonaute proteins (Elkayam *et al.* 2017) means that the model could plausibly be extended to cooperation between three adjacent sites.

Structural accessibility of the target site

The structural accessibility of a miRNA target site within the 3'-UTR, as determined by local secondary and tertiary structures, affects downregulation of the gene (Brown *et al.* 2006, Kertesz *et al.* 2007). Accessibility can be modelled by predicting the free energy lost by unpairing the target site (ΔG_{open}) and the free energy gained by miRNA binding the target (ΔG_{duplex}), and calculating the difference between them ($\Delta\Delta G$). Kertesz *et al.* (2007) found that the overall interaction energy $\Delta\Delta G$ correlated better with target repression than ΔG_{duplex} in a reporter assay. In single-molecule FRET experiments, RISC was most effective in scanning unstructured regions for seed matches (Chandradoss *et al.* 2015). Furthermore, a high-throughput *in vitro* binding assay of seed-matched targets revealed that RISC association was substantially slowed by the formation of stable secondary structures occluding the seed-binding site (Becker *et al.* 2017).

Effective miRNA target sites are disproportionately found in 3'-UTR regions with high AU content (Grimson *et al.* 2007, Nielsen *et al.* 2007). One hypothesis to explain this finding is that RISC interacts with AU-rich elements, enhancing binding to nearby target sites (Jing *et al.* 2005, Nielsen *et al.* 2007, Vasudevan & Steitz 2007). Alternatively, high AU content may decrease the prevalence of stable secondary structures in the 3'-UTR, due to the lower thermodynamic stability of AU base pairs compared to GC pairs. Grimson *et al.* (2007) argued that structural accessibility disappeared as a predictive factor for site efficiency when controlling for local AU content, but later work from the same lab (Agarwal *et al.* 2015) found that accessibility, measured as log unpaired probability with RNAplfold (Lorenz *et al.* 2011) had predictive value when controlling for other factors, including AU content. McGeary *et al.* (2019) found that the two nucleotides flanking the seed-binding site on either side had a substantial impact on *in vitro* binding affinities, with the hierarchy of contributions ($A \approx U > C > G$) inversely reflecting the ability of each nucleotide to stabilize an occlusive structure.

Finally, RNA-binding proteins (RBPs) can open up 3'-UTR structures and thus make miRNA target sites more accessible for RISC (Kedde *et al.* 2010). A recent study

showed that the number of RBP binding sites in the vicinity of miRNA target sites positively correlate with greater downregulation (Kim *et al.* 2021). This suggests that RBPs function globally to open up 3'-UTR secondary structures for RISC binding, without requiring a functional association between a specific RBP and miRNA-target pair.

Characteristics of the 3'-UTR

Shorter 3'-UTRs are favourable for miRNA targeting (Agarwal *et al.* 2015). Multiple explanations for this phenomenon have been proposed (Hong *et al.* 2009): shorter UTRs may be less structured overall, increasing the accessibility of target sites, or else greater site density in shorter UTRs may enable cooperative binding (see above). Additionally, sites close to the either end of a 3'-UTR generally confer more repression (Gaidatzis *et al.* 2007, Grimson *et al.* 2007, Majoros & Ohler 2007). This is possibly related to proximity of these sites to translation initiation complexes (with the 3' end brought close to the 5' end by mRNA looping, discussed in section 1.5). Alternatively, sites near the ends may be more accessible due to inhibition of secondary structure formation by ribosomes and poly(A)-binding proteins (Ruijtenberg *et al.* 2020). Sites very close to the 5' end of 3'-UTRs, within 15-20 nt of the stop codon, are rarely conserved, possibly due to interference by ribosomes and other parts of the translational machinery with RISC binding (Gaidatzis *et al.* 2007, Grimson *et al.* 2007, Majoros & Ohler 2007).

Another factor affecting miRNA-mediated regulation is the presence of 3'-UTR isoforms of the same transcript. Processing of mRNAs in eukaryotic cells include the addition of a 5' cap, splicing of introns, and cleavage and polyadenylation of the 3' end. Primary transcripts extend thousands of nucleotides beyond the polyadenylation signal (Core *et al.* 2008). As 54% of human genes have more than one polyadenylation site, and 51% of polyadenylation sites have heterogenous cleavage sites (Tian *et al.* 2005), a large number of genes produce 3'-UTR isoforms with identical coding sequences. This complicates predictions of miRNA target site efficiency, as parameters like UTR length and position of the site within the UTR become more difficult to assess. Furthermore, some sites may be present only in a subset of isoforms, whose relative expression levels can vary between cell types and developmental stages.

RNA editing

RNA editing refers to post-transcriptional modifications of the nucleotide sequence of an RNA molecule. Some miRNAs are affected by adenosine-to-inosine (A-to-I) editing, carried out by enzymes in the adenosine deaminase acting on the RNA (ADAR) family (Bass & Weintraub 1988). Since inosine usually pairs with cytidine rather than uridine, the substitution alters the base-pairing preferences of the miRNA and consequently its target selection. Pri-miRNAs can be edited by ADARs (Luciano *et al.* 2004, Blow *et al.* 2006). Known A-to-I substitutions in miRNAs predominantly occur in the brain, more commonly in human than in mouse (Landgraf *et al*. 2007), and can redirect targeting, particularly in the rare cases where a seed nucleotide is edited in a substantial fraction of the cellular miRNA pool (Kawahara *et al.* 2007). Intriguingly, A-to-I editing of seed nucleotides in miRNAs may affect targeting differently than A-to-G point mutations (Kume *et al.* 2014). Other kinds of RNA modifications have also been demonstrated to redirect targeting, including m^6A methylation (Konno *et al*. 2019) and o^6G modification (Seok *et al*. 2020).

Target-directed miRNA degradation

Although higher affinity between guide and target would generally be expected to result in stronger repression, an interesting exception can occur when the miRNA is perfectly or nearly perfectly base-paired to both the seed and 3'-pairing regions. This sometimes leads to degradation of the miRNA rather than the mRNA, through a process termed target-directed miRNA degradation (TDMD) (Ameres *et al*. 2010). Crystal structures of human AGO2 bound to TDMD-inducing sites show that these sites cannot be accommodated by RISC, resulting in the release of the miRNA 3' end from the protein, which exposes it to attack by ribonucleases (Sheu-Gruttadauria *et al*. 2019b).

2 Research Aims

The aim of my doctoral studies was to explore the relationship between structure and function in miRNA targeting, focusing on the role of pairing outside the seed region in recognition of target sites and degree of gene repression.

As a model miRNA for these studies, we chose *miR-34a*, a member of the broadly conserved miR-34/449 seed family, which is involved in a broad range of physiological processes (Rokavec *et al.* 2014).

The work is presented in four papers with complementary aims:

- Establish a method to determine the specific base-pairs formed between a miRNA and its target sites (Paper I).
- Elucidate the mechanistic underpinnings of the differential effect of *miR-34a* on individual mRNA targets (Paper II).
- Understand the combined impact of 3'-pairing and secondary structures in the mRNA on target site efficiency (Paper III).
- Identify favourable structural features for *miR-34a* 3'-pairing across its entire targetome (Paper IV).

3 Results and Discussion

In this section, I will summarize the key results and conclusions of each work. Papers III and IV were the core of my doctoral studies and are consequently discussed in greater depth here.

3.1 Paper I

The secondary structure of RNA molecules can be determined by selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) (Merino *et al*. 2005). This technique is based on chemical modification of nucleotides by covalent 2'-O adducts which preferentially react with unpaired residues in the RNA. When the modified sequences are reverse transcribed to complementary DNA (cDNA), the adduct either aborts the transcription or induces a mutation. After sequencing the cDNA, the size distribution of truncated reads or pattern of mutations allow for quantification of the frequency of modification for each nucleotide, referred to as its reactivity. This is negatively correlated with its propensity to form base-pairs with other nucleotides. The reactivities can be used to infer the most probable secondary structure of the RNA.

In this paper, we developed RNA-RNA binding by SHAPE (RABS) to study the interaction between miRNAs (either alone or bound to Argonaute) and target sites *in vitro*, allowing us to infer base-pairing in the miRNA-target complex. The target site of interest is put into a large stem loop in a scaffold RNA, referred to as the main loop, flanked by unpaired buffer nucleotides. The scaffold also contains two reference hairpins as sequencing controls and a reverse transcription primer binding site. The scaffold is incubated with the modifier 1-methyl-7-nitroisatoic anhydride (1M7) (Mortimer & Weeks 2007), by itself or in the presence of miRNA, followed by capillary electrophoresis Sanger sequencing. By comparing the reactivities of target sites incubated with and without miRNA, we can determine the most likely intermolecular base-pairing interactions. Comparing the differences between naked miRNA and RISC also provides information about the impact of Argonaute on the miRNA-target interaction. In principle, the RABS technique can be extended to any RNAs and RNA-protein complexes.

3.2 Paper II

Although the fundamental principles of target recognition by miRNAs, via pairing to the seed region, are established, less is known about how the structure of a particular miRNA-complex affects the final outcome of the interaction. In this paper, we combined affinity measurements by electromobility shift assays (EMSA), structural probing by RABS and quantification of gene repression in cells by luciferase reporter assays to obtain in-depth biochemical profiles of twelve *miR-34a* target site interactions. We selected previously validated target sites (Rokavec *et al.* 2014) for characterization with the aim of obtaining a variety of possible secondary structures of the miRNA-target complex, as predicted by MC-Fold (Parisien & Major 2008).

We cloned the target sites intro luciferase reporter plasmids and measured the repression of each target mediated by *miR-34a* in HEK293T cells. While the degree of downregulation of the targets varied significantly, with an approximately four-fold difference between the strongest and weakest, repression was only weakly correlated with the site's *in vitro* binding affinity for RISC, as measured by EMSA ($R^2 = 0.23$). A comparable correlation was present between repression and the structural accessibility of the seed-binding site in the luciferase mRNA, as predicted by RNAplfold (Lorenz *et al.* 2011) ($R^2 = 0.2$). Intriguingly, comparing binding affinities of RISC to the *miR-34a* guide RNA alone suggests that Argonaute has a bidirectional effect on the strength of target binding, weakening strong binders and strengthening weak ones. Overall, sites with bulges on the miRNA side within the target-miRNA complex, as determined by structural probing with RABS, were more strongly repressed than sites without bulges or target-side bulges. Molecular Dynamics simulations of *miR-34a* in complex with a miRNA-bulged site (*NOTCH1*) indicated that small bulges between the seed and 3'-pairing helices can be accommodated by Argonaute.

3.3 Paper III

Beyond sequence complementarity, the amount of repression conferred by a miRNA target site is affected by the presence of secondary structures in the mRNA (Brown *et al.* 2006, Kertesz *et al.* 2007). In this paper, we investigated the impact of mRNA structure on miRNA-mediated downregulation, focusing both on

structural elements within the miRNA-target complex and structures already present in the mRNA prior to miRNA binding.

We chose the interaction of *miR-34a* with its *SIRT1* binding site as a model to study the impact of different structural features on downregulation. The guide-target complex consists of a 7mer-A1 seed helix and an extensive 3'-pairing helix separated by four unpaired target nucleotides. A structural model of this miRNA-target complex has previously been solved by nuclear magnetic resonance spectroscopy (Baronti *et al.* 2020). This structure revealed that the ground state 7mer-A1 seed helix exists in equilibrium with a transient excited state (ES) in which the seed helix is extended by a closing 3' GU base-pair. If the seed is mutated to a canonical 8mer site, the resulting miRNA-target complex mimics the excited state. Structural probing by RABS confirmed the intermolecular base-pairing patterns determined by NMR for both the wild-type and ES mutant in complex with *miR-34a*.

We explored the impact of each part of *miR-34a* on site efficiency. To this end, we divided the 3'-pairing region into the central (g9-g12), supplementary (g13-g16) and tail (g17-g22) regions (Wee *et al.* 2012) and cloned luciferase reporter plasmids with mutant *SIRT1* target sites in which pairing to each miRNA region was disrupted. In order to understand how 3'-pairing interacts with other structural features, we generated these mutants for sites with either 7mer-A1 or 8mer seeds (reflecting the ground state and trapped excited state), with or without the four unpaired nucleotides in the target, and with or without the two mismatches present in the otherwise fully complementary 3'-pairing helix of *SIRT1*. We measured the amount of repression mediated by *miR-34a* overexpression of all 32 variant target sites in HEK293T cells. We found that pairing to the supplementary region was consequential regardless of other structural elements. Intriguingly, the tail region also had a strong impact and was sometimes necessary to achieve repression stronger than the seed-only baseline. Pairing to the tail region was generally more consequential for sites with 7mer-A1 seeds, indicating a greater reliance on 3'-pairing for the weaker seed type.

When investigating the impact of different 3'-pairing regions, we noticed that sites without possible pairing to the supplementary and tail regions, but retaining pairing to other regions, sometimes mediated less downregulation than seed-only sites. This suggests that even in a reporter construct, where short target sites are embedded in an otherwise identical sequence context, the effects of a miRNA are

not determined by base-pairing interactions with the target alone. We hypothesized that the discrepancies could be explained by the presence of unfavourable secondary structures in the mRNA. We found that the observed repression of luciferase reporters correlated better with the predicted binding energy of the site (using modules from the ViennaRNA package (Lorenz *et al*. 2011)) in a two-step model in which both target unfolding and miRNA binding were considered than a one-step model with only miRNA binding. Most notably, in the two-step model the correlation between luciferase repression and total predicted interaction energy of the seed-binding site alone was strong for sites with weak 3'-pairing, but minimal for sites with strong 3'-pairing. In other words, 3'-pairing rendered the target sites less sensitive to the (predicted) structural accessibility of the seed-binding site.

Next, we tested the proposed rescue effect of 3'-pairing on inaccessible seed-binding sites. We designed target sites for *miR-34a* where an 8mer seed-binding site can form a hairpin with a fully complementary downstream sequence. One site featured full 3'-pairing upstream of the seed, while the other offered seed pairing only. Structural probing of these sites by RABS showed that in the absence of *miR-34a*, the hairpins formed as predicted, rendering the seed-binding site fully inaccessible. For the site without 3'-pairing, the presence of *miR-34a* (alone or bound to AGO2) did not alter the reactivities of any nucleotides in the site, indicating that the target hairpin structure precludes miRNA binding. By contrast, when 3'-pairing was available upstream of the hairpin, addition of *miR-34a* led to decreased reactivities across the 3'-pairing site and increased reactivities in the target nucleotides sequestering the seed, indicating competition between RISC and target self-pairing for seed binding. In a luciferase reporter assay, the site without 3'-pairing was not downregulated compared to a scrambled seed control, while the site with 3'-pairing was substantially repressed. In summary, we present both structural and functional evidence for the ability of miRNA 3'-pairing to rescue sites with inaccessible seed-binding sites, which would otherwise be non-functional.

We proposed two models of the binding process, not incompatible with each other, which could plausibly explain our findings. In the first model, which does not require RISC to possess any abilities not described in prior work, the mRNA unfolds independently and allows the miRNA to bind to the seed. The presence of 3'-pairing slows down subsequent miRNA dissociation, compensating for the energetic penalty imposed by the necessity of target site unfolding. In the cell,

mRNA structure is variable due to interactions with RNA-binding proteins (Kim *et al.* 2021), ribosome movement (Ruijtenberg *et al.* 2020) and competition with other possible secondary structures, which means that even a stably self-paired target site may be exposed by chance often enough to enable miRNA binding. Even so, the fact that some inaccessible target sites were not repressed at all in our reporter assays shows that for this model to work, the pairing outside the seed must contribute a high degree of additional stability. In the second model, the miRNA first transiently interacts with the 3'-pairing site, allowing for the gradual replacement of target self-pairing by miRNA-target pairing via a strand displacement mechanism (Hong & Šulc 2019). This would allow for competitive binding even if the extant mRNA structure is highly stable, as single base-pair end fraying can be sufficient to enable strand displacement. On the other hand, it has never been directly demonstrated that RISC can bind target sites in this manner and structural evidence suggests that the 3'-end of the guide strand is not well-positioned to interact with the target in the absence of seed binding (Schirle *et al.* 2012, Schirle *et al.* 2014).

3.4 Paper IV

One of the central problems in assessing the role of pairing beyond the seed in microRNA targeting is the difficulty of knowing which sites are substantially affected by it in the cell. Seed-binding sites of the same type can mediate different levels of baseline repression depending on factors like their location in the 3'-UTR or their structural accessibility. Consequently, the observation that two target sites with potential 3'-pairing are comparably downregulated does not by itself imply that the 3'-pairing is equally important in both cases. Here, we used abasic modified nucleotides to remove the possibility of base-pairing to residues g13 and g14 in *miR-34a*. By disrupting pairing to the supplementary region without introducing alternative pairing possibilities, we were able to study the impact of these guide nucleotides on site efficiency by comparing the effects of the wild-type and mutant miRNAs on global *miR-34a* regulation in cells.

We measured the effects of overexpressing wild-type and mutant *miR-34a* in HEK293T cells, on the RNA level after 24 h incubation (with bulk RNA-seq) and on the protein level after 48 h incubation (with tandem mass tag mass spectroscopy (Branca *et al.* 2014)). Both variants exerted significant effects on the transcriptome and proteome, with a moderately strong correlation between changes in mRNA

and protein expression (R^2 = 0.55 for the wild-type and R^2 = 0.47 for the mutant), comparable to prior work (Baek *et al.* 2008). We found a strong correlation between the impact of wild-type and mutant *miR-34a* relative to the negative control on mRNA (R^2 = 0.86) and protein expression (R^2 = 0.91) of genes with at least one canonical seed-binding site, indicating that a minority of genes are affected by disruption of g13-g14 pairing. Intriguingly, while several genes were significantly more downregulated by wild-type *miR-34a* relative to the mutant, this was not the case for any genes at the protein level.

We wanted to identify structural features of target sites that contribute to effective 3'-pairing, as measured by the response at the mRNA level. We thus searched genes with a single canonical seed-binding site that were significantly downregulated by wild-type *miR-34a* compared to the negative control for predicted binding to the supplementary region, and quantified structural parameters such as GU pairing frequency and the offset between the seed and 3'-pairing helices. We found that *miR-34a* generally tolerates GU wobble pairs in the 3'-pairing helix, with the exception of position g14 (a guanidine residue in *miR-34a*). As a group, sites forming a GU pair to g14 did not mediate more repression than sites with no predicted supplementary pairing at all. We also examined the effect of the offset between the seed and 3'-pairing helices. Partially consistent with our results in Paper II, we found that *miR-34a* prefers small negative offsets (representing bulges on the miRNA side) or zero offset (perfect alignment between miRNA and target bases) to positive offsets (bulges on the target side). We validated our findings regarding GU pairing and offset preferences with designed target sites in luciferase reporter assays, obtaining results consistent with the RNA-seq data.

While GU wobble pairs are comparable to canonical AU pairs in terms of thermodynamic stability, they distort the shape of RNA helices (Varani & McClain 2000). GU pairs are not tolerated in the seed for this reason (Wee *et al.* 2012, Schirle *et al.* 2014) and the sensitivity of *miR-34a* to GU pairs at a particular position in the supplementary region could plausibly be explained by structural constraints imposed by Argonaute. It remains to be studied if GU pairs at g13 or g15 (an adenine and cytosine in *miR-34a*) are similarly deleterious, and if the impact of GU pairs in the supplementary region depends on the identity of neighbouring bases. Our finding that *miR-34a* prefers negative to positive offsets between the seed and 3'-pairing helices contrasts with observations from a recent high-throughput binding affinity study of 3'-compensatory sites (McGeary

et al. 2022), emphasizing the partially miRNA-specific nature of pairing preferences. Overall, our results demonstrate that abasic modification of miRNAs is a promising approach for investigating the impact of specific miRNA nucleotides on gene silencing.

4 Concluding Remarks and Future Perspectives

Although the study of miRNAs has come a long way since they were discovered around three decades ago, many questions about the mechanism and physiological role of this mode of gene regulation remain unanswered. Despite the apparent simplicity of their targeting mechanism, based on sense-antisense complementarity to around twenty nucleotides, the determinants of site recognition and binding affinity are not fully understood even when the miRNA-target interaction is considered in isolation, with no regard to cellular context. This brings into focus the extent to which the binding properties of the guide RNA are reshaped by its interaction with Argonaute proteins.

It is interesting to consider why miRNAs in humans and many other animals appear to have evolved away from consistently using the full guide sequence for target recognition, as plant miRNAs do. Speculatively, limiting the primary recognition motif to around eight nucleotides may allow for a broader set of potential targets for each miRNA, with pairing outside the seed working to sharpen differences between members of the same seed family (which are themselves differentially regulated over time), or to contribute increased affinity when required (for example, to compensate for unfavourable mRNA structures, which may be conserved for unrelated reasons).

The work presented here used a variety of techniques to study miRNA-target interactions, both detailed biochemical investigation of individual sites and statistical analyses of large datasets. What we have not studied are the kinetics of miRNA-mediated gene silencing. Due to the difficulty of measuring changes in mRNA and protein levels over time for more than a few genes at once, this has been relatively neglected in the study of miRNA targeting. As changes in the levels of specific proteins can have effects on vastly different timescales, miRNAs may use a variety of strategies to target classes of genes. New techniques and approaches to monitor mRNA or protein levels in cells over time could lead to a new understanding of the role of miRNAs in shaping gene expression patterns.

Our work, when combined with recent work by other researchers, reveals that miRNAs differ to a significant extent in their pairing preferences outside the seed. This means that target prediction algorithms may need new scoring metrics, taking into account the individual characteristics of each miRNA. Further study of

the structure of RISC and its binding mechanism could help us understand the relationship between guide RNA sequence and preference for different structural features in the guide-target complex.

5 Acknowledgements

First of all, I want to thank my supervisor **Emma Andersson** for giving me the opportunity to work on this project and the freedom to explore my own research ideas. The work took some unexpected turns over the years, in directions unfamiliar to both of us. I admire your ambition and creativity as a scientist.

I want to thank my co-supervisor **Katja Petzold** for all your support. I've enjoyed and learned a lot from our long, sometimes heated debates about RNA molecules.

I'm grateful to all the members of the Andersson lab over the years for all the fun times we've had. Thank you to **Simona** for holding the group together and for being the only person able to give me a nickname that stuck; **Katrin** for being a great friend with a sharp sense of humour; **Janek** for all our conversations inside and outside the lab, and for teaching me about cutting and pasting bits of DNA; **Noémi** for your thoughtfulness and all your fun ideas; **Anna**, our honourary lab member, it's too quiet here without you; **Ileana** for introducing me to the lab and for being great company when stuck in an elevator; **Paulina** for understanding the importance of good beer and strange movies; **Eli** for your energy and cheerfulness, I'm sure you'll finish both the post-doc and the 90 km in style; **Sandra** for organizing much-needed ski trips; **Bettina** for all our conversations during the few hours of the day when we are both awake; **Lenka** for your inspiring enthusiasm for all things scientific; **Agus** for joining me in the quest for a halfway functional coffee machine in Biomedicum; **Jia** for introducing us all to great restaurants; **Jingyan** for your kindness and generosity; **Afshan**, we ended up defending the same week, it's been fun sharing this journey with you.

Thanks also to all members of the Petzold lab and especially the miRNA team: **Sarah**, **Lorenzo**, **Lara**, **Hannes**, **Elnaz**, **Rubin**, **Walter**, **Ting** and **Joanna**.

My PhD has taken me to many places at Karolinska. Thanks to everyone at BioNut and CMB, and in the D6 and B9 quarters at Biomedicum.

Finally, I want to thank my friends and family for supporting me over the years.

6 References

Agarwal V, Bell GW, Nam JW, Bartel DP. (2015). Predicting effective microRNA target sites in mammalian mRNAs. eLife 4: e05005.

Alberti C, Cochella L. (2017). A framework for understanding the roles of miRNAs in animal development. Development 144: 2548-2559.

Alon U. (2007). Network motifs: theory and experimental approaches. Nature Reviews Genetics 8: 450–461.

Alvarez-Saavedra E, Horvitz HR. (2010). Many families of C. elegans microRNAs are not essential for development or viability. Current Biology 20: 367–373.

Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M *et al.* (2003). A uniform system for microRNA annotation. RNA 9: 277–279.

Ameres SL, Martinez J, Schroeder R. (2007). Molecular basis for target RNA recognition and cleavage by human RISC. Cell 130: 101-112.

Ameres SL, Horwich MD, Hung JH, Xu J, Ghildiyal M, Weng Z, Zamore PD. (2010). Target RNA-directed trimming and tailing of small silencing RNAs. Science 328: 1534–1539.

Aravind L, Walker DR, Koonin EV. (1999). Conserved domains in DNA repair proteins and evolution of repair systems. Nucleic Acids Research 27: 1223-1242.

Aravind L, Watanabe H, Lipman DJ, Koonin EV. (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. Proceedings of National Academy of Sciences 97: 11319-11324.

Azuma-Mukai A, Oguri H, Mituyama T, Qian ZR, Asai K, Siomi H, Siomi MC. (2008). Characterization of endogenous human argonautes and their miRNA partners in RNA silencing. Proceedings of the National Academy of Sciences 105: 7964-7969.

Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. (2008). The impact of microRNAs on protein output. Nature 455: 64–71.

Bail S, Swerdel M, Liu H, Jiao X, Goff LA, Hart RP, Kiledjian M. (2010). Differential regulation of microRNA stability. RNA 16: 1032–1039.

Baronti L, Guzzetti I, Ebrahimi P, Friebe Sandoz S, Steiner E, Schlagnitweit J, Fromm B, Silva L, Fontana C, Chen AA *et al.* (2020). Base-pair conformational switch modulates miR-34a targeting of Sirt1 mRNA. Nature 583: 139–144.

Bartel DP, Chen CZ. (2004). Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nature Reviews Genetics 5: 396-400.

Bartel DP. (2018). Metazoan microRNAs. Cell 173: 20-51.

Bass BL, Weintraub H. (1988). An unwinding activity that covalently modifies its double-stranded-RNA substrate. Cell 55: 1089-1098.

Baulcombe D. (2004). RNA silencing in plants. Nature 431: 356-363.

Becker WR, Ober-Reynolds B, Jouravleva K, Jolly SM, Zamore PD, Greenleaf WJ. (2019). High-throughput analysis reveals rules for target RNA binding and cleavage by AGO2. Molecular Cell 75: 741-755.

Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P, Izaurralde E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. Genes & Development 20: 1885-1898.

Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E *et al.* (2005). Identification of hundreds of conserved and nonconserved human microRNAs. Nature Genetics 37: 766–770.

Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E, Plasterk RHA. (2006). Diversity of microRNAs in human and chimpanzee brain. Nature Genetics 38: 1375–1377.

Berezikov E, Robine N, Samsonova A, Westholm JO, Naqvi A, Hung J, Okamura K, Dai Q, Bortolamiol-Becet D, Martin R *et al.* (2011). Deep annotation of *Drosophila melanogaster*

microRNAs yields insights into their processing, modification, and emergence. Genome Research 21: 203-215.
Berezikov E. (2011). Evolution of microRNA diversity and regulation in animals. Nature Reviews Genetics 12: 846-860.
Bernstein E, Caudy AA, Hammond SM, Hannon GJ. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409: 363-366.
Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, Mills AA, Elledge SJ, Anderson KV, Hannon GJ. (2003). Dicer is essential for mouse development. Nature Genetics 35: 215–217.
Bethune J, Artus-Revel CG, Filipowicz W. (2012). Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. EMBO Reports 13: 716–723.
Blow MJ, Grocock RJ, van Dongen S, Enright AJ, Dicks E, Futreal PA, Wooster R, Stratton MR. (2006). RNA editing of human microRNAs. Genome Biology 7: R27.
Bohmert K, Camus I, Bellini C, Bouchez D, Caboche M, Benning C. (1998). AGO1 defines a novel locus of *Arabidopsis* controlling leaf development. The EMBO Journal 17: 170-180.
Boland A, Tritschler F, Heimstadt S, Izaurralde E, Weichenrieder O. (2010). Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. EMBO Reports 11: 522–527.
Bosson AD, Zamudio JR, Sharp PA. (2014). Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. Molecular Cell 56: 347–359.
Branca RM, Orre LM, Johansson HJ, Granholm V, Huss M, Perez-Bercoff A, Forshed J, Kall L, Lehtiö J. (2014). HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. Nature Methods 11: 59-62.
Brennecke J, Stark A, Russell RB, Cohen SM. (2005). Principles of microRNA-target recognition. PLoS Biology 3: e85.
Briskin D, Wang PY, Bartel DP. (2020). The biochemical basis for the cooperative action of microRNAs. Proceedings of the National Academy of Sciences 117: 17764-17774.
Broughton JP, Lovci MT, Huang JL, Yeo GW, Pasquinelli AE. (2016). Pairing beyond the seed supports microRNA targeting specificity. Molecular Cell 64: 320–333.
Brown KM, Chu C, Rana TM. (2006). Target accessibility dictates the potency of human RISC. Nature Structural and Molecular Biology 12: 469-470.
Bråte J, Neumann RS, Fromm B, Haraldsen ABA, Tarver JE, Suga H, Philip C.J. Donoghue PCJ, Peterson KJ, Ruiz-Trillo I, Grini PE, Shalchian-Tabrizi K. (2018). Unicellular origin of the animal microRNA machinery. Current Biology 28: 3288–3295.
Campo-Paysaa F, Semon M, Cameron RA, Peterson KJ, Schubert M. (2011). microRNA complements in deuterostomes: origin of evolution of microRNAs. Evolution & Development 13: 15-27.
Chandradoss SD, Schirle NT, Szczepaniak M, MacRae IJ, Joo C. (2015). A dynamic search process underlies microRNA targeting. Cell 162: 96-107.
Chen K, Rajewsky N. (2007). The evolution of gene regulation by transcription factors and microRNAs. Nature Reviews Genetics 8: 93-103.
Chen Y, Boland A, Kuzuoğlu-Öztürk D, Bawankar P, Loh B, Chang CT, Weichenrieder O, Izaurralde E. (2014). A DDX6-CNOT1 complex and W-binding pockets in CNOT9 reveal direct links between miRNA target recognition and silencing. Molecular Cell 54: 737-750.
Chi SW, Zang JB, Mele A, Darnell RB. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature 460: 479–486.
Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE *et al*. (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes & Development 24: 992–1009.
Core LJ, Waterfall JJ, Lis JT. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322: 1845–1848.
Denzler R, McGeary SE, Title AC, Agarwal V, Bartel DP, Stoffel M. (2016). Impact of microRNA levels, target-site complementarity, and cooperativity on competing endogenous RNA-regulated gene expression. Molecular Cell 64: 565–579.
Didiano D, Hobert O. (2006). Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. Nature Structural & Molecular Biology 13: 849-851.

Diederichs S, Haber DA. (2007). Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. Cell 131: 1097-1108.
Djuranovic S, Nahvi A, Green R. (2012). miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. Science 336: 237–240.
Doench JG, Sharp PA. (2004). Specificity of microRNA target selection in translational repression. Genes & Development 18: 504-511.
Dueck A, Ziegler C, Eichner A, Berezikov E, Meister G. (2012). microRNAs associated with the different human Argonaute proteins. Nucleic Acids Research 40: 9850–9862.
Eichhorn SW, Guo H, McGeary SE, Rodriguez-Mias RA, Shin C, Baek D, Hsu SH, Ghoshal K, Villen J, Bartel DP. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. Molecular Cell 56: 104–115.
Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411: 494-498.
Elkayam E, Faehnle CR, Morales M, Sun J, Li H, Joshua-Tor L. (2017). Multivalent recruitment of human argonaute by GW182. Molecular Cell 67: 646–658.
Faehnle CR, Elkayam E, Haase AD, Hannon GJ, Joshua-Tor L. (2013). The making of a slicer: activation of human Argonaute-1. Cell Reports 3: 1901–1909.
Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP. (2005). The widespread impact of mammalian microRNAs on mRNA repression and evolution. Science 310: 1817–1821.
Fernandez-Valverde SL, Taft RJ, Mattick JS. (2010). Dynamic isomiR regulation in Drosophila development. RNA 16: 1881-1888.
Flores O, Kennedy EM, Skalsky RL, Cullen BR. (2014). Differential RISC association of endogenous human microRNAs predicts their inhibitory potential. Nucleic Acids Research 42: 4629-4639.
Frank F, Sonenberg N, Nagar B. (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. Nature 465: 818–822.
Friedman RC, Farh KK, Burge CB, Bartel DP. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Research 19: 92-105.
Fromm B, Billipp T, Peck LE, Johansen M, Tarver JE, King BL, Newcomb JM, Sempere LF, Flatmark K, Hovig E, Peterson KJ. (2015). A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. Annual Review of Genetics 49: 213–242.
Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M. (2007). Inference of miRNA targets using evolutionary conservation and pathway analysis. BMC Bioinformatics 8: 69.
Gantier MP, McCoy CE, Rusinova I, Saulep D, Wang D, Xu D, Irving AT, Behlke MA, Hertzog PJ, Mackay F, Williams BR. (2011). Analysis of microRNA turnover in mammalian cells following Dicer1 ablation. Nucleic Acids Research 39: 5692–5703.
Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, Baskerville S, Hammond SM, Bartel DP, Schier AF (2005). MicroRNAs regulate brain morphogenesis in Zebrafish. Science 308: 833–838.
Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Molecular Cell 27: 91-105.
Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature 455: 1193–1197.
Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. Cell 106: 23–34.
Grosswendt S, Filipchyk A, Manzano M, Klironomos F, Schilling M, Herzog M, Gottwein E, Rajewsky N. (2014). Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. Molecular Cell 54: 1042–1054.
Gu X, Su Z, Huang Y. (2009). Simultaneous expansions of microRNAs and protein-coding genes by gene/genome duplications in early vertebrates. Journal of Experimental Zoology B 312B: 164-170.

Guo H, Ingolia NT, Weissman JS, Bartel DP. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466: 835–840.
Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp AC, Munschauer M *et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141: 129–141.
Haley B, Zamore PD. (2004). Kinetic analysis of the RNAi enzyme complex. Nature Structural & Molecular Biology 11: 599-606.
Hauptmann J, Dueck A, Harlander S, Pfaff J, Merkl R, Meister G. (2013). Turning catalytically inactive human Argonaute proteins into active slicer enzymes. Nature Structural & Molecular Biology 20: 814–817.
Hausser J, Syed AP, Bilen B, Zavolan M. (2013). Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. Genome Research 23: 604-615.
Heimberg AM, Sempere LF, Moy VN, Donoghue PCJ, Peterson KJ. (2008). MicroRNAs and the advent of vertebrate morphological complexity. Proceedings of the National Academy of Sciences 105: 2946-2950.
Helwak A, Kudla G, Dudnakova T, Tollervey D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. (2013). Cell 153: 654–665.
Hendrickson DG, Hogan DJ, McCullough HL, Myers JW, Herschlag D, Ferrell JE, Brown PO. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. PLoS Biology 7: e1000238.
Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF & The Students of Bioinformatics Computer Labs 2004 and 2005. (2006). The expansion of the metazoan microRNA repertoire. BMC Genomics 7: 25.
Hong F, Šulc P. (2019). An emergent understanding of strand displacement in RNA biology. Journal of Structural Biology 207: 241–249.
Hong X, Hammell M, Ambros V, Cohen SM. (2009). Immunopurification of Ago1 miRNPs selects for a distinct class of microRNA targets. Proceedings of the National Academy of Sciences 106: 15085–15090.
Horman SR, Janas MM, Litterst C, Wang B, MacRae IJ, Sever MJ, Morrissey DV, Graves P, Luo B, Umesalma S *et al.* (2013). Akt-mediated phosphorylation of argonaute 2 downregulates cleavage and upregulates translational repression of MicroRNA targets. Molecular Cell 50: 356-367.
Hornstein E, Shomron N. (2006). Canalization of development by microRNAs. Nature Genetics 38: 520-524.
Hur JK, Zinchenko MK, Djuranovic S, Green R. (2013). Regulation of Argonaute slicer activity by guide RNA 3′ end interactions with the N-terminal lobe. Journal of Biological Chemistry 288: 7829–7840.
Hutvágner G, McLachlan J, Pasquinelli AE, Bálint E, Tuschl T, Zamore PD. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science 293, 834–838.
Hutvágner G, Zamore PD. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. Science 297: 2056–2060.
Inui M, Martello G, Piccolo S. (2010). MicroRNA control of signal transduction. Nature Reviews Molecular Cell Biology 11: 252-263.
Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T, Tomari Y. (2010). Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Molecular Cell 39: 292-299.
Jacob F, Monod J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology 3: 318-356.
Jin HY, Oda H, Chen P, Yang C, Zhou X, Kang SG, Valentine E, Kefauver JM, Liao L, Zhang Y *et al.* (2017). Differential sensitivity of target genes to translational repression by miR-17~92. PLoS Genetics 13: e1006623.
Jing Q, Huang S, Guth S, Zarubin T, Motoyama A, Chen J, Di Padova F, Lin SC, Gram H, Han J. (2005). Involvement of microRNA in AU-rich element-mediated mRNA instability. Cell 120: 623–634.

Jonas S, Izaurralde E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. Nature Reviews Genetics 16: 421-433.
Jones-Rhoades MW, Bartel DP, Bartel B. (2006). MicroRNAs and their regulatory roles in plants. Annual Review of Plant Biology 57: 19–53.
Kataoka Y, Takeichi M, Uemura T. (2001). Developmental roles and molecular characterization of a *Drosophila* homologue of Arabidopsis Argonaute1, the founder of a novel gene superfamily. Genes to Cells 6: 313–325.
Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. (2007). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science 315: 1137–1140.
Kedde M, van Kouwenhove M, Zwart W, Oude Vrielink JA, Elkon R, Agami R. (2010). A Pumilio-induced RNA structure switch in p27-3' UTR controls miR-221 and miR-222 accessibility. Nature Cell Biology 12: 1014–1020.
Keene JD. (2007). RNA regulons: coordination of post-transcriptional events. Nature Reviews Genetics 8: 533-543
Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. (2007). The role of site accessibility in microRNA target recognition. Nature Genetics 39: 1278-1284.
Khvorova A, Reynolds A, Jayasena SD. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell 115: 209–216.
Kim S, Kim S, Chang HR, Kim D, Park J, Son N, Park J, Yoon M, Chae G, Kim YK *et al*. (2021). The regulatory impact of RNA-binding proteins on microRNA targeting. Nature Communications 12: 5057.
Kobayashi H, Singer RH. (2022). Single-molecule imaging of microRNA-mediated gene silencing in cells. Nature Communications 13: 1435.
Konno M, Koseki J, Asai A, Yamagata A, Shimamura T, Motooka D, Okuzaki D, Kawamoto K, Mizushima T, Eguchi H *et al*. (2019). Distinct methylation levels of mature microRNAs in gastrointestinal cancers. Nature Communications 10: 3888.
Koonin EV, Dolja VV, Krupovic M. (2015). Origins and evolution of viruses of eukaryotes: the ultimate modularity. Virology 479-480: 2–25.
Koonin. (2017). Evolution of RNA- and DNA-guided antivirus defense systems in prokaryotes and eukaryotes: common ancestry vs convergence. Biology Direct 12: 5.
Kuchenbauer F, Mah SM, Heuser M, McPherson A, Rüschmann J, Rouhi A, Berg T, Bullinger L, Argiropoulos B, Morin RD *et al.* (2011). Comprehensive analysis of mammalian miRNA* species and their role in myeloid cells. Blood 118: 3350–3358.
Kume H, Hino K, Galipon J, Ui-Tei K. (2014). A-to-I editing in the miRNA seed region regulates target mRNA selection and silencing efficiency. Nucleic Acids Research 42: 10050–10060.
Kwak PB, Tomari Y. (2013). The N domain of Argonaute drives duplex unwinding during RISC assembly. Nature Structural & Molecular Biology 19: 145–151.
Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschi T. (2001). Identification of novel genes coding for small expressed RNAs. Science 294: 853-858.
Lai EC, Tam B, Rubin GM. (2005). Pervasive regulation of Drosophila Notch target genes by GY-box-, Brd-Box-, and K-box-class microRNAs. Genes & Development 19: 1067-1080.
Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M *et al.* (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129:1401–1414.
Lau NC, Lim NP, Weinstein EG, Bartel DP. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science 294: 858-862.
Lee RC, Feinbaum RL, Ambros V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75: 843-854.
Lee RC, Ambros V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. Science 294: 862-864.
Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S *et al.* (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature 425: 415–419.
Lee Y, Jeon K, Lee JT, Kim S, Kim VN. (2002). MicroRNA maturation: stepwise processing and subcellular localization. The EMBO Journal 21: 4663-4670.

Lewis BP, Burge CB, Bartel DP. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120:15–20.
Li Y, Wang F, Lee J, Gao, F. (2006). MicroRNA-9a ensures the precise specification of sensory organ precursors in *Drosophila*. Genes & Development 20: 2793–2805.
Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP. (2003). Vertebrate microRNA genes. Science 299: 1540.
Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433: 769–773.
Lingel A, Simon B, Izaurralde E, Sattler M. (2004). Nucleic acid 3′-end recognition by the Argonaute2 PAZ domain. Nature Structural & Molecular Biology 11: 576–577.
Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua-Tor L, Hannon GJ. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science 305: 1437–1441.
Loeb GB, Khan AA, Canner D, Hiatt JB, Shendure J, Darnell RB, Leslie CS, Rudensky AY. (2012). Transcriptome-wide miR-155 binding map reveals widespread noncanonical microRNA targeting. Molecular Cell 48: 760–770.
Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. (2011). ViennaRNA package 2.0. Algorithms in Molecular Biology 6: 26.
Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA *et al.* (2005). MicroRNA expression profiles classify human cancers. Nature 435: 834-838.
Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu CI. (2008). The birth and death of microRNA genes in Drosophila. Nature Genetics 40: 351-355.
Luciano DJ, Mirsky H, Vendetti NJ, Maas S. (2004). RNA editing of a miRNA precursor. RNA 10: 1174–1177.
Ma JB, Ye K, Patel DJ. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature 429: 318–322.
Ma JB, Yuan YR, Meister G, Pei Y, Tuschl T, Patel DJ. (2005). Structural basis for 5′-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. Nature 434: 666–670.
MacRae IJ, Doudna JA. (2007). Ribonuclease revisited: structural insights into ribonuclease III family enzymes. Current Opinion in Structural Biology 17: 138-145.
Majoros WH, Ohler U. (2007). Spatial preferences of microRNA targets in 3' untranslated regions. BMC Genomics 8: 152.
Mathys H, Basquin J, Ozgur S, Czarnocki-Cieciura M, Bonneau F, Aartse A, Dziembowski A, Nowotny M, Conti E, Filipowicz W. (2014). Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. Molecular Cell 54: 751-765.
Mayya VK, Flamand MN, Lambert AM, Jafarnejad SM, Wohlschlegel JA, Sonenberg N, Duchaine TF. (2021). microRNA-mediated translation repression through GYF-1 and IFE-4 in *C. elegans* development. Nucleic Acids Research 49: 4803-4815.
McGeary SE, Lin KS, Shi CY, Pham TM, Bisaria N, Kelley GM, Bartel DP. (2019). The biochemical basis of microRNA targeting efficacy. Science 366: eaav1741.
McGeary SE, Bisaria N, Pham TM, Wang PY, Bartel DP. (2022). MicroRNA 3'-compensatory pairing occurs through two binding modes, with affinity shaped by nucleotide identity and position. eLife 11: e69803.
McJunkin K, Ambros V. (2017). A microRNA family exerts maternal control on sex determination in C. elegans. Genes & Development 31: 422-437.
Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T. (2004). Human argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Molecular Cell 15: 185–197.
Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. (2005). RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). Journal of the American Chemical Society 127: 4223–4231.

Moore MJ, Scheel TKH, Luna JM, Park CY, Fak JJ, Nishiuchi E, Rice CM, Darnell RB. (2015). MiRNA-target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. Nature Communications 6: 8864.
Moran Y, Fredman D, Praher D, Li XZ, Wee LM, Rentzsch F, Zamore PD, Technau U, Seitz H. (2014). Cnidarian microRNAs frequently regulate targets by cleavage. Genome Research 24: 651–663.
Moran Y, Agron M, Praher D, Technau U. (2017). The evolutionary origin of plant and animal microRNAs. Nature Ecology and Evolution 1: 0027.
Moretti F, Kaiser C, Zdanowicz-Specht A, Hentze MW. (2012). PABP and the poly(A) tail augment microRNA repression by facilitated miRISC binding. Nature Structural & Molecular Biology 19: 603-608.
Moroz LL, Kocot KM, Citarella MR, Dosung S, Norekian TP, Povolotskaya IS, Grigorenko AP, Dailey C, Berezikov E, Buckley KM *et al.* (2014). The ctenophore genome and the evolutionary origins of neural systems. Nature 510: 109–114.
Mortimer SA, Weeks, KM. (2007). A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. Journal of the American Chemical Society 129: 4144–4145.
Muddashetty RS, Nalavadi VC, Gross C, Yao X, Xing L, Laur O, Warren ST, Bassell GJ. (2011). Reversible inhibition of PSD-95 mRNA translation by miR-125a, FMRP phosphorylation, and mGluR signaling. Molecular Cell 42: 673-688.
Nakanishi K, Weinberg DE, Bartel DP, Patel DJ. (2012). Structure of yeast Argonaute with guide RNA. Nature 486: 368–374.
Nakanishi K, Ascano M, Gogakos T, Ishibe-Murakami S, Serganov AA, Briskin D, Morozov P, Tuschl T, Patel DJ. (2013). Eukaryote-specific insertion elements control human ARGONAUTE slicer activity. Cell Reports 3: 1893-1900.
Nakanishi K. (2022). Anatomy of four human Argonaute proteins. Nucleic Acids Research 50: 6618-6638.
Nguyen TA, Jo MH, Choi YG, Park J, Kwon SC, Hohng S, Kim VN, Woo JS. (2015). Functional anatomy of the human Microprocessor. Cell 161: 1374-1387.
Nielsen CB, Shomron N, Sandberg R, Hornstein E, Kitzman J, Burge CB. (2007). Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. RNA 13: 1894–1910.
Okamura K, Phillips M, Tyler D, Duan H, Chou Y, Lai EC. (2008). The regulatory activity of microRNA* species has substantial influence on microRNA and 3′ UTR evolution. Nature Structural & Molecular Biology 15: 354–363.
Olovnikov I, Chan K, Sachidanandam R, Newman DK, Aravin AA. (2013). Bacterial argonaute samples the transcriptome to identify foreign DNA. Molecular Cell 51: 594–605.
Parisien M, Major F. (2008). The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature 452: 51–55.Park MS, Phan HD, Busch F, Hinckley SH, Brackbill JA, Wysocki VH, Nakanishi K. (2017). Human argonaute3 has slicer activity. Nucleic Acids Research 45: 11867–11877.
Parker JS, Roe SM, Barford D. (2004). Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. The EMBO Journal 23: 4727–4737.
Parker JS, Roe SM, Barford D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature 434: 663–666.
Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Müller P *et al.* (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature 408: 86-89.
Peterson KJ, Dietrich MR, McPeek MA. (2009). MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. BioEssays 31: 736–747.
Piriyapongsa J, Mariño-Ramírez L, Jordan IK. (2007). Origin and evolution of human microRNAs from transposable elements. Genetics 176: 1323-1337.
Prochnik SE, Rokhsar DS, Aboobaker AA. (2007). Evidence for a microRNA expansion in the bilaterian ancestor. Development Genes and Evolution 217: 73-77.
Rehwinkel J, Behm-Ansmant I, Gatfield D, Izaurralde E. (2005). A crucial role for GW182 and the DCP1: DCP2 decapping complex in miRNA-mediated gene silencing. RNA 11: 1640-1647.

Reichholf B, Herzog VA, Fasching N, Manzenreither RA, Sowemimo I, Ameres SL. (2019). Time-resolved small RNA sequencing unravels the molecular principles of microRNA homeostasis. Molecular Cell 75: 756–768.
Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JG, Rougvie AE, Horvitz HR, Ruvkun G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. Nature 403: 901-906.
Ro S, Park C, Young D, Sanders KM, Yan W. (2007). Tissue-dependent paired expression of miRNAs. Nucleic Acids Research 35: 5944-5953.
Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. (2004). Identification of mammalian microRNA host genes and transcription units. Genome Research 14: 1902-1910.
Rokavec M, Li H, Jiang L, Hermeking H. (2014). The p53/miR-34 axis in development and disease. Journal of Molecular Cell Biology 6: 214–230.
Rouya C, Siddiqui N, Morita M, Duchaine TF, Fabian MR, Sonenberg N. (2014). Human DDX6 effects miRNA-mediated gene silencing via direct binding to CNOT1. RNA 20: 1398-1409.
Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. Cell 127: 1193–1207.
Ruijtenberg S, Sonneveld S, Cui TJ, Logister I, de Steenwinkel D, Xiao Y, MacRae IJ, Joo C, Tanenbaum ME. (2020). mRNA structural dynamics shape Argonaute-target interactions. Nature Structural & Molecular Biology 27: 790-801.
Saetrom P, Heale BS, Snove O, Aagaard L, Alluin J, Rossi JJ. (2007). Distance constraints between microRNA target sites dictate efficacy and cooperativity. Nucleic Acids Research 35: 2333–2342.
Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. Genes & Development 20: 2214–2222.
Salomon WE, Jolly SM, Moore MJ, Zamore PD, Serebrov V. (2015). Single-molecule imaging reveals that Argonaute reshapes the binding properties of its nucleic acid guides. Cell 162: 84-95.
Sapkota S, Pillman KA, Dredge BK, Liu D, Bracken JM, Kachooei SA, Chereda B, Gregory PA, Bracken CP, Goodall GJ. (2023). On the rules of engagement for microRNAs targeting protein coding regions. Nucleic Acids Research 51: 9938-9951.
Schirle NT, MacRae IJ. (2012). The crystal structure of human Argonaute2. Science 336: 1037-1040.
Schirle NT, Sheu-Gruttadauria J, Chandradoss SD, Joo C, MacRae IJ. (2015). Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets. eLife 4: e07646.
Schirle NT, Sheu-Gruttadauria J, MacRae IJ. (2014). Structural basis for microRNA targeting. Science 346: 608-613.
Schmiedel JM, Klemm SL, Zheng Y, Sahay A, Blüthgen A, Marks DS, van Oudenaarden A. (2015). MicroRNA control of protein expression noise. Science 348: 128-132.
Schratt GM, Tuebing F, Nigh EA, Kane CG, Sabatini ME, Kiebler M, Greenberg ME. (2006). A brain-specific microRNA regulates dendritic spine development. Nature 439: 283–289.
Schwarz DS, Hutvágner G, Du T, Xu Z, Aronin N, Zamore PD. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell 115: 199–208.
Sempere LF, Cole CN, McPeek MA, Peterson KJ. (2006). The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. Journal of Experimental Zoology B 306: 575-588.
Seok H, Lee H, Lee S, Ahn SH, Lee HS, Kim GD, Peak J, Park J, Cho YK, Jeong Y *et al.* (2020). Position-specific oxidation of miR-1 encodes cardiac hypertrophy. Nature 584: 279–285.
Shabalina SA, Koonin EV. (2008). Origins and evolution of eukaryotic RNA interference. Trends in Ecology and Evolution 23: 578-587.
Sheng G, Zhao H, Wang J, Rao Y, Tian W, Swarts DC, van der Oost J, Patel DJ, Wang Y. (2014). Structure-based cleavage mechanism of *Thermus thermophilus* Argonaute DNA guide strand-mediated DNA target cleavage. Proceedings of the National Academy of Sciences 111: 652–657.

Sheu-Gruttadauria J, Xiao Y, Gebert LFR, MacRae IJ. (2019a). Beyond the seed: structural basis for supplementary microRNA targeting by human Argonaute2. The EMBO Journal 38: e101153.
Sheu-Gruttadauria J, Pawlica P, Klum SM, Wang S, Yario TA, Schirle Oakdale NT, Steitz JA, MacRae IJ. (2019b). Structural basis for target-directed microRNA degradation. Molecular Cell 75: 1243-1255.e7.
Sonenberg N, Hinnebusch AG. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell 136: 731-745.
Song JJ, Smith SK, Hannon GJ, Joshua-Tor L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305: 1434–1437.
Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature 508: 66–71.
Suzuki HI, Katsura A, Yasuda T, Ueno T, Mano H, Sugimoto K, Miyazono K. (2015). Small-RNA asymmetry is directly driven by mammalian Argonautes. Nature Structural & Molecular Biology 22: 512–521.
Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin EV, Patel DJ, van der Oost J. (2014a). The evolutionary journey of Argonaute proteins. Nature Structural & Molecular Biology 21: 743–53.
Swarts DC, Jore MM, Westra ER, Zhu Y, Janssen JH, Snijders AP, Wang Y, Patel DJ, Berenguer J, Brouns SJ, van der Oost J. (2014b). DNA-guided DNA interference by a prokaryotic Argonaute. Nature 507: 258–61.
Taft R, Pheasant M, Mattick JS. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. BioEssays 29: 288-299.
Tian B, Hu J, Zhang H, Lutz CS. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Research 33: 201–212.
Vagin VV, Sigova A, Li CJ, Seitz H, Gvozdev V, Zamore PD. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science 313: 320–324.
Varani G, McClain WH. (2000). The G·U wobble base pair. EMBO Reports 1: 18–23.
Vasquez-Rifo A, Jannot G, Armisen J, Labouesse M, Bukhari SIA, Rondeau EL, Miska ES, Simard MJ. (2012). Developmental characterization of the microRNA-specific *C. elegans* Argonautes alg-1 and alg-2. PLOS ONE 7: e33750.
Vasudevan S, Steitz JA. (2007). AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. Cell 128: 1105–1118.
Vella MC, Choi EY, Lin SY, Reinert K, Slack FJ. (2004). The *C. elegans* microRNA let-7 binds to imperfect let-7 complementary sites form the lin-41 3'UTR. Genes & Development 18: 132-137.
Wang D, Zhang Z, O'Loughlin E, Lee T, Houel S, O'Carroll D, Tarakhovsky A, Ahn NG, Yi R. (2012). Quantitative functions of argonaute proteins in mammalian development. Genes & Development 26: 693–704
Wang YL, Sheng G, Juranek S, Tuschl T, Patel DJ. (2008). Structure of the guide-strand-containing argonaute silencing complex. Nature 456: 209–213.
Wee LM, Flores-Jasso CF, Salomon WE, Zamore PD. (2012). Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. Cell 151: 1055-1067.
Wheeler BM, Heimberg AM, Moy VN, Sperling EA, Holstein TW, Heber S, Peterson KJ. (2009). The deep evolution of metazoan microRNAs. Evolution & Development 11: 50-68.
Wightman B, Burglin TR, Gatto J, Arasu P, Ruvkun G. (1991). Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development. Gene and Development 5: 1813–1824.
Wightman B, Ha I, Ruvkun G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. Cell 75: 855-862.
Xiao Y, MacRae IJ. (2020). Robust differential microRNA targeting driven by supplementary interactions in vitro. RNA 26: 162-174.
Yan Y, Acevedo M, Mignacca L, Desjardins P, Scott N, Imane R, Quenneville J, Robitaille J, Feghaly A, Gagnon E *et al*. (2018). The sequence features that define efficient and specific hAGO2-dependent miRNA silencing guides. Nucleic Acids Research 46: 8181-8196.
Yekta S, Shih IH, Bartel DP. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science 304: 594–596.

Yi R, Qin Y, Macara IG, Cullen BR. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes & Development 17: 3011-3016.
Yuan Y, Pei Y, Ma J, Kuryavyi V, Zhadina M, Meister G, Chen H, Dauter Z, Tuschi T, Patel D. (2005). Crystal structure of *A. aeolicus* argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Molecular Cell 19: 405–419.
Yuan Z, Sun X, Liu H, Xie J. (2011). MicroRNA genes derived from repetitive elements and expanded by segmental duplication events in mammalian genomes. PLOS ONE 6: e17666.
Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W. (2004). Single processing center models for human Dicer and bacterial RNase III. Cell 118: 57–68.
Zolotarov G, Fromm B, Legnini I, Ayoub S, Polese G, Maselli V, Chabot PJ, Vinther J, Styfhals R, Seuntjens E *et al.* (2022). MicroRNAs are deeply linked to the emergence of the complex octopus brain. Science Advances 8: eadd9938.