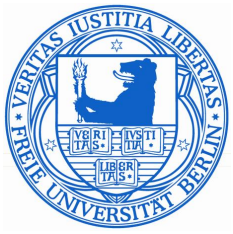
STRUCTURAL AND FUNCTIONAL ANALYSIS OF LIN28-MEDIATED INHIBITION OF LET-7 MIRNA BIOGENESIS

Dissertation zur Erlangung des akademischen Grades des
Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin



vorgelegt von

FLORIAN MAYR
aus Lauingen a.d. Donau

Januar 2013

Die vorliegende Arbeit wurde von November 2009 bis Januar 2013 unter der Leitung von PROF. DR. UDO HEINEMANN am Max-Delbrück-Centrum für Molekulare Medizin, Berlin-Buch, angefertigt.

1. Gutachter: PROF. DR. UDO HEINEMANN
2. Gutachter: PROF. DR. MARKUS WAHL

Disputation am 14. Mai 2013

ABSTRACT

Lin28 is an essential RNA-binding protein that is ubiquitously expressed in embryonic stem cells. Its physiological function has been linked to regulation of differentiation, development, oncogenesis as well as glucose metabolism. Recently, emerging evidence has revealed that Lin28 mediates these pleiotropic functions by inhibiting *let-7* miRNA biogenesis and by modulating the translation of target mRNAs.

In this PhD thesis, the structural and mechanistic basis of Lin28-mediated inhibition of *let-7* biogenesis was analyzed. Lin28 binds to the terminal loop (pre-element) of precursor (pre) *let-7* miRNA thereby impairing a cleavage by the ribonuclease III Dicer. Both Lin28 RNA-binding domains (RBDs), a cold-shock domain (CSD) and a retroviral-type Zn-knuckle domain (ZKD), were essential for pre-*let-7* binding and Dicer inhibition. A systematic binding analysis demonstrated that both domains bind to single-stranded (ss) nucleic acids with the ZKD mediating specific binding to a conserved GGAG motif, while the CSD showed an overall low sequence specificity with a slight preference for pyrimidine-rich oligonucleotides. Crystal structures of Lin28 CSDs in complex with hexa- and heptathymidine as well as hexauridine revealed the molecular basis for the limited sequence specificity, as binding of ssDNA/RNA was dominated by unspecific base-stacking interactions.

Further electrophoretic mobility shift assays with pre-*let-7* and Lin28 variants confirmed the importance of the GGAG motif, as mutations within this motif or the ZKD impaired both binding and inhibition of Dicer mediated pre-*let-7* processing. However, only the isolated CSD, but not the ZKD, could bind to pre-*let-7* alone. Using site-directed mutagenesis in combination with a time-resolved RNA remodeling assay, I could show that Lin28 binds in a stepwise manner to pre-*let-7*. After initial binding of the CSD, a structural change within pre-*let-7* is induced leading to melting of Dicer cleavage site and facilitating subsequent binding of the ZKD to the conserved GGAG motif. Thereby Lin28 can recognize all *let-7* members despite their structural diversity and ensure specific inhibition of their biogenesis.

Apart from competitive inhibition, Lin28 is also known to promote polyuridylation of pre-*let-7* thereby labeling it for degradation. Using co-immunoprecipitation and *in vitro* uridylation assays, I identified two retroviral-type CCHC Zn-knuckles in TUT4 that are essential for pre-*let-7* uridylation in a Lin28-dependent manner. On the Lin28 level, both the C-terminus as well as the two RBDs were indispensable for promoting pre-*let-7* uridylation.

ZUSAMMENFASSUNG

Lin28 ist ein essentielles RNA-Bindeprotein, welches ubiquitär in embryonalen Stammzellen exprimiert wird und mit dessen Hilfe induziert pluripotente Stammzellen (iPSC) erzeugt werden können. Seine physiologische Funktion besteht in der Regulation vieler zellulärer Prozesse wie Zelldifferenzierung, Wachstum und Entwicklung als auch Onkogenese. Auf molekularer Ebene inhibiert Lin28 die *let-7* miRNA Biogenese und moduliert darüber hinaus die Translation diverser mRNAs.

Das Ziel dieser Arbeit lag in der strukturellen und funktionellen Untersuchung der Lin28•pre-*let-7* Interaktion, um auf molekularer Ebene zu verstehen, wie Lin28 die *let-7* Biogenese inhibiert. Es konnte gezeigt werden, dass Lin28 pre-*let-7* spezifisch über dessen terminale Schleife bindet und auf diese Weise eine Prozessierung durch die Ribonuklease Dicer verhindert. Diese Prozesse erforderten die Anwesenheit der beiden Lin28 RNA-Bindedomänen (RBDs), einer Kälteschockdomäne (CSD) und einer retroviralen Zinkfingerdomäne (ZKD). Eine systematische Analyse der Bindungsspezifitäten dieser offenbarte, dass beide Domänen einzelsträngige (ss) DNA und RNA binden, wobei die ZKD eine hohe Spezifität gegenüber einem konservierten GGAG-Motiv aufwies. Im Gegensatz dazu zeigte die isolierte CSD eine sehr breite Spezifität mit einer leichten Präferenz für Pyrimidin-reiche Sequenzen. Kristallstrukturen von Lin28 CSDs im Komplex mit ssDNA/RNA-Oligonukleotiden enthüllten, dass deren Bindung im Wesentlichen durch unspezifische Basen-Stapelung erfolgt während kaum Basen-spezifische Wasserstoffbrückenbindungen ausgebildet werden.

Elektrophoretische Bindungsstudien von Lin28 Varianten mit pre-*let-7* bestätigten die Bedeutung des GGAG-Motivs für die Bindung an pre-*let-7*. Mutationen innerhalb des GGAG-Motivs als auch der Lin28 ZKD beeinträchtigten sowohl die Bindung als auch die Blockierung der pre-*let-7* Prozessierung durch Dicer. Interessanterweise konnte jedoch nur die isolierte CSD, nicht aber die ZKD, an pre-*let-7* binden. Weitere Mutagenesestudien zusammen mit einem RNA-Remodellierungstest zeigten, dass Lin28 in einem mehrstufigen Prozess an pre-*let-7* bindet. Demnach wird nach einer anfänglichen Bindung der CSD eine strukturelle Änderung in pre-*let-7* induziert, die zu einem Aufschmelzen der Dicer-Schnittstelle führt. Dadurch wird eine Bindung der ZKD an das konservierte GGAG-Motiv erleichtert. Auf diese Weise kann Lin28 alle *let-7* Familienmitglieder trotz deren strukturellen Diversität erkennen und deren Biogenese spezifisch inhibieren.

Neben der kompetitiven Inhibierung ist Lin28 auch dazu in der Lage die Poly-Uridylierung von *let-7* zu stimulieren, wodurch diese für den Abbau markiert wird. Durch Co-Immunopräzipitation und *in vitro* Uridylierungstests konnten zwei essentielle CCHC Zn-Finger in TUT4 identifiziert werden, die essentiell für die Lin28-abhängige pre-*let-7* Uridylierung sind. Auf Lin28 Ebene waren sowohl der C-Terminus als auch die beiden RBDs für die pre-*let-7* Uridylierung erforderlich.

PUBLICATIONS

Mayr, F., Schuetz, A., Doege, N., Heinemann, U. The Lin28 cold-shock domain remodels pre-let-7 miRNA. *Nucleic Acids Res*, *40*, 7492-7506 (2012).

Mayr, F., Heinemann, U. Cold-shock domains - versatile molecular modules for single-stranded nucleic acid recognition and remodeling. *Biomolecular forms and functions*, 220-228. World Scientific Publishing Company (2012).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 MICRORNAS AND THEIR DISCOVERY

The controlled expression of genes is essential for all organisms and thus is tightly regulated by several different mechanisms. In the last twenty years, small microRNAs (miRNAs) have emerged as important regulators of gene expression in eukaryotes that affect several cellular processes such as tissue development, differentiation, proliferation, apoptosis, innate and adaptive immune response [1, 2, 3, 4, 5, 6, 7]. MiRNAs are endogenously expressed 21 to 25 nt long RNAs that negatively regulate gene expression at a post-transcriptional level [8, 9].

The first miRNA, *lin-4*, was discovered in *Caenorhabditis elegans (C. elegans)* almost 20 years ago. *lin-4* encodes for a small RNA that forms a characteristic hairpin precursor. Unlike protein encoding genes, the primary RNA transcript is not translated but further processed to an about 22 nt long RNA. Worms that contained mutations in *lin-4* failed to execute certain developmental stages. Further studies revealed, that *lin-4* is a post-transcriptional repressor of *lin-14* and *lin-28*, genes important for the temporal patterning of *C. elegans* larval development. *lin-4* mediates the translational repression of *lin-14* and *lin-28* mRNAs via complementary base pairing in the 3'-untranslated regions (3' UTR) [4, 10, 11]. A few years later, a second miRNA called *let-7* was discovered that controls the transition of *C. elegans* larval stage L4 to the adult state [12]. Like *lin-4*, *let-7* also binds to partially complementary sites within the 3' UTR of its target genes *lin-41* and *hbl-1* and thereby represses their translation and thus promotes differentiation [12, 13, 14, 15]. Moreover, it was shown that *let-7* miRNAs and their target *lin-41* are evolutionarily conserved across various animal species [16]. Since then, a large number of miRNA similar to *lin-4* and *let-7* have been discovered in animals, plants and viruses [17, 18, 19]. Recent developments such as deep sequencing [20] along with advanced computational prediction algorithms [21, 22] allowed the identification of less abundant miRNAs as well as their potential target sites. Currently, the miRNA database (miRBase) predicts up to 1600 human miRNAs though less than 1000 haven been experimentally confirmed [23, 24, 25].

1.2 MIRNA CHARACTERISTICS

MiRNAs have several unique features in common and are mainly classified based on their biogenesis pathway (see Section 1.3 and Figure 1). In contrast to small interfering (si) RNAs that typically arise from long exogenous double stranded (ds) RNAs taken up by the cell or transcribed from viral vectors, miRNAs are endogenously expressed and contain hairpin structures within the primary transcript (reviewed in [26]). Genes encoding for miRNAs are non-

randomly distributed over the entire genome at inter- and intragenic locations and are often arranged in clusters that share a common transcript. Based on their genomic location miRNAs can be grouped into three major categories: i) intronic miRNAs in non-coding transcripts (~40%, e.g. *miR-14*), ii) exonic miRNAs in non-coding transcripts (~10%, e.g. *let-7*) iii) intronic miRNAs in protein coding transcripts (~40%, e.g. *miR-7*) [27, 28, 25, 29].

Apart from their different origins, in animals miRNA and siRNA differ with respect to their sequence complementarity to their target mRNAs. While siRNA show a perfect complementarity, miRNAs predominantly bind via their seed sequence to their target mRNAs. This seed sequence is highly conserved within a miRNA family and normally spans nt 2 to 7 at the 5′ end of mature miRNAs. However, seed pairing alone with the target mRNA is only rarely sufficient for an efficient post-transcriptional repression of the target mRNA [30, 31]. In addition, an imperfect pairing of the seed could be compensated in some cases by extensive base pairing at the 3′ end [32] or the central part of the miRNA [33] (reviewed in [34]).

1.3 MIRNA BIOGENESIS

1.3.1 *Nuclear processing of primary miRNA transcripts*

Transcription of miRNA genes is mostly mediated by RNA polymerase II and less frequently by RNA polymerase III (Figure 1) [35, 36]. The resulting primary miRNA transcripts (pri-miRNAs) are often capped and polyadenylated [28]. Moreover, they form one or several hairpin structures that are typically composed of a terminal loop (pre-element, preE), a 33 bp long ds stem containing mismatches or bulges and flanking single-stranded (ss) RNA segments. The pri-miRNA is then processed by the nuclear Microprocessor complex formed by the RNase III Drosha and the dsRNA-binding protein DGCR8 (DiGeorge critical region 8, also known as Pasha in invertebrates) [37, 38, 39, 40]. DGCR8 as well as a ds stem and its flanking regions were shown to be essential for pri-miRNA processing. Moreover, DGCR8 stabilizes Drosha via direct protein-protein interactions and thus directs cleavage of Drosha 11 bp from the ds/ss junction, resulting in an approximately 65 bp long precursor miRNA (pre-miRNA) with a characteristic 2 nt 3′ overhang (Figure 2) [41, 42, 43, 44]. Further studies reported that the Microprocessor crops pri-miRNA co-transcriptionally and therefore prior to splicing [45, 46, 47].

Alternatively, pre-miRNA like hairpin structures (mirtrons) can also be generated without Drosha directly from short introns of protein-coding mRNAs. They are directly produced as a side product of the splicing reaction and therefore can bypass Drosha processing [48, 49, 50].

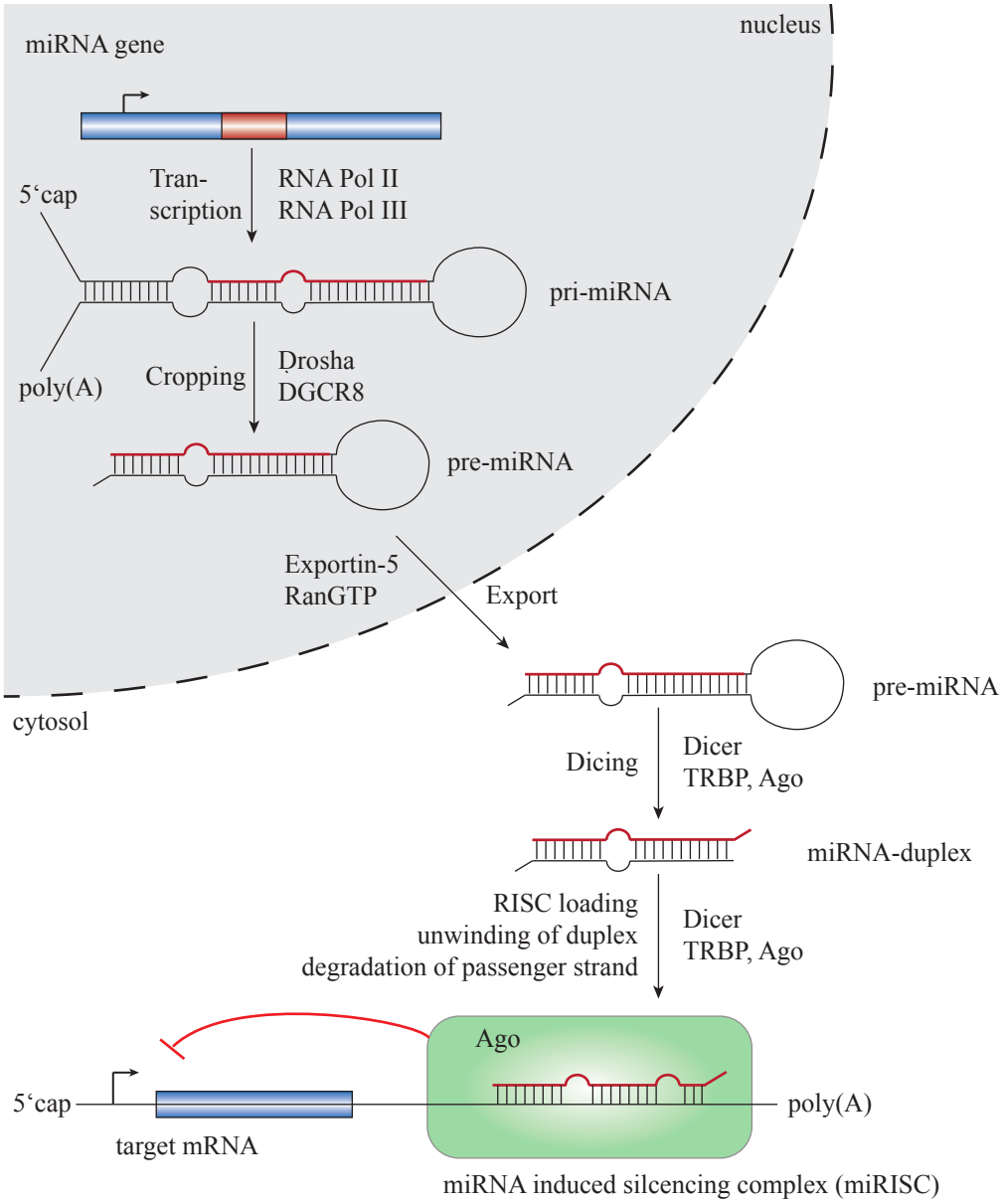


Figure 1: Canonical miRNA biogenesis pathway. Primary miRNA transcripts (pri-miRNAs) are transcribed by RNA polymerase II or RNA polymerase III and subsequently capped and polyadenylated. Then the Microprocessor complex composed of the RNase III Drosha and the dsRNA-binding protein DGCR8 cleaves pri-miRNAs approximately 11 bp from the ss/ds junction. The resulting hairpin precursor miRNA (pre-miRNA) is recognized and transported to the cytoplasm by Exportin-5•RanGTP where a second RNase III called Dicer removes the terminal loop to generate a miRNA duplex. TRBP and Ago associate with Dicer to mediate the processing and subsequent transfer of the guide strand (red) into the miRNA induced silencing complex (miRISC), where it directs miRISC to silence the target mRNA. The second strand of the miRNA duplex, also called passenger strand, is degraded.

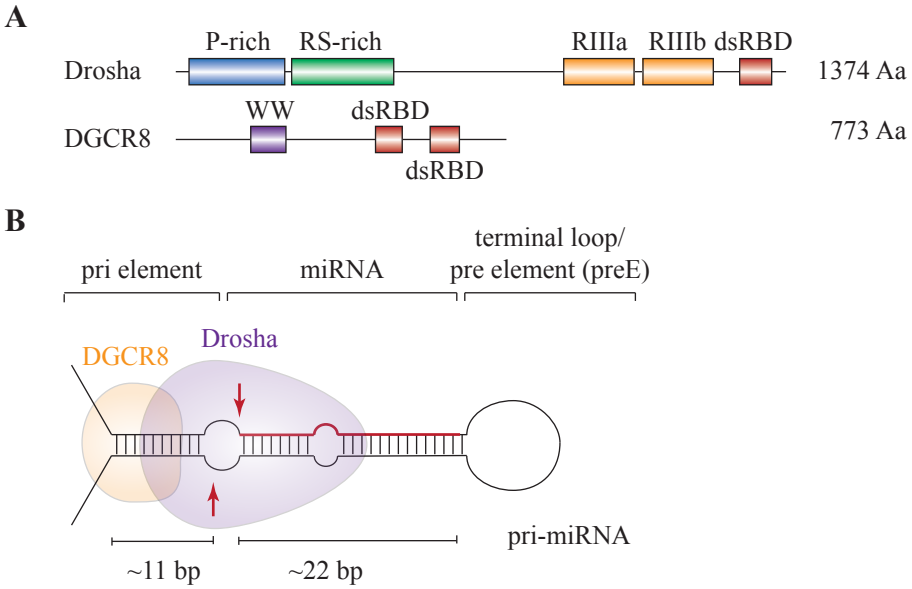


Figure 2: Maturation of pri-miRNA by the Microprocessor complex. (**A**) Schematic representation of the domain organization of the Microprocessor forming proteins Drosha and DGCR8. P-rich: Pro-rich domain, RS-rich: Arg/Ser-rich domain, RIIIa, RIIIb: tandemly arranged RNase III domains, dsRBD: dsRNA-binding domain, WW: Trp-rich domain. (**B**) DGCR8 binds to both the ss flanking sequences and the upper stem of pri-miRNAs. Furthermore, it stabilizes Drosha via direct interactions of its WW-domain with Drosha's P-rich domain. As a result, Drosha specifically cleaves the stem loop 11 bp away from the ss/ds junction. The resulting ~ 65 nt long pre-miRNA contains a 2 nt 3′ overhang, a hallmark of all type III ribonuclease cleavage products.

1.3.2 *Nuclear export of precursor miRNAs*

Exportin-5 in complex with RanGTP mediates export of pre-miRNAs to the cytoplasm [51]. In the nucleus, the small GTPase Ran is present in its GTP bound form and thus associates with Exportin-5 to form a stable export complex. After pre-miRNA loading and export to the cytoplasm via a nuclear core complex, Ran bound GTP is hydrolyzed by a Ran GTPase activating protein (RanGAP). Thereupon, the export complex dissociates and the pre-miRNA is released (reviewed in [52, 53]).

Recent structural studies of Exportin-5•RanGTP in complex with pre-*miR-30a* showed that the protein export factor specifically recognizes the 2 nt 3′ overhang of pre-miRNAs via several salt bridges and hydrogen bonds in a deep tunnel (Figure 3). Additional unspecific electrostatic contacts are mediated predominantly with the phosphate backbone of the ds stem and are thus sequence independent [54]. No contacts are mediated with the preE. Taken together, the specific recognition of the 3′ overhang and a defined length of the ds stem are sufficient for Exportin-5 to export the pre-miRNA in a RanGTP dependent manner. Moreover, Exportin-5•RanGTP probably protects miRNAs from degradation by RNases, as a knockdown of Exportin-5 led to decreased levels of mature miRNAs but not to an accumulation of nuclear pre-miRNAs [51, 55, 56].

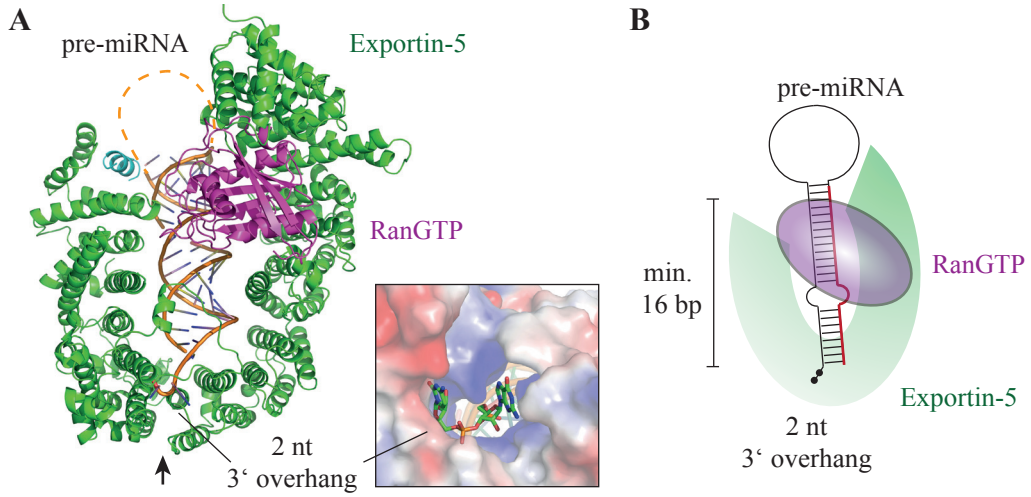


Figure 3: Nuclear export of pre-miRNAs by Exportin-5•RanGTP. (**A**) X-ray crystal structure of Exportin-5•RanGTP in complex with pre-*miR-30a* (PDB-ID 3A6P). The pre-miRNA is bound in the inside of Exportin-5•RanGTP via weak and unspecific electrostatic interactions with the ds stem. In contrast, a basic tunnel at the bottom of the complex mediates strong interactions with the 2 nt 3′ overhang via various salt bridges and hydrogen bonds. (**B**) Schematic drawing of Exportin-5•RanGTP•pre-miRNA interactions. At least 16 bp of the ds stem and the 2 nt 3′ overhang are required for Exportin-5 mediated export [57].

1.3.3 *Cytoplasmic processing by Dicer*

Once in the cytoplasm, a second ribonuclease III called Dicer cleaves the pre-miRNA close to the end of the ds stem thereby releasing an approximately 22 nt miRNA duplex and the preE [58, 59, 60]. Recent structural and biochemical studies revealed that Dicer specifically recognizes the 2 nt 3′ overhang created by Drosha by means of its PAZ domain [61]. The length of cleaved miRNA duplex is determined by a long α-helical connector helix spanning the distance between Dicer's PAZ domain and the active site of the intramolecular RNase III dimer. In the crystal structure of *Giardia intestinalis* Dicer, the RNase III active site is located ~ 65 Å away from the 3′ end corresponding to about 25 dsRNA base pairs (Figure 4 B). This is consistent with the length of cleavage products. Like all RNase III enzymes, Dicer catalyzes a staggered cleavage of the dsRNA leading to a miRNA duplex containing a 2 nt 3′ overhang at each end.

Dicer is highly conserved among almost all eukaryotic species. Compared to *Giardia* Dicer, human Dicer contains an additional DEXD/H-box helicase domain, a domain of unknown function (DUF283) and a dsRNA-binding domain (Figure 4 A). Recent EM studies with human Dicer revealed an L-shaped particle with the helicase domain forming a clamp-like structure next to the RNase III active site, which facilitates recognition of preEs (Figure 4 C) [62].

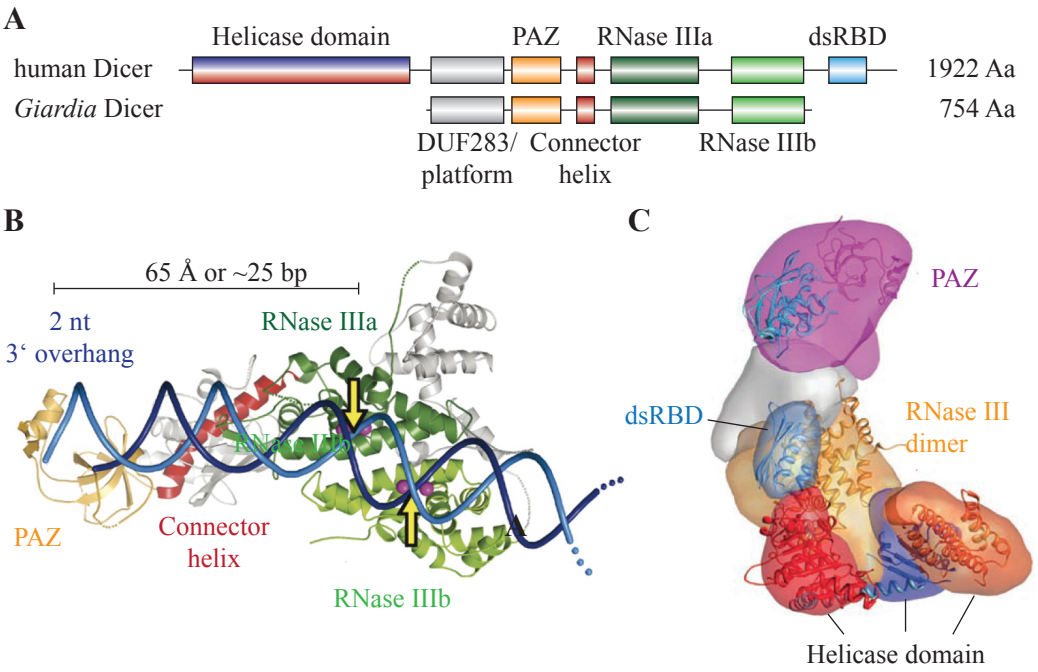


Figure 4: Structural basis of pre-miRNA processing by Dicer. (**A**) Human Dicer contains a DEAD-box RNA helicase, a platform (DUF283), a connector helix, a PAZ domain and a dsRNA-binding domain (dsRBD) in addition to its tandemly arranged RNase III domains. (**B**) Model of *Giardia intestinalis* Dicer (PDB-ID 2FFL) with bound dsRNA. The PAZ domain binds the 2 nt 3′ overhang and thereby positions the 3′ OH end 65 Å away from the RNase III active site. Figure modified from [63]. (**C**) Architecture of human Dicer as obtained by EM [62].

1.3.4 *MiRNA-induced silencing complex (miRISC) loading*

Following Dicer cleavage, the miRNA duplex must be separated into: i) the functional guide strand, which selectively guides the miRNA-induced silencing complex (miRISC) to target mRNAs, and ii) the passenger strand that is subsequently degraded [64]. The process of selective incorporation depends on thermodynamic stability of the duplex termini. Typically, the miRNA with less stable base pairs at its 5′ end is loaded into miRISC and serves as guide strand [65]. In addition, structural features such as the position of mismatches and bulges also contribute to strand selection [66, 67].

To enable miRISC loading, human Dicer associates with the dsRNA-binding proteins TRBP (Tar RNA binding protein) or PACT (protein activator of the protein kinase R). TRBP and/or PACT were demonstrated to recruit Argonaute (Ago) to the complex, thereby forming the RISC loading complex (RLC) [68, 69, 70]. The complex can also be formed prior to pre-miRNA binding whereby TRBP stabilizes Dicer and facilitates cleavage of the pre-miRNAs [71, 68, 70]. Though the precise mechanism of RLC mediated miRISC loading is not known, it was suggested that the miRNA duplex is partially released from Dicer and repositioned to favor selective loading of the guide strand's 3′ end within Ago's PAZ domain. After repositioning, Dicer preferentially binds to the less stable end of the miRNA duplex, while TRBP binds to the more stable end [72]. This

brings the 3′ end of the guide strand in close proximity to Ago2 and thus facilitates a directional transfer to Ago (Figure 5).

How the guide strand finally is unwound and separated from the passenger is not known yet. Several helicases like p68, p72 or Mov10 were shown to associate with miRISC components and thus might support the unwinding reaction [73, 74, 75, 76]. However, as miRISC loading was shown to be ATP-independent [77] it is likely that the helicases only have a supportive function. Furthermore, it was proposed that Ago2 cleave the passenger strand directly via its slicer activity [78, 79].

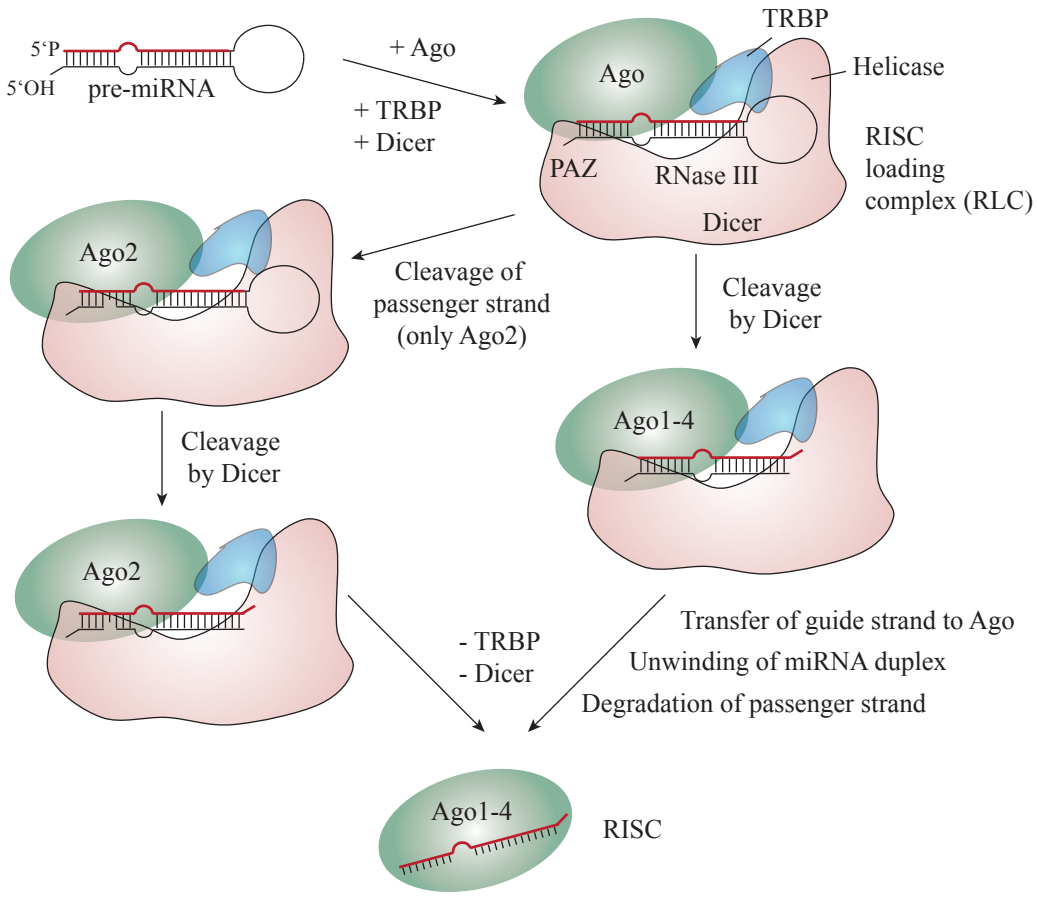


Figure 5: Formation of the miRNA induced silencing complex (miRISC). Dicer associates with TRBP and Ago to form the ternary RISC loading complex (RLC). After cleavage by Dicer, the miRNA duplex is partially released and repositioned with the help of TRBP to enable Ago loading. Only one strand serves as guide strand, while the other strand (passenger strand) is degraded. For some miRNAs, Ago2 cleaves the passenger strand prior to cleavage by Dicer.

1.4 MIRNA-MEDIATED GENE SILENCING

Among the three proteins that form the RLC only Ago is required in the active miRISC complex. Ago proteins are composed of an N-terminal, a PAZ, a Mid and a PIWI domain. As mentioned above, the PAZ domain specifically binds to the 2 nt 3′ overhang [80], while the Mid domain recognizes the 5′ end

phosphate of mature miRNAs [81, 82]. The PIWI domain has a high structural similarity to the RNase H superfamily. All four human Ago proteins (Ago1-4) are involved in RISC loading, however only Ago2 has catalytic residues within its PIWI domain to perform endonucleolytic cleavage of target mRNAs [83, 84].

Although the first discovered miRNA-binding sites were exclusively found in the 3′ UTR of mRNAs (see Section 1.1), an experimental genome-wide based study of Ago mRNA-binding sites revealed that miRNA-binding occurs to 37% in the exon coding regions and to about 60% in the 3′ UTR [85]. However, miRNA-binding sites in coding regions were demonstrated to have a lower regulatory efficacy compared to those in the 3′ UTR [86, 87]. Additionally, conservation of miRNA-binding sites, their structural accessibility, as well as the number of binding sites within the mRNA influences their functionality.

Typically, target mRNAs are not degraded since the lack of complementarity between miRNAs and their target mRNAs prevents the mRNA from being cut by Ago2. However, perfect pairing of miRNA with associated target mRNA led to endonucleolytic cleavage and rapid degradation of target mRNAs [88, 89]. Beside this, two different mechanisms for miRNA-mediated gene silencing have been described: i) inhibition of translation of target mRNA ii) destabilization and subsequent degradation of mRNAs [90] (see Figure 6).

1.4.1 *MiRNA-mediated repression of translation*

The miRISC complex was shown to repress translation by interfering with eIF4E-cap recognition and subsequent recruitment of the 40S ribosome [91, 92]. The eIF4E (eukaryotic initiation factor 4E) shares some similarities to Ago's central domain and thus it was proposed that Ago can compete with eIF4E in binding the 7-methylguanylate cap (m^7G) [93]. Additionally, miRISC complexes were shown to recruit eIF6, a factor that is known to antagonize association between the 60S and 40S ribosomal subunits [94]. A number of recent studies provided evidence that miRNA represses translation at postinitiation steps by inhibiting ribosome elongation [95, 96, 97]. This hypothesis relies on an early study in *C. elegans*, in which the *lin-4* miRNA targets *lin-14* and *lin-28* mRNAs stably associated with polysomes [98, 99]. The presence of mRNAs in polysomes is normally associated with a strong translational activity. However, in these studies *lin-14* and *lin-28* mRNAs were poorly translated, indicating that translation is inhibited at postinitiation steps due to the presence of *lin-4* miRNA. Further studies suggested that miRNA induces a ribosome drop off during elongation or promotes the degradation of nascent peptides.

1.4.2 *MiRNA-mediated destabilization of mRNA*

In addition to inhibition of translation, miRNAs are also known to destabilize target mRNAs and thereby facilitate their subsequent degradation. For example, human *miR-125a* and *miR-125b* were shown to reduce Lin28 mRNA abundance through an accelerated deadenylation of the poly(A) tail [100]. A recent study used ribosome profiling along with total mRNA quantification to examine the impact of miRNAs with respect to mRNA destabilization versus translational

repression [101]. According to this study, miRNAs downregulated gene expression predominantly through mRNA destabilization, since lowered mRNA levels could explain most of the decreased protein production (84%). Moreover, mRNAs that were not degraded were less efficiently translated.

MiRISCs initiate mRNA destabilization through deadenylation via CAF1-CCR4-NOT, a deadenylase complex that is recruited by the miRISC associated protein GW182 (Gly-Trp repeat-containing protein of 182 kDa). Thereupon, the poly(A) binding protein (PABP) dissociates and thus allows removal of m^7G cap structure by DCP2 and its cofactors. Finally, the deadenylated and decapped mRNA is degraded by cellular 5′-3′ exonucleases like XRN1 (reviewed in [102]).

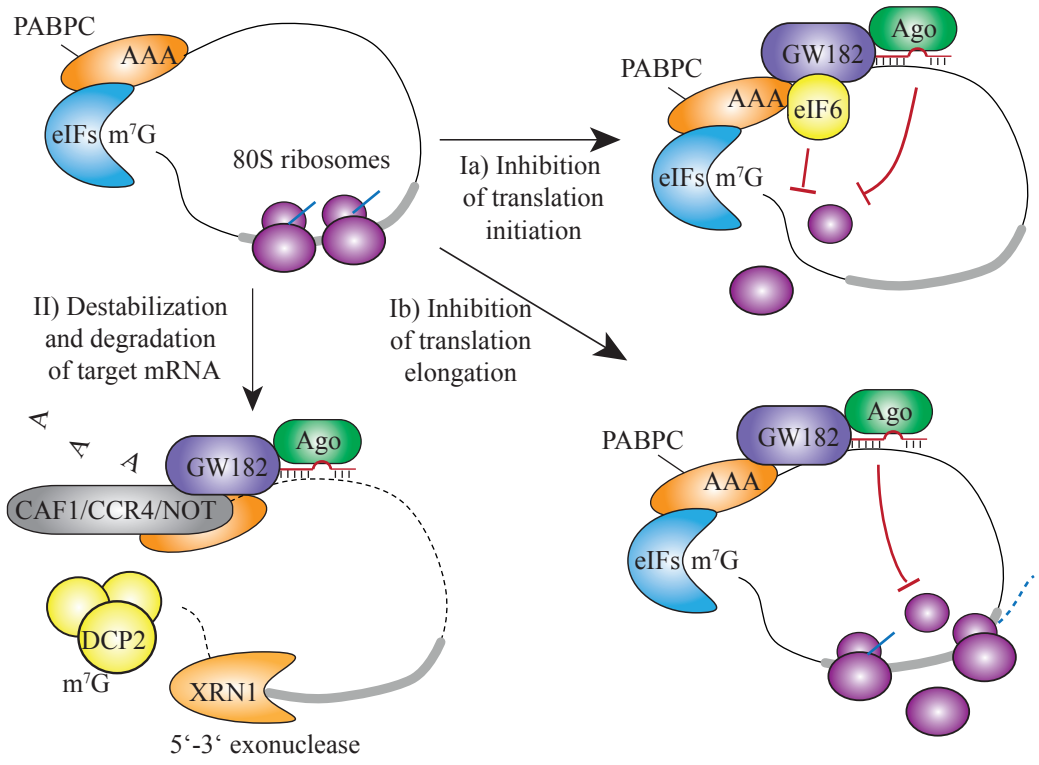


Figure 6: Mechanisms of miRNA mediated gene silencing. (**Ia**) The miRISCs can repress translation initiation by interfering with cap binding by eIF4E and recruitment of 40S ribosomal subunit. Additionally, they can prevent the 80S ribosomal complex formation by recruiting eIF6. (**Ib**) Translational repression by miRISC can also be mediated at postinitiation steps by inhibiting ribosome elongation, forcing ribosome drop off and facilitating degradation of the nascent peptide. (**II**) In most cases miRISCs destabilize target mRNAs and facilitate their subsequent degradation. Therefore GW182 recruits the deadenylate complex CAF1-CCR4-NOT and the decapping complex containing DCP2 and its cofactors. Decapped and deadyenaled mRNAs are rapidly degraded by cellular 5′-3′ endonucleases like XRN1. Abbreviations: eIFs: eukaryotic translation initiation factors, PABPC: poly(A) binding protein C.

1.4.3 *Processing bodies (P-bodies)*

P-bodies (sometimes also called GW-bodies) are dynamic cytoplasmic aggregates of translationally repressed mRNAs [103]. Their function is associated with storage and decay of repressed mRNAs. Consequently, components of mRNA decay and repression machinery are substantially enriched within these subcellular foci. Recently a link to the miRISC component was found, as Ago and GW182 proteins were shown to localize to these foci [104, 105]. Furthermore, knockdown of GW182 or inhibition of miRNA biogenesis caused a disappearance of P-bodies [106, 107]. Since P-bodies typically contain factors that are necessary for deadenylation (e.g. CCR4,CAF1,NOT), decapping (e.g. DCP1) or mRNA degradation (XRN1 5'-3' exoribonuclease), it is thought that these foci are primary sites of miRNA-mediated gene silencing and form as a consequence of this process [108].

1.5 MECHANISMS FOR REGULATING MIRNA PROCESSING

The canonical miRNA biogenesis pathway can be regulated at multiple steps by distinct mechanisms. These mechanisms lead to altered levels of mature miRNAs and comprise transcriptional control of miRNA genes as well as a regulation of miRNA processing factors. The latter is mediated by regulatory proteins that can be subdivided into four groups: i) Drosha binding/associated proteins, ii) Dicer binding associated proteins, iii) regulation by RNA-editing, iv) pre-element (preE) binding proteins (see Figure 7).

1.5.1 *Regulation of pri-miRNA processing by Drosha binding proteins*

Drosha can associate with at least 20 proteins in the Microprocessor complex [39]. These additional factors can enhance or decrease the processing efficacy for certain miRNAs [109, 110, 111, 112, 113].

The DEAD-box RNA helicases p68 and p72 are part of the Microprocessor complex and stimulate Drosha processing. However, they can also alter the processing of various pri-miRNAs by providing a scaffold for other proteins that either enhance or inhibit pri-miRNA processing by Drosha. For example, association of p68 with the tumor suppressor p53 stimulated Drosha processing of *miR-16-1*, *miR-143* and *miR-145* in response to DNA damage [112]. A similar effect was described for SMAD, a signal transducer of the TGFβ (tumor growth factor) and BMP (bone morphogenetic protein) signaling pathway. After stimulation with TGFβ or BMP4, SMAD localizes to the nucleus where it regulates transcription of target genes and associates with pri-*miR-21*, p68 and Drosha. As a result, pri-*miR-21* processing was significantly increased [110].

On the other hand, p68 and p72 were shown to selectively inhibit pri-miRNA processing in response to estrogen. Upon activation, the estrogen receptor was recruited by p68/p72 to the Microprocessor complex leading to a decreased processing of pre-*miR-16*, pre-*miR-143*, pre-*miR-195* and pre-*miR-125a* [113]. A heterodimeric complex between the nuclear factors NF45 and NF90 negatively regulates the processing of a number of miRNAs including pri-*miR-21* and pri-

let-7a. These factors were shown to bind to the ds stem and block the accessibility for DGCR8.

A recent study provided evidence that pri-miRNA processing by Dicer can occur co-transcriptionally and can be enhanced by a component of the nuclear capping complex called ARS2 (Arsenite-resistence protein 2) [114]. ARS2 was shown to interact directly with Dicer thereby coupling co-transcriptional capping and pri-miRNA processing. Furthermore, depletion of ARS2 lead to reduced levels of antiproliferative miRNAs such as *let-7*, *miR-21* and *miR-155* [115].

1.5.2 *Regulation of pre-miRNA processing by Dicer binding proteins*

Apart from its helper functions in Dicer processing and RISC loading, TRBP can selectively influence the processing of certain miRNAs in a phosphorylation-dependent manner. Paroo and colleagues showed that TRBP is phosphorylated by the mitogen-activated protein kinase (MAPK) signaling pathway [116]. Phosphorylated TRBP stabilized the RLC and resulted in reduced levels of antiproliferative miRNAs like *let-7a* while levels of growth-promoting miRNAs were increased.

1.5.3 *Regulation of miRNA biogenesis by miRNA editing*

The post-transcriptional change of RNA sequences by deamination of adenosine to inosine (RNA editing) can influence pri- and pre-miRNA processing by Drosha and Dicer, respectively. RNA editing is normally mediated by adenosine deaminases (ADAR) and results in altered base pairing and structural properties. The first miRNA that was linked to RNA editing was pri-*miR-22*; A-to-I editing was observed within the ds stem close to the Dicer cleavage site [117]. Since then, a variety of other pri- and pre-miRNAs were shown to be edited by ADAR1/ADAR2 and typically linked to reduced Drosha or Dicer processing [118, 119, 120]. Furthermore, editing of pri-*miR-142* led to increased degradation by the ribonuclease Tudor-SN, which preferentially cleaves I-U base pairs [121]. However, depending on the site of deamination, miRNA editing was also shown to have a stimulating effect on Drosha processing and thus might redirect mature miRNAs to new cellular mRNA targets [122, 119].

1.5.4 *Regulation of miRNA biogenesis by pre-element (preE) binding proteins*

Proteins that bind directly to miRNAs can also regulate miRNA processing. For example, the alternative splice factor hnRNP A1 (heterogeneous ribonucleoprotein A1) was shown to promote processing of pri-*miR-18a* by binding to its preE and inducing a structural change [123]. The structural rearrangements probably facilitate a subsequent binding of the Microprocessor complex and stimulate pri-miRNA cleavage by Drosha [124, 125]. Conversely, binding of hnRNP A1 to the terminal loop of pri-*let-7a* was associated with reduced Drosha processing by antagonizing the binding of the KH-type splicing regulatory protein (KSRP) [126]. In the absence of hnRNP A1, KSRP recognizes a GGG-motif within the

terminal loop of pri- and pre-*let-7a* and promote processing by Drosha and Dicer probably by favorable protein-protein interactions [127].

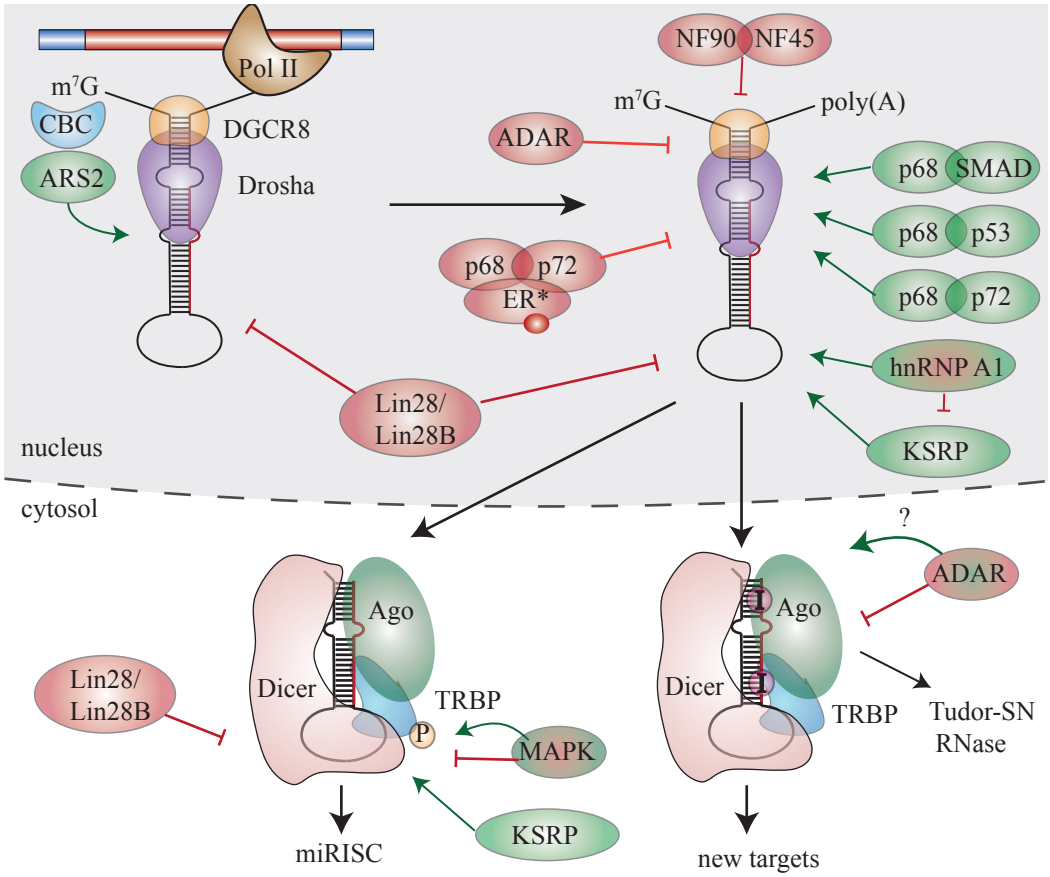


Figure 7: Regulators of miRNA processing. Several factors regulate biogenesis of different miRNAs at multiple levels. ARS2 binds to the cap binding complex (CBC) and recruits Drosha/DGCR8 to the nascent transcripts of pri-*miR-21*, pri-*miR-155* or pri-*let-7* thereby promoting their co-transcriptional processing. The RNA-helicases p68 and p72 are part of the Microprocessor complex and stimulate pri-miRNA processing. In addition, they act as a scaffold for other regulatory proteins that either increase or decrease Drosha processing for certain miRNAs. ADARs deaminate adenosine in a number of pri- and pre-miRNAs and thus inhibit their processing and subject them to degradation by the Tudor-SN ribonuclease. Site-dependent deamination is also known to enhance processing and lead to mature miRNAs with different mRNA targets. Proteins that bind to the terminal loop can selectively influence processing of certain miRNAs. For example, Lin28/Lin28B binds to the terminal loop of pre-*let-7* miRNAs thereby inhibiting Dicer or Drosha-mediated processing. By contrast, the splicing factor hnRNP A1 was shown to bind to a conserved motif within the preE of pri-*miR-18a* leading to enhanced processing. In addition, hnRNP A1 antagonizes KSRP in binding and promotes the biogenesis of pri- and pre-*let-7a*. Dicer processing is further regulated by MAPK-dependent phosphorylation of TRBP. Phosphorylated TRBP was shown to stabilize the RLC and thus enhanced processing of growth promoting miRNAs.

Pri- and pre-*let-7* biogenesis are negatively regulated by the pluripotence factor Lin28 or its paralog Lin28B. These proteins bind to the preE of both pri- and pre-miRNAs and inhibit processing by Drosha and Dicer [128, 129, 130]. More-

over, Lin28 promotes 3′ end polyuridylation of pre-miRNAs leading to constant inhibition of Dicer processing. Furthermore, the poly(U) tail serves as a signal for cellular nucleases and thus facilitates degradation of pre-*let-7* miRNAs [131, 132]. A more detailed description of the Lin28/*let-7* regulatory circuit is described in following sections.

1.6 *let-7* miRNA

As already described in Section 1.1, *let-7* miRNA is essential in *C. elegans*, as mutations within this gene cause developmental abnormalities and lethality [133, 134]. The *let-7* family is highly conserved in animal species and some bilaterians [16]. They typically comprise several genes even though the number of *let-7* copies differs considerably from species to species. While *C. elegans* only contains one *let-7* gene, the human genome encodes for 13 different *let-7* precursors that are processed into 10 different mature miRNAs (reviewed in [135], Table 1).

Table 1: *let-7* family members are highly conserved in animals but differ in the total number of *let-7* copies between organisms.

	C. elegans (nematode)	*D. rerio* (zebrafish)	*X. tropicalis* (frog)	*M. musculus* (mouse)	*H. sapiens* (human)
let-7a	1	6	1	2	3
let-7b	-	1	1	1	1
let-7c	-	2	1	2	1
let-7d	-	2	-	1	1
let-7e	-	1	2	1	1
let-7f	-	1	1	2	2
let-7g †	-	2	1	1	1
let-7h	-	1	-	-	-
let-7i	-	1	1	1	1
let-7j	-	1	-	-	-
miR-98	-	-	1	1	1
miR-202 ‡	-	1	1	2	1
total	1	19	10	14	13

Numbers in the table represent the number of *let-7* copies in each corresponding species.
† The sequences of *let-7g* differ between human/mouse and zebrafish/frog.
‡ Even though all species are labeled *miR-202*, each has a different sequence.

The *let-7* family is defined by the seed sequence (nt 2 to 7, GAGGUA), which is highly conserved even between different species. The preE is hardly conserved except for a GGAG or GGAG-like motif that was shown to be essential for Lin28 binding and terminal uridylation [132]. Despite low sequence conservation, members of the pre-*let-7* family tend to form small stems within their

preE, thereby creating an hairpin and an internal loop. However, these structures are not stable, and for most pre-*let-7* miRNAs, multiple preE secondary structures are predicted (Figure 8).

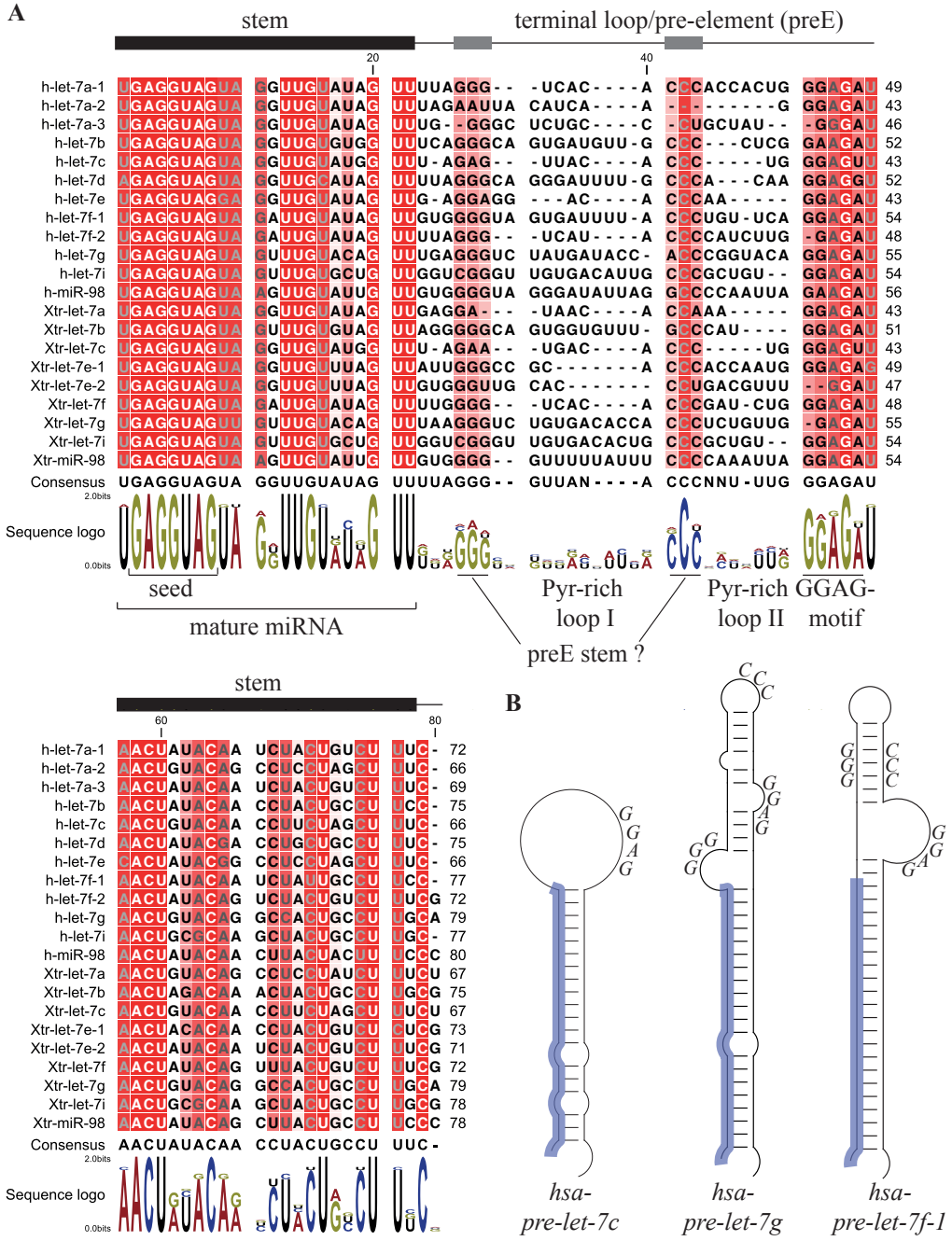


Figure 8: Sequence alignment of pre-*let-7* miRNAs. (**A**) The *let-7* family is defined by their seed sequence GAGGUAG from nt 2 to 8. While the seed and mature *let-7* sequences are well conserved even between species, the preEs show little sequence conservation except for a GGAG motif. In addition, pre-*let-7* preEs form diverse secondary structures that mostly contain a hairpin and internal loop as well as a small ds stem. (**B**) Schematic representation of the lowest energy secondary structure of *hsa-let-7c*, *hsa-let-7g* and *hsa-let-7f* according to CLC-workbench 3.6.5. The guide strand is indicated in blue.

In humans, *let-7* family members are arranged in 9 different clusters predominantly located in intergenic positions. However, not all *let-7* family members exhibit the same expression pattern. In addition, they might have different targets, since not all *let-7* miRNAs can compensate for each other, indicating that they are involved in different biological processes [136]. As of yet, *let-7* miRNAs are known to regulate stem cell differentiation, neuromusculature development [137, 138], limb development [139], cell proliferation and differentiation [140, 141]. In addition, *let-7* is a potential tumor suppressor, since it negatively regulates proto-oncogenes and cell cycle regulators such as Ras, Hmga2, c-Myc, Cdc25a, Cyclin D, Cdk6 and Lin28 [142, 143, 144, 145, 130]. Conversely, the transcription of the major *let-7* cluster is repressed by the proto-oncogenic transcription factor c-myc, which explains why *let-7* typically appears late in development [146].

1.7 LIN28

1.7.1 *General features*

Lin28 (cell LINeage abnormal 28) is a conserved RNA-binding protein in higher eukaryotes that regulates several important cellular functions associated with development, glucose metabolism, differentiation and pluripotency. It was first described as a heterochronic gene in *C. elegans* since mutations within *lin-28* disturbed the developmental progression of the worm [133]. Further experiments revealed that Lin28 is expressed early in nematode embryonic and larval development, but its expression decreases as differentiation proceeds (see Section 1.1) [11, 99]. Similarly, the mammalian homologs Lin28A and Lin28B encode for basic 23 or 28 kDa proteins that are highly expressed in embryonic stem cells (ESC) but are downregulated upon differentiation of ESCs into embryoid bodies [147]. Conversely, Lin28 could be used to reprogram adult human fibroblasts to induced pluripotent stem cells in cooperation with Oct4, Sox and Nanog [140]. Furthermore, knockdown of Lin28A expression in mouse ESC led to loss of Oct4 and Nanog expression, indicating an impaired self-renewal potential [132]. These observations, in conjunction with the fact that *lin-28a* and *lin-28b* are reactivated in about 15% of all analyzed cancers [128], underscore the functional relevance of Lin28 in maintaining and reconstituting pluripotency. Recently, Lin28 was also linked to the regulation of metabolic processes. Overexpression of Lin28 in mice was associated with increased insulin sensitivity and glucose metabolism [148], while a depletion of Lin28 resulted in insulin resistance and glucose intolerance [149].

On the molecular level, Lin28 acts both as a negative regulator of *let-7* biogenesis and as a post-transcriptional regulator of mRNA translation (reviewed in [150]). Both activities strongly depend on its two RNA-binding domains (RBDs): an N-terminal cold-shock domain (CSD) and a C-terminal Zn-knuckle domain (ZKD) composed of two tandemly arranged retroviral-type CCHC Zn-knuckles (Figure 9). The combination of both RBDs is unique in animals. In addition, Lin28 proteins contain low complexity regions at the C-terminus, one bipartite nuclear localizing sequence (NLS, in Lin28A and Lin28B) as well as one pu-

tative nucleolar localizing sequence (NoNLS, in Lin28B) [151]. Despite their NLS sequences, Lin28 predominantly localizes to the cytosol with only minor accumulation in the nucleus [152, 153, 130]. Within the cytosol Lin28 associates with poly(A) binding proteins and translating mRNAs, whereas under stress conditions, Lin28 localizes to stress granules and P-bodies. Upon mutation of both Lin28 RBDs, Lin28 accumulated in the nucleus, suggesting that it exits the nucleus in complex with bound RNA [153].

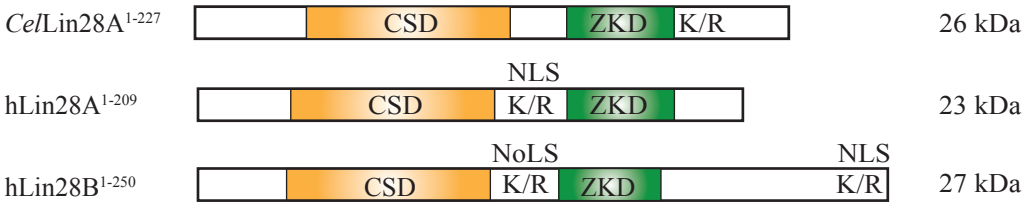


Figure 9: Lin28 domain organization. Lin28 contains two RNA-binding domains (RBDs): an N-terminal cold-shock domain (CSD) and a C-terminal Zn-knuckle domain (ZKD) comprised of two retroviral type CCHC Zn-knuckles. Additionally, Lin28 harbors Lys/Arg (K/R)-rich stretches and putative nuclear (NLS) and nucleolar localization sequences (NoLS).

1.7.2 *Lin28 inhibits let-7 biogenesis and promotes pre-let-7 uridylation*

As previously discussed for *C. elegans* larval development, Lin28 and mature *let-7* exhibit opposing expression patterns. This inverse relationship between Lin28 and *let-7* miRNA is also present in mammalian cells, whereby Lin28 is mainly expressed in undifferentiated cells, and mature *let-7* is only detectable upon differentiation or tissue development. Furthermore, levels of pri-*let-7* remain constant throughout the entire differentiation or development process suggesting a negative regulation of *let-7* biogenesis by Lin28 in stem or progenitor cells [154, 155, 156, 141]. Purification of pre-*let-7* bound complexes and subsequent analysis via mass spectrometry revealed that both human Lin28 paralogs (Lin28A and Lin28B) specifically associate with pre-*let-7* *in vivo* [129, 128]. Furthermore, *in vitro* purified Lin28 could inhibit pri- and pre-*let-7* processing by Drosha and Dicer by binding to the upper stem and preE [157, 129] (Figure 10). Mutations in the Lin28 CSD and in the ZKD impaired pre-*let-7* binding and inhibition of Dicer processing, suggesting a competitive relationship between Lin28 and Dicer. Moreover, a recent study provided evidence that Lin28 induces a structural change within pre-*let-7*'s preE, thereby leading to opening of the ds stem including Dicer cleavage site [158].

An additional inhibition mode of *let-7* processing was revealed by Heo and colleagues, which irreversibly targets pre-*let-7* to a decay pathway [131]. They demonstrated that Lin28 induces polyuridylation of pre-*let-7*'s 3′ overhang. Polyuridylated pre-*let-7* is resistant to Dicer cleavage given that Dicer normally recognizes the 2 nt 3′ overhang via its PAZ domain. Thus Dicer is unable to recognize the elongated 3′ overhang and to process pre-*let-7*. Furthermore, polyuridylated RNAs normally recruit 3′-5′ exonucleases and thus serve as a signal for degradation [159, 160]. Indeed, polyuridylated pre-*let-7* was more

rapidly degraded than unmodified pre-*let-7* [131]. Subsequent studies revealed that polyuridylation of pre-*let-7* is catalyzed by the non-canonical poly(A) polymerase TUT4 (terminal uridyl transferase 4) and to a minor extent by TUT7 in a Lin28-dependent manner [132, 161]. Interestingly, these enzymes catalyze mono-uridylation of pre-miRNAs with a 1 nt 3′ overhang (like most pre-*let-7* family members) in the absence of Lin28, thereby enhancing Dicer mediated processing [162]. However, in the presence of Lin28, pre-*let-7* and other miRNA containing a GGAG-motif within their preE were subjected to polyuridylation. Upon mutation of this motif, both Lin28 binding and polyuridylation were impaired, indicating that the GGAG-motif is essential for both processes [132] (Figure 10).

In *C. elegans* a similar mechanism for inhibiting pre-*let-7* processing has been recently reported [163]. The poly(U) polymerase PUP2 was shown to polyuridylate pre-*let-7* in a Lin28-dependent fashion, thereby avoiding premature expression of mature *let-7* during larval development. However, regulation at the Dicer step could not explain the oscillating expression pattern of different *let-7* forms. During larval stage L2 only pri-, but not pre-*let-7*, accumulates in wild-type (WT) *C. elegans*, but not in a *lin-28* mutant [164]. Subsequent RNA and chromatin immunoprecipitation assays revealed a specific interaction between Lin28 and pri-*let-7* that co-transcriptionally inhibits pri-*let-7* processing by Drosha.

This co-transcriptional interaction between Lin28 and pri-*let-7* is also valid for human ESCs [164]. Moreover, it was suggested that Lin28B predominantly localizes to the nucleoli, where it sequesters pri-*let-7*, thereby preventing Drosha processing in the nucleus [151]. Thus, Lin28 seems to obviate precocious expression of mature *let-7* during early development and differentiation, by interfering with both the Drosha and Dicer complexes, and by targeting pre-*let-7* towards degradation. Conversely, upon differentiation of stem or progenitor cells, *let-7* ensures constant downregulation of Lin28 by binding to the 3′ UTR of Lin28 and its promoting transcription factor c-Myc [130].

1.7.3 *Lin28-mediated regulation of translation*

Besides regulation of *let-7* biogenesis, Lin28 can interact with various mRNAs and modulate their translation. Polesskaya *et al.* revealed that Lin28 can associate with polysomes and enhance translation of a number of mRNAs in differentiating myoblasts [166]. Among the first mRNA targets that had been identified was IGF-2 (insulin-like growth factor 2), a major growth and differentiation factor in muscle tissue. Further evidence was provided that Lin28 recruits IGF-2 mRNA to polysomes and enhances its translation, via interactions with components of the translation initiation complexes. Subsequent studies revealed a number of additional mRNA targets of Lin28 in mouse ESCs such as H2a (Histone 2a), HMGA1, Cyclin A, Cyclin B, Cdk4 and Oct4 [167, 168, 169, 170]. An association of Lin28 with most of these mRNAs was connected with enhanced translation, suggesting that Lin28 maintains pluripotency by stimulating the translation of corresponding cell cycle effectors. Further genome-wide studies suggested that Lin28 facilitates translation of genes important for growth and survival in human ESCs, by recruiting a RNA helicase A (RHA) to polysomes

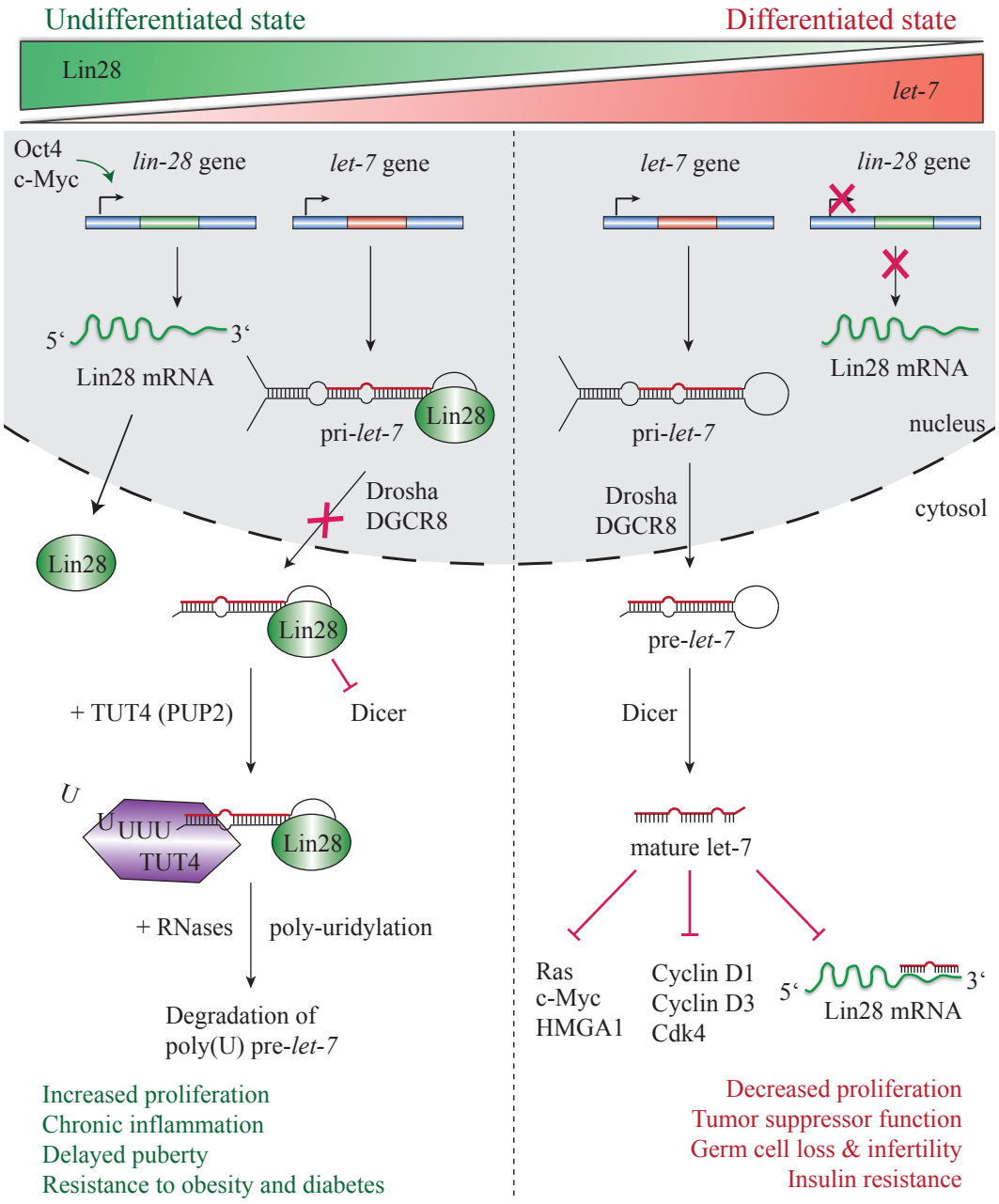


Figure 10: Lin28/*let-7* regulatory axis. In undifferentiated cells, Lin28 is highly expressed and blocks the biogenesis of *let-7* miRNA. By binding to the preE of pri- or pre-*let-7*, neither Drosha nor Dicer can process the corresponding *let-7* precursor. In addition, Lin28 recruits TUT4 (or PUP2 in *C. elegans*) to pre-*let-7* and promotes it 3' end polyuridylation. Polyuridylated pre-*let-7* cannot be cleaved by Dicer and thus serves as a signal for cellular 3'-5' ribonucleases (RNases). Upon differentiation, Lin28 expression is reduced, which leads to increased levels of mature *let-7*. The latter silences gene expression of proto-oncogenes (Ras, c-Myc, HMGA2) and cell cycle progression factors (Cyclin D1 and D3, Cdk4) and Lin28 itself, thereby establishing a positive-feedback loop. In addition to differentiation, the Lin28/*let-7* regulatory network is apparently involved in several cellular processes like proliferation, oncogenesis, development and physiology, as well as metabolism (recently reviewed in [165].

[171, 172]. Additional mutagenesis studies revealed that the C-terminal part of Lin28 is required for RHA interactions, while mutations in the ZKD only impaired the stimulatory impact on translation, but not protein-protein interactions [170].

Very recently, two genome-wide crosslinking and immunoprecipitation coupled with high-throughput sequencing (CLIP-seq) studies were conducted in human and mouse ESCs as well as somatic cells [173, 174]. These experiments revealed that Lin28 binding was observed in 25% of all human transcripts [173]. Surprisingly, only a small fraction of the identified targets could be traced back to miRNAs (0.07%), while the majority was mapped to mRNAs (42%) and ribosomal RNAs (17%) [174]. However, considering the different populations of each species in the cell, the validity of these numbers remains to be seen. Additionally, both studies showed some discrepancies with respect to preferential mRNA interactions of Lin28 and the related physiological function. According to Wilbert *et al.* [173] Lin28 preferentially associates with mRNAs encoding for RNA processing and splicing factors, thereby enhancing their translation. In contrast, Cho *et al.* reported preferential interaction with mRNAs that are destined for the endoplasmic reticulum. The latter led to translational repression of the corresponding mRNAs by reducing ribosome occupancy [174].

1.8 COLD-SHOCK DOMAINS (CSDS)

CSDs are highly abundant nucleic acid-binding domains that are widely distributed in bacteria, animals and plants. In eukaryotes, auxiliary domains in addition to the CSD have broadened the biological function of CSD-containing proteins. With this architecture, they fulfill pleiotropic functions mainly related to RNA metabolism and regulate complex biological processes such as stress tolerance, development and differentiation (reviewed in [175]).

CSD-containing proteins were first discovered in bacteria since members of this family are induced during bacterial cell response to cold-shock conditions (reviewed in [176, 177]. These so-called major cold-shock proteins (Csps) are known to bind pyrimidine-rich ssRNA/DNAs with dissociation constants in the subnanomolar to micromolar range [178, 179, 180, 181, 182, 183]. They work as RNA chaperones that destabilize RNA secondary structures, which may impede transcription or translation [184, 185]. In addition, *E. coli* CspA, CspC and CspE were characterized as transcription antiterminators by destabilizing RNA hairpins in mRNAs that otherwise act as Rho-independent transcriptional terminators [186].

From a structural perspective, the CSD belongs to the oligonucleotide/oligosaccharide-binding (OB) fold and consist of a β-barrel, composed of five antiparallel β-strands. The CSD typically comprises 70 amino-acid residues and is characterized by a prominent amphipathic surface. One side of the protein forms a hydrophobic platform surrounded by individual basic side chains and thus accommodates the moderately conserved ribonucleoprotein motifs RNP1 (GF/YGFI) an RNP2 (DV/LFVH) (Figure 11 A,B).

Crystal structures of hexathymidine (dT_6) bound to bacterial *Bacillus subtilis* CspB (*Bs*-CspB) and *Bacillus caldolyticus* Csp (*Bc*-Csp) revealed a common mode

of DNA-binding to CSDs [180, 182]. Binding is dominated by stacking interactions between aromatic protein sidechains and thymine bases, while very few specific hydrogen bonds are mediated with the bases. Very recently, a similar binding mode was observed for RNA-oligonucleotides bound to *Bs*-CspB (Figure 11 C) [187]. Since the sugar-phosphate backbone is surface exposed, and barely involved in protein•RNA interactions, no substantial difference in binding mode and strength between ssDNA and ssRNA oligonucleotides was observable. Only the additional hydrophobic contribution of the 5-methyl group of T in comparison to U led to higher binding affinities of dT_7 compared to rU_7.

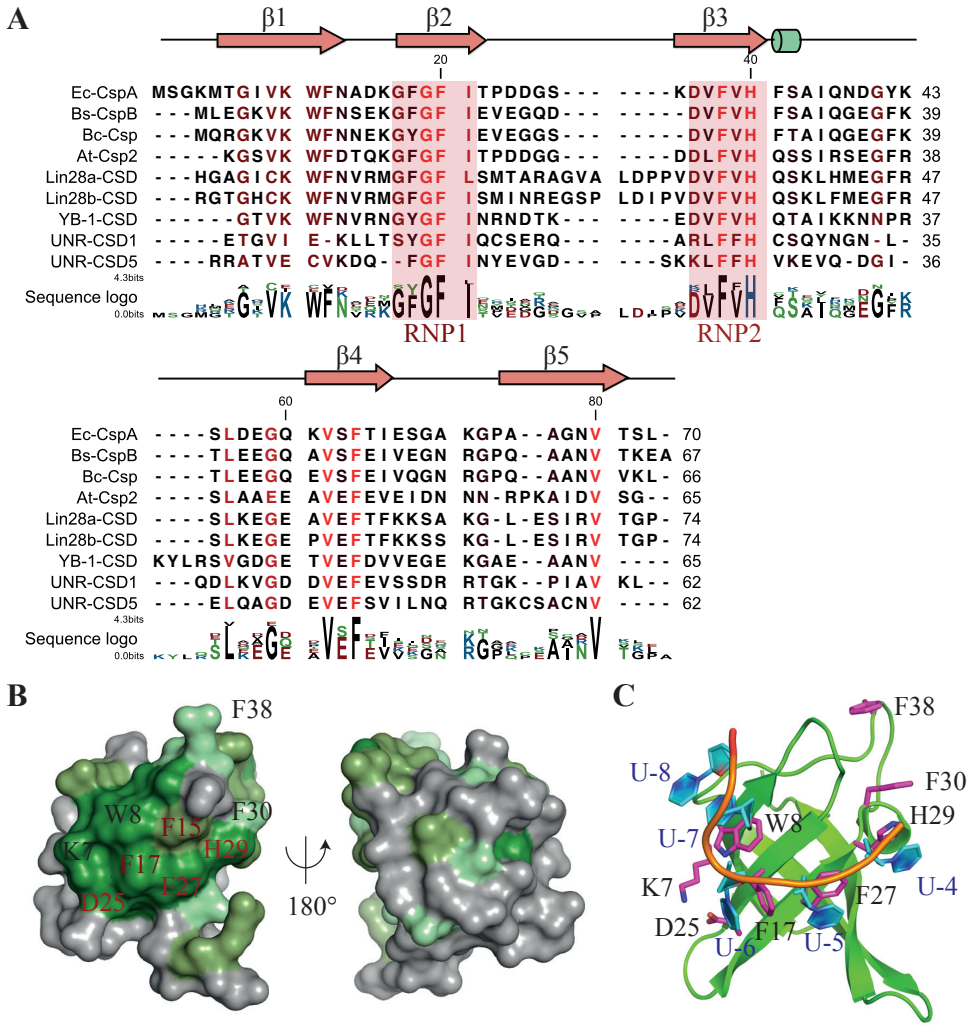


Figure 11: (**A**) Structure based sequence alignment of Lin28 CSDs with other eukaryotic and bacterial CSDs. The secondary structure of *Bs*-CspB (PDB 19C0) is indicated above the sequences. Conserved amino acid residues are labeled in dark-red (100% conserved) to light red (at least 70% conserved). (**B**) Sequence conservation mapped to the surface of *Bs*-CspB. Depending on their relative sequence conservation, residues were colored from dark green (>90%) over olive green (>70%) and pale green (>50%) to gray. (**C**) Structure of *Bs*-CspB•rU_6 (PDB 3PF5). Conserved residues important for RNA binding are depicted as sticks.

1.9 RETROVIRAL-TYPE CCHC ZN-KNUCKLES

Retroviral-type CCHC Zn-knuckles are known to bind ssRNA and ssDNA sequences [188]. Originally, they were identified as repeating Cys-X_2-Cys-X_4-His-X_2-Cys motifs (X: variable amino acid) in retroviral nucleocapsid (NC) proteins that form zinc-binding domains [189, 190]. Like Lin28, HIV-1 NC contains two tandemly arranged CCHC Zn-knuckles forming a functional Zn-knuckle domain (ZKD) that specifically binds GGAG or GGUG stem loops within the HIV-1 ψ packaging signal [191] (Figure 12 A, B). NMR solution structures of HIV-1 ZKD with these stem loops revealed that each Zn-knuckle specifically recognizes one G via a hydrophobic pocket and specific hydrogen bonds with the bases [191, 192, 193]. In addition, the N-terminal tail forms a 3_{10} helix that packs against the C-terminal part and thus interacts nonspecifically with the ds stem (Figure 12 C). It is likely that the Lin28 ZKD mediates specific binding to the conserved GGAG motif within the terminal loop of pre-*let-7*. Recent studies support this hypothesis, since ZKD-mutations impaired pre-*let-7* binding and uridylation by TUT4 [132, 194].

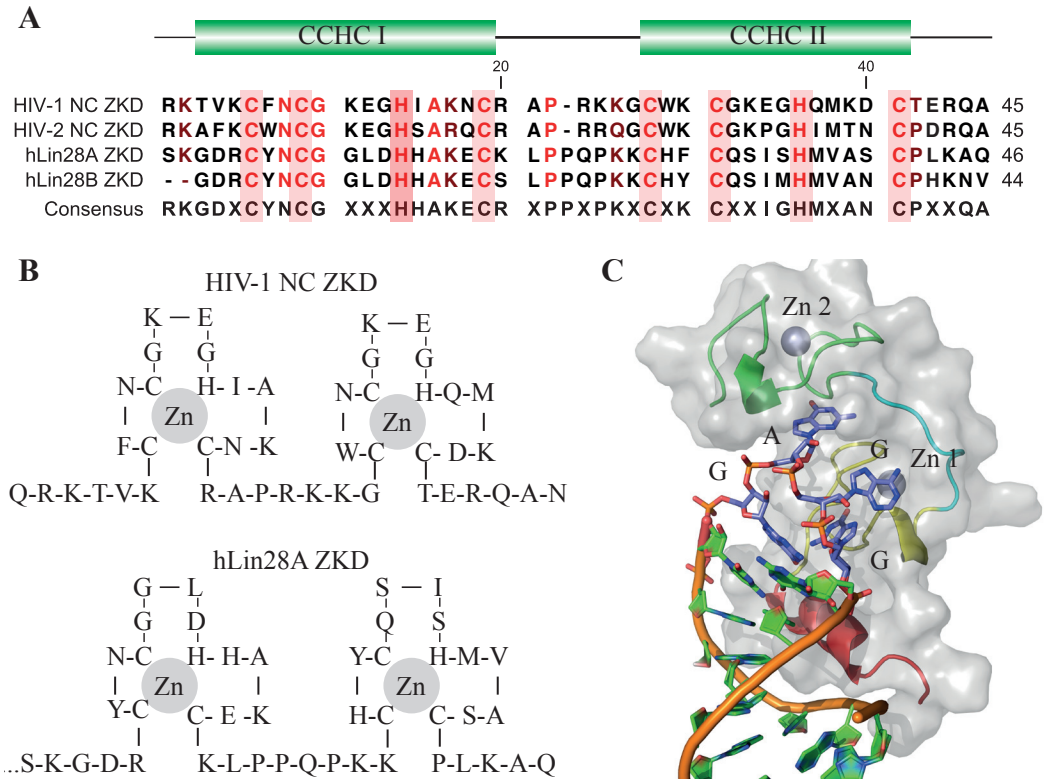


Figure 12: Comparison between HIV nucleocapsid protein (NC) ZKDs and Lin28 ZKDs. (**A**) Sequence alignment of HIV-1 NC, HIV-2 NC, hLin28A and hLin28B ZKDs. The chelating Cys and His residues of the CCHC Zn-knuckles are shaded in red. Conserved residues are labeled from dark red (100% type-conserved) to light red (70% type-conserved). (**B**) Secondary structure of HIV-1 NC ZKD and hLin28A ZKD. (**C**) Cartoon representation of a NMR solution structure from HIV-1 NC ZKD bound to the SL3 stem loop recognition element of the retroviral Ψ RNA packaging signal (PDB-ID 1A1T). The ZKD specifically binds to the GGAG tetraloop, which is shown in sticks.

The two CCHC Zn-knuckles can act independently from each other since they are separated by a flexible linker which can vary in length. A recent study demonstrated that each Zn-knuckle recognizes single G's and thus the length of linker between the knuckles affects spacing between the G's [188]. Interestingly, the terminal uridyl transferases TUT4 and TUT7 also contain two CCHC ZKDs despite that the linker between the two Zn-knuckles (37 amino acids) is much longer than for Lin28 ZKD (7 amino acids).

1.10 SCOPE OF THIS WORK

Lin28 is an essential regulator that controls several important cellular processes like cell differentiation, development and metabolism. Its pleiotropic functions are mediated both by antagonizing *let-7* miRNA biogenesis and modulating mRNA translation of various genes. Both processes are strongly dependent on its two RBDs, a cold-shock domain (CSD) and a Zn-knuckle domain (ZKD). The combination of both domains is unique in animals and thus gives rise to the question how these two domains act in concert to specifically recognize the appropriate RNA targets.

Therefore, the goal of this doctoral thesis was to elucidate the molecular mechanisms of Lin28-mediated *let-7* precursor binding and processing inhibition. To pursue this task detailed structural and functional analyses of Lin28 nucleic acid and protein-protein interactions were performed to address the following questions.

- What is the structural basis of Lin28 binding to pre-*let-7* miRNA?
- What are the contributions of the isolated Lin28 RBDs in RNA binding?
- How can Lin28 specifically recognize twelve different *let-7* precursors and inhibit their processing by Drosha or Dicer?
- How does Lin28 promote polyuridylation of pre-*let-7* miRNA via TUT4/PUP2?

2 MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 *Instruments*

Instruments used in this thesis are listed in Appendix A, Table 16.

2.1.2 *Chemicals and media*

Chemicals and media were purchased from the following companies: Carl Roth (Karlsruhe, D), Sigma-Aldrich (St. Louis, USA), Calbiochem (Darmstadt, D), Jena Bioscience (Jena, D), GE Healthcare (Piscataway, USA), Roche Applied Science (Penzberg, D), Biochrom (Berlin, D), Becton Dickinson (Franklin Lakes, USA), AppliChem (Darmstadt, D) and Serva (Heidelberg, D).

2.1.3 *Oligonucleotides*

Primer and DNA-oligonucleotides were obtained from Biotez (Berlin, D), Eurofins MWG Operon (Ebersberg, D), Biomers (Ulm, D) and Integrated DNA-Technologies (Coralville, USA). RNA-oligonucleotides were obtained from Thermo Scientific/Dharmacon (Lafayette, USA) and Biomers (Ulm, D).

2.1.4 *Enzymes*

Enzymes used in this study are listed in Appendix A, Table 17.

2.1.5 *Antibodies*

Antibodies used in this study are listed in Appendix A, Table 17.

2.1.6 *Kits*

Kits used in this study are listed in Appendix A, Table 18.

2.1.7 *Bacterial strains*

- *E. coli* DH5α $T1^R$ (Promega, Mannheim, D)
- *E. coli* DB3.1 $T1^R$ (Invitrogen, Carlsbad, USA); for propagation of plasmids containing the *ccdB* gene

- *E. coli* Rosetta 2 (DE3) $T1^R$ (Novagen, Darmstadt, D) with *pRARE2* plasmid (Cm^R) containing the tRNA genes *argU, argW, ileX, glyT, leuW, proL, metT, thrT, tyrU and thrU*

2.1.8 *Plasmids*

- pQLinkH (Ap^R, N-terminal heptahistidine-tag and TEV protease cleavage site, [195])
- pSKB2-LNB (Kan^R, N-terminal hexahistidine-tag and PreScission protease cleavage site, O. Daumke, MDC Berlin)
- pENTR4 (Kan^R, entry vector, for LR recombination with pDEST-vectors, Invitrogen, Carlsbad, USA)
- pDEST15 (Ap^R, N-terminal GST-tag, Invitrogen, Carlsbad, USA)
- pDEST17 (Ap^R, N-terminal hexahistidine-tag, Invitrogen, Carlsbad, USA)
- pDEST53 (Ap^R, N-terminal GFP-tag, Invitrogen, Carlsbad, USA)
- pDEST57 (Ap^R, N-terminal hexahistidine-NusA-tag, Invitrogen, Carlsbad, USA)
- pcDNA 3.1/nV5 (Ap^R, N-terminal V5-epitope, Invitrogen Carlsbad, USA)
- pcDNA 3.1/nFlag (Ap^R, N-terminal Flag-epitope, manually modified)
- pDEST-*renilla* (Ap^R, N-terminal protein A (PA) *renilla* luciferase (RL), Anup Arumughan, MDC Berlin, D)
- pDEST-*firefly* (Ap^R, N-terminal V5-epitope *firefly* luciferase (FL)-tag, Anup Arumughan, MDC Berlin, D)

2.1.9 *Composition of growth media and buffers*

The composition of all used growth media and buffers used in this work are listed in Appendix A, Table 19 and Appendix A, Table 20 to 23.

2.2 MOLECULAR BIOLOGY METHODS

2.2.1 *Polymerase chain reaction (PCR)*

DNA fragments for cloning and *in vitro* transcription were amplified with *Pfu* Plus polymerase (Roboklon, Berlin, D) according to standard procedures [196]. Constructs that were subsequently cloned into pENTR4 included the coding sequence for the TEV protease cleavage site in the forward primer. For colony PCR [196] and for other verification PCR, *Taq* polymerase (Roboklon, Berlin, D) was used.

2.2.2 *Agarose gel electrophoresis*

Depending on the sizes of corresponding DNA fragments, 0.5 to 2.0% agarose gels supplemented with 0.5 µg/ml ethidium bromide were prepared and developed in TAE buffer according to standard procedures [196]. Sizes of linear DNA fragments were analyzed using a UV lamp and estimated by comparison with a 100 bp or 1 kbp standard, respectively (Roboklon, Berlin, D).

2.2.3 *DNA restriction*

DNA was cut with type II restriction endonucleases from New England Biolabs (NEB, Frankfurt am Main, D) for ligation of amplified PCR fragments into the corresponding vector as described in the manufacturer's instructions. Double digests were used to enable directional cloning. Typically, the reactions were incubated for 1 to 2 h at 37 °C.

2.2.4 *DNA purification*

DNA fragments obtained from PCR reactions and restriction digests were purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, D) according to the manufacturer's protocol. If the PCR reaction generated multiple DNA fragments, the fragment with the correct size was excised with a scalpel from the agarose gel and purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, D) according to the manufacturer's instructions.

2.2.5 *Plasmid preparation*

Plasmids were prepared using the QIAPrep Spin Miniprep Kit or the QIAGEN Plasmid Maxiprep Kit (both Qiagen, Hilden, D) according to the manufacturer's protocol.

2.2.6 *Ligation*

A fivefold molar excess of insert was added to about 50 ng cut plasmid and incubated with T4 DNA ligase (NEB, Frankfurt am Main, D) according to the manufacturer's manual. Concentration of inserts and plasmids were determined using the absorption at 260 nm.

2.2.7 *Preparation of chemically competent E. coli*

Chemically competent *E. coli* bacteria were prepared according to [197].

2.2.8 *Transformation of chemically competent E. coli*

For transformation, the corresponding plasmid was incubated with 50 µl of competent cells for 5 to 15 min, followed by a heat shock at 42 °C for 60 s. After

the addition of 500 µl LB medium the cells were incubated at 37 °C for 45 min under constant agitation and then streaked on LB agar plates containing the appropriate antibiotics. *E. coli* DH5α was used for plasmid propagation and *E. coli* Rosetta 2 (DE3) for protein expression. Vectors encoding the *ccdB* gene required DB3.1 cells for plasmid propagation.

2.2.9 *Storage of E. coli bacteria*

Prior to storage of bacteria at -80 °C, 1 ml of a LB overnight bacteria culture was mixed with 0.8 ml sterile 86% (w/v) glycerol.

2.2.10 *LR recombination*

Inserts that were initially cloned into the entry vector pENTR4 (Section 2.1.8) were subsequently transferred into the appropriate destination vector via LR recombination. The LR recombination reaction was carried out with 150 ng of entry and destination vector in a total volume of 10 µl at room temperature for approximately 1 h.

2.2.11 *Site-directed mutagenesis*

Site-directed mutagenesis was conducted as described in the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, USA) using *Pfu* Plus polymerase (Roboklon, Berlin, D).

2.2.12 *Protein constructs*

An overview of all cloned constructs encoding for the protein of interest can be found in Appendix A, Table 24.

2.2.13 *Antibiotics*

The following antibiotics with the denoted concentrations were used in this study.

- Ampicillin (Ap), 100 µg/ml
- Chloramphicenol (Cm), 34 µg/ml
- Kanamycin (Kan), 50 µg/ml

2.3 EXPRESSION AND PURIFICATION OF PROTEINS

Purification of histidine-tagged fusion proteins comprised the following steps: I) Ni/Zn-NTA affinity chromatography, II) cleavage and removal of tag by hexahistidine-tagged TEV protease, III) reapplication of cleaved protein to Ni/Zn

affinity chromatography, IV) cation exchange chromatography using a heparin column, V) size-exclusion chromatography. For proteins containing Zn-fingers, Ni^{2+} HisTrap columns (GE Healthcare, Piscataway, USA) were charged with Zn^{2+} according to the manufacturer's protocol. Additionally, purification buffers were supplemented with 10 µM $ZnSO_4$ for these proteins. For GST-tagged proteins, the tag was cleaved on-column. Ni/Zn-NTA and GSH affinity chromatography were conducted on a peristaltic P-1 FPLC system. Cation exchange chromatography and size-exclusion chromatography (SEC) were performed on an Äkta explorer. The time course of the purification was monitored by measuring the absorption at 260 and 280 nm, and peak fractions were analyzed by SDS-PAGE (see Section 2.4.2). All chromatographic procedures were conducted on ice or at 4 °C. A detailed description of the individual chromatographic steps is given in following sections (Section 2.3.4 to 2.3.8).

2.3.1 *Small-scale protein overexpression and solubility test*

Overexpression conditions, solubility and binding to affinity matrix were tested for desired constructs in a small-scale format using 100 ml cultures. A single *E. coli* Rosetta 2 (DE3) colony containing the desired expression plasmid was grown in 10 ml LB medium supplemented with the appropriate antibiotics overnight. The pre-culture was diluted 100-fold in 100 ml selective TB or LB medium and grown at 37 °C under constant shaking to an OD_{600} of 0.8. After cooling to either 17 °C or 25 °C, protein overexpression was induced with 0.5 mM IPTG, and growth was continued for about 16 h. Cells were centrifuged for 10 min at 7000 g and 4 °C and the pellet was resuspended in 15 ml lysis buffer and sonified 6 times with a 50% pulse for 15 s on ice. The cell lysate was centrifuged at 50000 g and 4 °C for 30 min, sterile filtered through an 0.2 µm filter and applied on a pre-equilibrated gravity-flow column containing either 0.5 ml Ni-NTA agarose or GSH sepharose beads. The columns were washed with at least 10 column volumes (CV) Ni-NTA or GST wash buffer and eluted with 2 CV of the appropriate elution buffer (Ni-NTA or GST elution buffer). For documentation, 10 µl samples of each step were taken, diluted with 2x SDS sample buffer, boiled for 10 min at 95 °C and subjected to SDS-PAGE (see Section 2.4.2).

2.3.2 *Large-scale protein overexpression*

Depending on the construct, 4 to 12 l TB medium supplemented with antibiotics were inoculated with an overnight *E. coli* Rosetta 2 (DE3) culture containing the desired expression construct. Cells were incubated at 37 °C with shaking to an OD_{600} of 0.8, cooled to 17 °C or 25 °C and induced with 0.5 mM IPTG. After 16 h cells were harvested by centrifugation at 7000 g for 15 min and stored at -80 °C.

2.3.3 *Cell lysis*

Cell pellets were thawed on ice and resuspended in lysis buffer (3 ml per 1 g cell pellet) supplemented with one tablet of Complete Mini-EDTA free protease inhibitor as well as a DNase I or benzoase (Roche, Penzberg, D) and lysozyme (Sigma-Aldrich St. Louis, USA). The suspension was stirred at 4 °C and disrupted by passage through a microfluidizer for at least three times. The lysate was then centrifuged at 50000 g and 4 °C for 30 min and filtered through a 0.2 µm filter.

2.3.4 *Ni-NTA affinity chromatography*

Filtered cell lysates containing recombinant 7xHis or 6xHis-tagged proteins were applied on 5 ml HisTrap columns (GE Healthcare, Piscataway, USA) equilibrated with 5 CV of Ni-NTA equilibration buffer. After extensive washing with at least 10 CV of Ni-NTA wash buffer, the protein was eluted with 2 to 5 CV Ni-NTA elution buffer containing 250 mM imidazole.

2.3.5 *Tag removal*

For removal of affinity tags, all fusions proteins carried a specific recognition site for TEV protease (ENLYFQ|GS). TEV protease was recombinantly purified by a one-step Ni-NTA affinity purification and contained an non-cleavable N-terminal hexahistidine-tag for subsequent removal.

Eluted fractions from the Ni-NTA affinity chromatography were pooled and dialyzed against Ni-NTA equilibration buffer in the presence of 1 mg TEV protease per 10 mg of recombinant protein overnight. The next day the dialyzed protein solution was subjected to another Ni-NTA affinity chromatography step in the presence of 10 µM imidazole to avoid unspecific binding. The flow-through fractions containing protein of interest were collected and used for subsequent purification steps.

2.3.6 *GSH affinity chromatography and on-column cleavage*

Recombinant proteins fused to the 27 kDa GST-tag and a TEV protease cleavage site were purified using 5 ml GST-Trap columns (GE Healthcare, Piscataway, USA).

For purification, GST-Trap columns were equilibrated with 5 CV lysis buffer before applying cell lysate to the columns. The affinity matrix was subsequently washed with 20 CV of GST wash buffer, followed by 2 CV GST on-column cleavage buffer. $MgCl_2$ and ATP were included within the GST wash buffer to eliminate co-purifications such as Hsp70 chaperones [198]. To elute the desired protein, the GST-tag was cleaved on column, using 1 mg of recombinant TEV protease per 10 mg of recombinant protein. The reaction was performed over night in 4 CV GST on-column cleavage buffer under constant circulation. The following day, cleaved recombinant protein was collected by applying 2 CV of GST wash buffer to the column and subsequently dialyzed for at least

2 h against Ni-NTA loading buffer. To remove His-tagged TEV protease, the dialyzed protein solution was subjected to Ni-NTA affinity chromatography where the desired protein was collected in flow-through fractions.

2.3.7 *Cation exchange chromatography*

RNA-binding proteins like Lin28 still contain nucleic acid contamination after initial affinity purification steps. To remove these contaminations, a cation exchange column was used. Presence of anionic sulfate groups in heparin columns mimic the polyanionic structure of nucleic acids and can thus be used for selective binding of positively charged RNA-binding proteins, while negatively charged nucleic acids do not bind to the column and are washed away.

Before application to a pre-equilibrated 5 ml HiTrap Heparin column (GE Healthcare, Piscataway, USA), the protein solution was dialyzed against heparin loading buffer. This buffer contained a low salt concentration (50 mM NaCl) and a pH (20 mM Bis-Tris, pH 6.5) below the desired protein's pI to ensure that it was positively charged and could bind to the column. Bound Lin28 was eluted with a linear gradient from 0.05 to 1 M NaCl in the same buffer (heparin elution buffer). The elution profile was monitored by measuring the absorption at 260 and 280 nm.

2.3.8 *Size-exclusion chromatography (SEC)*

To remove aggregates and higher oligomeric species, eluted fractions from cation exchange chromatography (see Section 2.3.7) were concentrated and subsequently injected on pre-equilibrated Superdex columns (GE Healthcare, Piscataway, USA) at a flow rate between 0.5 and 2 ml/min. Depending on the size of the proteins, either Superdex 75 (optimal separation range 3-70 kDa) or Superdex 200 columns (optimal separation range 10-600 kDa) were used. Small sample volumes (less than 4 ml) were applied on 16/60 columns, while sample volumes up to 10 ml were applied to Superdex 26/60 columns (diameter in mm/length in cm). Peak fractions were pooled and concentrated to approximately 10 mg/ml.

2.3.9 *Concentration of protein solutions*

Protein solutions were concentrated using Amicon Centrifugal Filter Devices (Merck Millipore, Billerica, USA) with a molecular weight cut off of 3, 10 or 30 kDa according to the manufacturer's protocol.

2.3.10 *Protein storage*

For long-term storage, purified protein was divided into 50, 100 or 200 µl aliquots and flash-frozen with liquid N_2. Aliquots were stored at -80 °C.

2.4 BIOCHEMICAL AND BIOPHYSICAL METHODS

2.4.1 *Protein concentration determination*

To exclude errors derived from minor RNA contamination, protein concentrations of RNA-binding proteins were determined using a modified Bradford and Lowry assay, according to the manufacturer' protocol (Bio-Rad Protein Assay, Hercules, USA; assay adapted from [199, 200]). The concentrations of all other proteins were determined according to Beer's law (Equation 1) by their absorbance at 280 nm (A_{280}) and their molar extinction coefficient ϵ_{280} calculated with ProtParam [201].

$$c_{protein} = \frac{A_{280}}{\epsilon_{280} \cdot d} \tag{1}$$

$c_{protein}$	protein concentration (M)
ϵ_{280}	molar extinction coefficient at 280 nm in ($M^{-1}cm^{-1}$)
d	path length in (cm)

2.4.2 *Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)*

The presence and purity of proteins were determined by discontinuous SDS-PAGE. Depending on the size of the protein between 12 - 19% SDS-polyacrylamide gels were cast according to standard procedures [196]. Prior to SDS-PAGE, samples were mixed with SDS sample buffer and boiled for 5 min before loading the gel. Electrophoresis were conducted with 1x SDS running buffer and developed at 240 V for 35 min. To estimate the size of protein bands, Precision Plus (Bio-Rad, Hercules, USA) or Mark 12 (Life Technologies, Carlsbad, USA) molecular weight standards were used. Gels were stained with Coomassie brilliant blue solution R-250 for 20 min followed by destaining for 1 h or overnight in destaining solution.

2.4.3 *Right-angle light scattering (RALS)*

In this method, an analytic gel filtration column is combined with a coupled RALS system and refractive index (RI) detector. This technique allows calculation of the absolute molecular mass and oligomeric state of a protein at any point in the chromatogram. For each protein sample, 100 µl of a 3 mg/ml solution was injected on a pre-equilibrated Superdex 75 or 200 10/300 GL column at a flow rate of 0.5 ml/min in SEC buffer I, II or III, respectively.

2.4.4 *Circular dichroism (CD) spectroscopy*

To estimate the secondary structure and folding of different protein samples, each protein of interest was dialyzed against CD buffer overnight and diluted to 0.1 mg/ml. For each sample at least three CD spectra ranging from 260 nm

to 190 nm at a resolution of 0.5 nm were recorded and averaged. Each measurement was performed in a 0.5 or 1 mm quartz cuvette (Hellma, Mühlheim, D) at 20 °C on a Chirascan spectropolarimeter (Applied Photophysics, London, UK). After subtraction of the buffer spectrum from each protein spectrum, the measured ellipticity Θ_{meas} (in mdeg) was recalculated in mean molar ellipticity per amino acid (Θ_{MRW}) according to [202] (Equation 2). Secondary-structure estimations were performed using the CDSSTR algorithm [203, 204, 205] and protein reference set 7 within Dichroweb [206].

$$\Theta_{MRW} = \frac{\Theta_{meas} \cdot M_p}{10 \cdot c \cdot d \cdot N} \tag{2}$$

Θ_{MRW}	mean molar ellipticity per amino acid ($deg \cdot cm/dmol$)
Θ_{meas}	measured ellipticity (mdeg)
M_p	molar mass of protein (g/mol)
N	number of amino acids in protein
$c_{protein}$	protein concentration (mg/ml)
d	path length (cm)

2.4.5 *Isothermal titration calorimetry (ITC)*

ITC experiments were conducted at 15 °C in ITC buffer on a VP-ITC MicroCalorimeter with an active cell volume of 1.4 ml. Prior to the experiment, all proteins or oligonucleotides were dialyzed against ITC buffer. Depending on the interaction partners, protein concentrations between 5 to 50 µM in the cell were titrated with at least a tenfold molar excess of protein or DNA/RNA ligand in the syringe. After an initial temperature equilibration, injections of either 8 µl or 10 µl were repeated every 270 s. All binding isotherms and thermodynamic parameters were fit and calculated using the MicroCal ORIGIN software (OriginLab Corporation, Northampton, USA).

2.4.6 *Fluorescence quencher assay*

The intrinsic fluorescence of *Xtr*Lin28B CSD Trp39 is quenched upon binding to DNA and RNA oligonucleotides. Therefore it was used as a probe to examine the nucleic acid binding specificities of this protein domain. 50 nM *Xtr*Lin28B CSD were mixed with increasing concentrations of ligand in a total volume of 1 ml in ITC buffer supplemented with 40 µM N-acytyl tryptophanamide to avoid unspecific association of the protein to the quartz surface. At each step 10 µl ligand from a concentrated stock solution were added and equilibrated for 1 min under constant stirring at 20 °C. The fluorescence of Trp39 was subsequently measured at 350 nm upon excitation at 280 nm and corrected for inner-filter effects, buffer, oligonucleotide fluorescence and dilution. The total increase in volume after the completed titration was less than 10%. The changes

of fluorescence were plotted against the concentration of oligonucleotide and analyzed according to the following binding equation [207, 208] (Equation 3).

$$Q = Q_{max} \frac{A - \sqrt{A^2 - 4 \cdot n \cdot [P_0] \cdot [L_0]}}{2 \cdot [L_0]} \tag{3}$$

A	$A = K_D + [P_0] + [L_0]$
Q	quenching of the intrinsic Trp39 fluorescence at each step
n	stoichiometric ratio of protein•ligand complex
K_D	equilibrium dissociation constant (M)
$[P_0]$	concentration of *Xtr*Lin28B CSD (M)
$[L_0]$	concentration of oligonucleotide ligand (M))

2.4.7 *Co-immunoprecipitation*

Plasmids containing V5-tagged Lin28 and Flag-tagged TUT4 were transiently expressed in HEK293 for 48 h (Section 2.7.2). After cell lysis using HEK293 cell lysis buffer and a 23 gauge needle, the cell extract was incubated with 10 µl of anti-Flag antibody-conjugated agarose beads (anti-Flag M2 affinity gel, Sigma-Aldrich, St. Louis, USA) under constant rotation for 2 h at 4 °C. Bound protein complexes were subsequently eluted using Flag-peptide (Sigma) and subjected to SDS-PAGE and Western blot analyses. To check, whether the interaction depends on the presence of pre-*let-7* miRNA, 1 µg of *in vitro* transcribed pre-*let-7* miRNA was added to individual HEK293 cell extracts after centrifugation.

2.4.8 *Western blot*

After SDS-PAGE, proteins were transferred from an unstained polyacrylamide gel onto a polyvinylidene fluoride (PVDF) membrane (GE Healthcare, Piscataway, USA) as previously described [209]. Prior to usage, PVDF membranes were activated by soaking for 1 min in 100% methanol. Then, blotting buffer was applied onto gel and membrane and the transfer was conducted at 60 mA for 50 min in a semi-dry blotting apparatus.

2.4.9 *Immunodetection*

Detection of proteins immobilized on PVDF membranes was performed using specific antibodies against the epitope tags or against the protein itself. After blocking the membrane with blocking buffer, primary antibody was diluted in PBS-T (anti-FLAG: 1:6000, anti-V5: 1:5000) and incubated at room temperature for 2 h or overnight. The membrane was washed three times for 15 min with 20 ml PBS-T. Then, the secondary antibody, which is coupled to horseradish peroxidase (HRP), was diluted 1:10000 in PBS-T and incubated with the membrane under gentle agitation for 1 h. The membrane was washed three times for 10 min before visualized as described in the Enhanced Chemiluminscence (ECL) Plus Western Blotting Detection Kit (GE Healthcare, Piscataway, USA).

For the detection of His-tagged proteins, an anti His-HRP conjugate (Qiagen, Hilden, D) was used according to the manufacturer's protocol.

2.5 RNA BIOCHEMISTRY

2.5.1 *In vitro transcription*

For *in vitro* transcription of pre-*let-7* miRNAs, the T7 Transcription Kit (Roboklon, Berlin, D) was used according to the manufacturer's instructions. All DNA templates were fused to a T7 RNA polymerase promoter (TAATACGACTCACTATAG) prior to cloning into pENTR4. After PCR amplification, 1 to 2 µg linear DNA template was added to *in vitro* transcription buffer supplemented with 5 U of heat stable pyrophosphatase and 50 U of ribonuclease inhibitor (both from Roboklon, Berlin, D) in a total volume of 100 µl. The reaction was started by addition of 500 U of T7 RNA polymerase and incubated for 4 h at 37 °C. After stopping the reaction by heating at 95 °C for 5 min, the RNA was purified using MicroSpin G-50 columns (GE Healtchare, Piscataway, USA) and checked for purity by denaturing urea PAGE (see Section 2.5.5).

For large-scale *in vitro* transcription, the small-scale reaction conditions were up-scaled to 10 ml and conducted overnight using 0.2 mg/ml 6xHis-tagged T7 RNA polymerase. RNA was concentrated, purified using a Superose 12/300 GL column (GE Healthcare, Piscataway, USA) and checked for purity by denaturing urea PAGE.

2.5.2 *DNA and RNA concentration determination*

Concentrations of ssDNA and ssRNA oligonucleotides were determined by measuring the absorbance at 260 nM with a Nanodrop spectrophotometer according to Beer´s law (see equation Equation 1). Molar extinction coefficients for each oligonucleotide were calculated using OligoCalc [210].

2.5.3 *RNA labeling*

For electrophoretic mobility shift assays (EMSA) RNA radiolabeling was performed during *in vitro* transcription using 10 µCi [α-^{32}P] ATP (Perkin Elmer, Waltham, USA) per reaction (see Section 2.5.1). Otherwise, the 5' end of RNA was labeled in the presence of [γ-^{32}P] ATP and polynucleotide kinase (PNK) using the Ambion KinaseMax 5' End Labeling Kit (Life Technologies, Carlsbad, USA) according to the manufacturer's protocol. Prior to the labeling reaction, RNA 5'-phosphate was removed by incubating with calf intestine phosphatase (NEB, Ipswich, USA) for 30 min at 37 °C. Finally, radiolabeled RNA was purified using MicroSpin G-50 columns (GE Healthcare, Piscataway, USA).

2.5.4 *Electrophoretic mobility shift assay (EMSA)*

Binding reactions were conducted using the indicated Lin28 concentrations and 1 nM [α-^{32}P] ATP labeled pre-*let-7* miRNA as a probe in a total volume of

20 µl in EMSA binding buffer. To ensure optimal hairpin formation, the used miRNAs were heated at 95 °C for 2 min followed by snap cooling on ice for 3 min. For each EMSA experiment, up to 10 protein samples were diluted with EMSA binding buffer and mixed with the appropriate RNA. After 30 min of incubation at room temperature, EMSA samples were mixed in a 1:1 ratio with EMSA loading buffer and resolved on a native 10% (w/v) polyacrylamide gel. The gel was run at 200 V until the bromo phenolblue marker reached the bottom of the gel. RNA was visualized by phosphor imaging, and band intensities of scanned gels were quantified using ImageQuant (GE Healthcare, Piscataway, USA). The total amount of bound RNA in each binding reaction was normalized against the unbound RNA in the absence of recombinant protein. Data were fit to a one-site specific binding function with Hill slope, taking a possible cooperative binding into account. The following binding equation of the nonlinear curve fitting method of GraphPad Prism (version 5.0a for Mac, GraphPad Software, La Jolla, USA) was used (Equation 4).

$$B = \frac{B_{max} \cdot [P]^h}{K_D + [P]^h} \tag{4}$$

B	ratio of protein•RNA complex versus the total amount of RNA as a function of protein concentration (%)
B_{max}	total amount of RNA/maximal number of binding sites (%)
$[P]$	protein concentration (M)
h	Hill coefficient
K_D	the dissociation constant (M)

2.5.5 *Denaturing urea polyacrylamide gel electrophoresis*

To analyze the size of ssRNA products form *in vitro* pre-miRNA processing and uridylation assays (Section 2.5.6 and 2.5.9), denaturing urea PAGE was used. 31x38.5 cm gels were prepared with 10% (w/v) TBE acrylamide solution supplemented with 8 M urea according to standard procedures [196]. After pre-running the gel in 1x TBE buffer for 45 min, samples were diluted with 2x TBE urea buffer and heated at 95 °C for 5 min before loading the gel. Electrophoresis was conducted in 1x TBE urea buffer at 1500 V and maximal 40 W in a Model S2 Sequencing Apparatus system (Analytik Jena, Jena, D) for 1.5-2 h. RNA was visualized by phosphor imaging, and band intensities of scanned gels were quantified using ImageQuant (GE Healthcare, Piscataway, USA).

2.5.6 *In vitro pre-miRNA processing assay*

10000 cpm of [γ-^{32}P] 5'-end labeled pre-miRNA (Section 2.5.3) was incubated with different concentrations of Lin28 variants for 30 min on ice in Dicer reaction buffer. Then, 2 U of recombinant human Dicer (Genlantis, San Diego, USA) were added to each sample and incubated for 2 h at 37 °C. The reaction was stopped by mixing 4 µl Dicer stop solution (from recombinant human Dicer enzyme kit, Genlantis San Diego, USA) with 20 µl sample and heated at 95 °C for

5 min. All samples were resolved by 10% (w/v) TBE-urea PAGE and visualized by autoradiography (Section 2.5.5).

2.5.7 *RNA remodeling assay*

For determination of the RNA remodeling activity of Lin28, a truncated *Xtr*pre-*let-7g** was synthesized by Dharmacon (Life Technologies, Carlsbad, USA). This construct contained a 5'-dabcyl and a 3'-fluorescein label that are in close proximity to each other according to secondary structure prediction using Mfold [211]. The remodeling reaction was started with 200 nM *Xtr*pre-*let-7g** in a total volume of 1 ml at 20 °C before increasing amounts of Lin28 variants were added stepwise. At each step 100 µl of the appropriate stock solution were added and equilibrated for 1 min under constant stirring before the fluorescence of fluorescein was measured at 522 nm upon excitation at 495 nm. All experiments were performed in RNA remodeling buffer using a Cary-Eclipse fluorescence spectrofluorimeter. The fluorescence was corrected for inner filter effects, buffer and dilution.

2.5.8 *Kinetic measurements of RNA remodeling*

The kinetics of Lin28 mediated *Xtr*pre-*let-7g** remodeling were measured using rapid mixing of 30 µM Lin28 variant and 200 nM *Xtr*pre-*let-7g** at a 1:1 ratio. Prior to the experiment, protein and RNA solutions were dialyzed against RNA remodeling buffer and applied to two separate 2 ml syringes. Remodeling kinetics were followed by total fluorescence above 515 nm upon excitation at 490 nm using a 5 nm bandwidth. All experiments were performed at 20 °C in RNA remodeling buffer on a Chirascan SF.3 spectrometer (Applied Photophysics, UK, London). The path length of the observation chamber was 10 mm. Kinetic curves were measured at least 12 times under identical conditions, averaged and fit either to a mono- (Equation 5) or bi-exponential equation (Equation 6).

$$F(t) = A \cdot e^{-k \cdot t} + C \tag{5}$$

$$F(t) = A_{fast} \cdot e^{-k_{fast} \cdot t} + A_{slow} \cdot e^{-k_{slow} \cdot t} + C \tag{6}$$

$F(t)$	fluorescence intensity at time t
A, A_{fast}, A_{slow}	fluorescence amplitude
k, k_{fast}, k_{slow}	rate constants (s^{-1})
$\tau, \tau_{fast}, \tau_{slow}$	time constants, reciprocal values of rate constants (s)
C	constant representing the fluorescence at time zero

2.5.9 *In vitro uridylation assay*

PUP2 mediated 3'-uridylation of pre-*let-7* miRNA was assayed using [α-^{32}P] ATP and non-labeled *in vitro* transcribed pre-*let-7* miRNA. For the reaction,

10 µM purified PUP2 was added to 100 nM pre-*let*-7 in the presence and absence of 10 µM Lin28. The reaction was conducted for 1 h at 37 °C in a total volume of 20 µl containing *in vitro* uridylation buffer supplemented with ~100 nM [α-^{32}P] ATP and 10 U ribonuclease inhibitor (Roboklon, Berlin, D). All samples were resolved by 10% (w/v) TBE-urea PAGE and visualized by autoradiography (Section 2.5.5).

2.6 PROTEIN CRYSTALLIZATION AND STRUCTURE DETERMINATION

All structures of proteins and protein•nucleic acid complexes were determined by x-ray crystallography. In the following sections practical aspects of crystallization, data collection, molecular replacement, atomic model building, refinement and structure validation are presented. A more detailed overview about the theoretic background is given in Appendix B.

2.6.1 *Protein crystallization*

All crystallization screens were performed using the vapor-diffusion method by mixing equal volumes of protein and reservoir buffer at either 4 °C or 20 °C. Initial crystallization screens were set up in 96-well plates using a semi-automatic dispensing system [212] with the following precipitation screens: JCSG, Classic, Classic II, Classic Lite, pH-Clear, pH-Clear II, PEG, PEG II, PEG-Ion, PACT, MPD, Anions, Cations, ProteinComplex, Nucleix (Qiagen, Hilden, D). In these setups typically 200-400 nl protein solution with concentrations between 5 and 20 mg/ml were mixed at a 1:1 ratio with precipitant solution. The reservoir volume was 75 µl. All initial crystallization setups were regularly photographed and evaluated using the program OBSERVATION (developed by M. Hoeschen). Initial hits were reproduced and optimized using hanging-drop 24-well plate setups. In these trials, the previously observed crystallization conditions were systematically changed by varying the concentration of protein and precipitant as well as the pH. For each well 1 µl protein solution was mixed with an equal volume of reservoir solution on a siliconized cover slip. The cover slip was then attached on a pre-greased 24 well crystallization plate whose reservoirs were filled with 500 µl precipitant solution.

Single crystals of hLin28B CSD (approximately 120 µm x 80 µm x 80 µm) were obtained at 4 °C by mixing 11 mg/ml protein in 20 mM Tris/HCl pH 8.0, 340 mM NaCl, 5 mM DTT with 2 M $(NH_4)_2SO_4$, 0.2 M NaCl and 0.1 M MES pH 6.5 as reservoir solution.

*Xtr*Lin28B CSD was crystallized at 20 °C with 10 mg/ml protein in 20 mM Tris/HCl pH 8.0, 360 mM NaCl, 5 mM DTT and 2.5 M sodium acetate, 0.1 M HEPES pH 7.0 as reservoir solution. Hexagonal crystals appeared after 2 d and finally grew to a size of approximately 120 µm x 80 µm x 80 µm. For freezing, single crystals were soaked in a cryoprotectant consisting of reservoir solution supplemented with 25-30% (v/v) glycerol and subsequently flash-frozen in liquid nitrogen.

2.6.2 *Crystallization of protein•nucleic acid complexes*

For co-crystallization of protein•nucleic acid complexes, purified *Xtr*Lin28B CSD was mixed with a 1.2-fold molar excess of the corresponding oligonucleotide and applied on a Superdex 75 16/60 or Superose 12 10/300 GL column (both GE Healthcare, Piscataway, USA) to separate the complex from unbound oligonucleotides. Peak fractions were pooled and concentrated to 5 or 10 mg/ml protein and immediately used for crystallization trials.

*Xtr*Lin28B CSD•dT_6 was crystallized at 4 °C using 0.75 mM complex in 20 mM Tris/HCl pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 1 mM DTT and a reservoir solution comprising 0.1 mM Bis-Tris pH 5.5, 15% (w/v) PEG 3350, 0.1 M sodium thiocyanate. Crystals appeared after 2 d and grew to a final size of approximately 250 µm x 60 µm x 20 µm.

*Xtr*Lin28B CSD•dT_7 was crystallized at 4 °C using 1 mM complex in 20 mM Tris pH 7.5, 100 mM KCl, 1 mM DTT and a reservoir solution comprising 17% (w/v) PEG 3350, 0.2 M sodium thiocyanate. The crystals were similar in size and shape as those of *Xtr*Lin28B CSD•dT_6.

Crystals of *Xtr*Lin28B CSD•rU_6 were obtained at 4 °C using 1 mM complex in 20 mM Tris pH 7.5, 100 mM KCl, 1 mM DTT and a reservoir solution comprising 20% (w/v) PEG 3350, 0.2 M potassium thiocyanate. The crystals grew to a final size of 250 µm x 100 µm x 80 µm within 7 d.

2.6.3 *Data collection*

X-ray diffraction data of single crystals were collected using the oscillation method with a ϕ increment of 1° under cryo conditions at BESSY II (Berlin, D) [213]. All datasets were recorded at a wavelength of 0.9184 Å on beamline BL 14.1 (equipped with a Rayonics MX-225 CCD detector) or BL 14.2 (equipped with MAR165 CCD detector). Initial indexing and determination of an optimal data collection strategy was performed using iMOSFLM [214]. The recorded diffraction data for *Xtr*Lin28B CSD crystals were indexed, integrated and scaled with XDS [215], or with HKL2000 (hLin28B) [216].

2.6.4 *Molecular replacement*

For *Xtr*Lin28B CSD the phase problem was solved by molecular replacement using Phaser [217] and *Salmonella typhimurium* cold shock protein (3I2Z) as a template. The structure of hLin28B CSD and the *Xtr*Lin28B CSD•nucleic acid complexes were solved similarly but using the structure of apo *Xtr*Lin28B CSD as a template.

2.6.5 *Atomic model building, refinement and validation*

Initial atomic models of proteins were built into the electron density by the program ARP/wARP [218]. Atomic models for DNA and RNA oligonucleotides were placed manually into the electron density using COOT [219]. The initial models were subsequently refined in multiple rounds of TLS refinement

(translation, libration and screw-rotation, [220]) and restrained refinement with REFMAC 5.5 [221]. TLS groups were determined using the TLS Motion Determination (TLSMD) webserver [222]. Due to its high resolution, *Xtr*Lin28B CSD could be refined using anisotropic B-factors. The graphics program COOT was used for model building and structure analysis [219]. All atom contacts and geometry of the atomic models were evaluated using the Molprobity server [223, 224].

2.6.6 *Structure analysis and alignments*

Molecular drawings and root-mean-square deviation (RMSD) values were created or calculated with PyMOL Molecular Graphics System (Version 1.3, Schrödinger, LLC, Heidelberg, D). Hydrogen bonding was analyzed using the University of California at San Francisco Chimera program [225]. Amino acid, DNA and RNA sequences were aligned using CLC Genomics workbench 3.6.5 (CLC bio, Katrinebjerg, Denmark).

2.7 CELL BIOLOGICAL METHODS

2.7.1 *Cell culture*

HEK293 cells were cultivated in DMEM high glucose medium (Gibco/Life Technologies, Carlsbad, USA) supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 100 U/ml penicillin and 0.1 mg/ml streptomycin. Cultures were constantly maintained at 37 °C and 5% CO_2 and split every 2-3 d. For splitting, medium was removed and cells were washed with PBS. Cells were detached from the surface by incubating with 0.05% Trypsin/EDTA (Gibco/Life Technologies, Carlsbad, USA) for 2 min and then resuspended in the appropriate medium. To remove traces of Trypsin/EDTA, cells were centrifuged for 5 min at 500 g and resuspended in the corresponding medium. New culture plate were seeded at ratios from 1:3 to 1:5.

2.7.2 *Transient expression in eukaryotic cells*

HEK293 cells were co-transfected with Flag-TUT4 and V5-Lin28 encoding plasmids pcDNA3.1/nFlag-DEST-TUT4 and pcDNA3.1/nV5-DEST-Lin28. For transfection, 2×10^6 cells were seeded per 10 cm culture vessel 48 h prior to the experiment. Cells were transfected using 10 µg DNA and DNAfectin 2100 (ABM, Crestwood Place, Canada) according to the manufacturer's protocol. After 48 h, cells were collected in HEK293 lysis buffer and passed through a 23 gauge needle at least 15 times. Cell debris was removed by centrifugation and the supernatant used for co-immunoprecipitation assays (Section 2.4.7).

2.7.3 *Luminescence-based mammalian interactome assay*

Luminescence-based mammalian interactome assays (LUMIER) were performed as recently described [226] with minor modifications. The coding sequence of

the suspected interaction partners human/*C. elegans* Lin28 and TUT4/PUP2 were cloned into pENTR4 and subsequently transferred into pDEST-*renilla* and pDEST-*firefly* via LR recombination (Section 2.2.10). After co-transfection of the plasmids into HEK293 cells, Protein A (PA)-*renilla* luciferase (RL)-tagged Lin28 and V5-*firefly* luciferase (FL)-tagged TUT4/PUP2 proteins were transiently expressed for 48 h. Then cells were harvested and lysed in HEK293 lysis buffer supplemented with Complete protease inhibitor cocktail (Roche, Penzberg, D), RNase-free DNase I (Qiagen, Hilden, D) and RiboLock RNase inhibitor (Thermo Fisher Scientific, Vilnius, Lithuania) to avoid pre-*let-7* degradation. Expression levels of the fusion proteins were analyzed by measuring luciferase activity in the lysate. Potential protein complexes were co-immunoprecipitated using IgG (Jackson Immunoresearch, Suffolk, UK), immobilized to high binding 96-well plates (Greiner, Kremsmuenster, Austria). After washing the bound complexes with PBS, binding of V5-FL-tagged TUT4/PUP2 to PA-RL-tagged Lin28 was analyzed using the *firefly* luciferase activity in a TECAN Infinite M200 luminescence plate reader (Tecan, Männedorf, CH). Simultaneously, *renilla* luciferase activity was measured as a control. All luciferase activity measurements were performed using the Dual Luciferase Assay System (Promega, Fitchburg, USA) according to the manufacturer's protocol. Each measurement was repeated at least three times.

To immobilize IgG to high binding 96-well plates, each well was incubated with 100 µl coating solution containing sheep gamma globulin (Jackson Immunoresearch, Suffolk, UK) for 3 h under constant agitation. After removing coating solution, 300 µl TP solution were added to each well and incubated for 1 h. Wells were washed three times with 300 µl TBS-T and subsequently coated with 100 µl capture solution containing anti-sheep IgG. Prior to usage, plates were washed for another three times with TBS-T and once with HEK293 lysis buffer.

3 RESULTS

3.1 BIOCHEMICAL CHARACTERIZATION OF LIN28

3.1.1 *Screening for soluble protein*

To obtain structural and functional data of Lin28, cDNA from *C. elegans*, human, mouse and *Xenopus tropicalis (Xtr)* Lin28 were purchased, PCR-amplified and cloned into the bacterial expression vectors pQLinkH or pDEST15. Furthermore, various truncated constructs were designed lacking unstructured regions that are detrimental to crystallization. Typically all constructs harbored at least one of the two RBDs. After sequence verification, plasmids were transformed into *E. coli* Rosetta 2 DE(3) and each gene was expressed in small-scale format as described in Section 2.3.1 to assess protein solubility. Most of the constructs displayed good overexpression and solubility at 17 °C and 20 °C and could be enriched using Ni-NTA Agarose or GSH beads. Only some hexahistidine-tagged mLin28B and Lin28B constructs were not soluble under these conditions. A summary about the numbers of cloned, soluble and purified constructs is given in Table 2.

Table 2: Summary of cloned, recombinantly overexpressed, soluble and purified constructs.

GENE	CLONED	SOLUBLE	PURIFIED
*Cel*Lin28	24	21	15
hLin28A	15	12	9
hLin28B	13	9	8
mLin28A	5	3	2
mLin28B	4	0	0
*Xtr*Lin28A	9	7	5
*Xtr*Lin28B	35	33	32

3.1.2 *Protein purification*

Soluble constructs were overexpressed in large-scale format (see Section 2.3.2) and purified within three days in a five step process (see Section 2.3). After cell lysis, His-tagged proteins were purified by Zn/Ni-NTA affinity chromatography and incubated with TEV protease overnight to remove the N-terminal tag from the protein. Following cleavage and dialysis, cleaved Lin28 protein was separated from uncut protein and cleaved tag by a second Zn/Ni-NTA affinity chromatography. Most purified Lin28 constructs still contained DNA or RNA

contamination at this step as judged by their A_{260}/A_{280} ratio. To remove these, the protein solution was dialyzed against a low salt buffer (typically 50 mM to 100 mM NaCl or KCl) with a slightly acidic pH value (6.5) before application to a heparin column. Under these conditions, all Lin28 constructs (pIs between 7.5 and 9.5) are positively charged and consequently bound to the negatively charged column. A further decrease of pH led to a visible precipitation of protein, probably due to protonation of the chelating Zn-knuckle's His, and loss of the central Zn^{2+} ion. Elution of pure Lin28 protein was subsequently achieved by increasing the salt concentration. A final size-exclusion chromatography step removed remaining aggregates and contaminants to obtain almost homogeneous protein. As an example, the purification of *Xtr*Lin28B CSD$^{27\text{-}114}$ is shown in Figure 13.

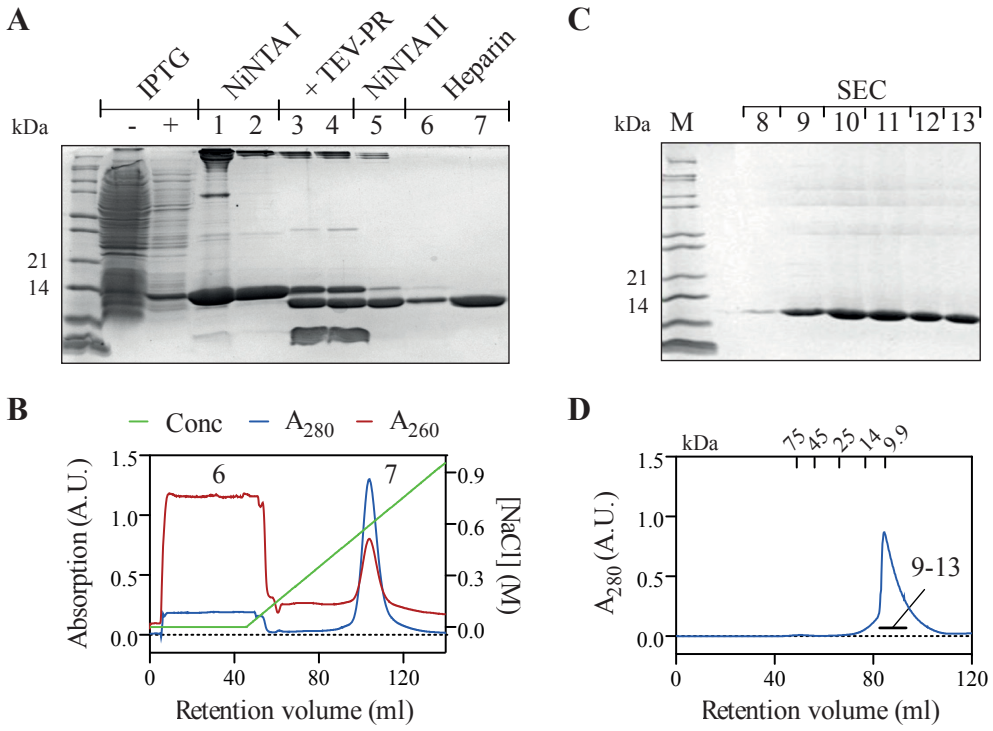


Figure 13: Typical Lin28 purification procedure. (**A**) 15% SDS-PAGE of various samples taken during the purification of *Xtr*Lin28B$^{27-114}$CSD. -/+ IPTG: whole-cell bacterial lysates before and after induction; 1, 2: imidazole eluate containing His-tagged *Xtr*Lin28B$^{27-114}$CSD; 3, 4: imidazole eluate after 4 h or 16 h incubation with TEV protease; 5: flow-through of Ni-NTA sepharose column in the presence of 10 mM imidazole; 6: flow-through of heparin cation exchange chromatography; 7: elution fraction of heparin column. (**B**) Heparin cation exchange chromatogram. Flow-through fractions mainly contained RNA- and DNA-contaminations with $A_{260}/A_{280} > 2$ while in the elution fraction hardly any RNA- and DNA-contaminations were detectable ($A_{260}/A_{280} < 0.7$). (**C**) 15% SDS-PAGE of samples from peak fractions of an S75 size-exclusion run with purified *Xtr*Lin28B$^{27-114}$CSD. (**D**) Superdex 75 size-exclusion chromatogram. The indicated fractions were pooled and concentrated for further experiments.

All Lin28 constructs were purified in the same way. For both *Xenopus tropicalis* paralogs *Xtr*Lin28A and *Xtr*Lin28B as well as their isolated domains, high

yields could be obtained (between 3 mg to 10 mg protein per l medium). Hence, most of the following biochemical experiments were conducted with these constructs. Moreover, in preparative SEC runs, all proteins eluted as single peaks at retention volumes corresponding to their molecular mass (Figure 14).

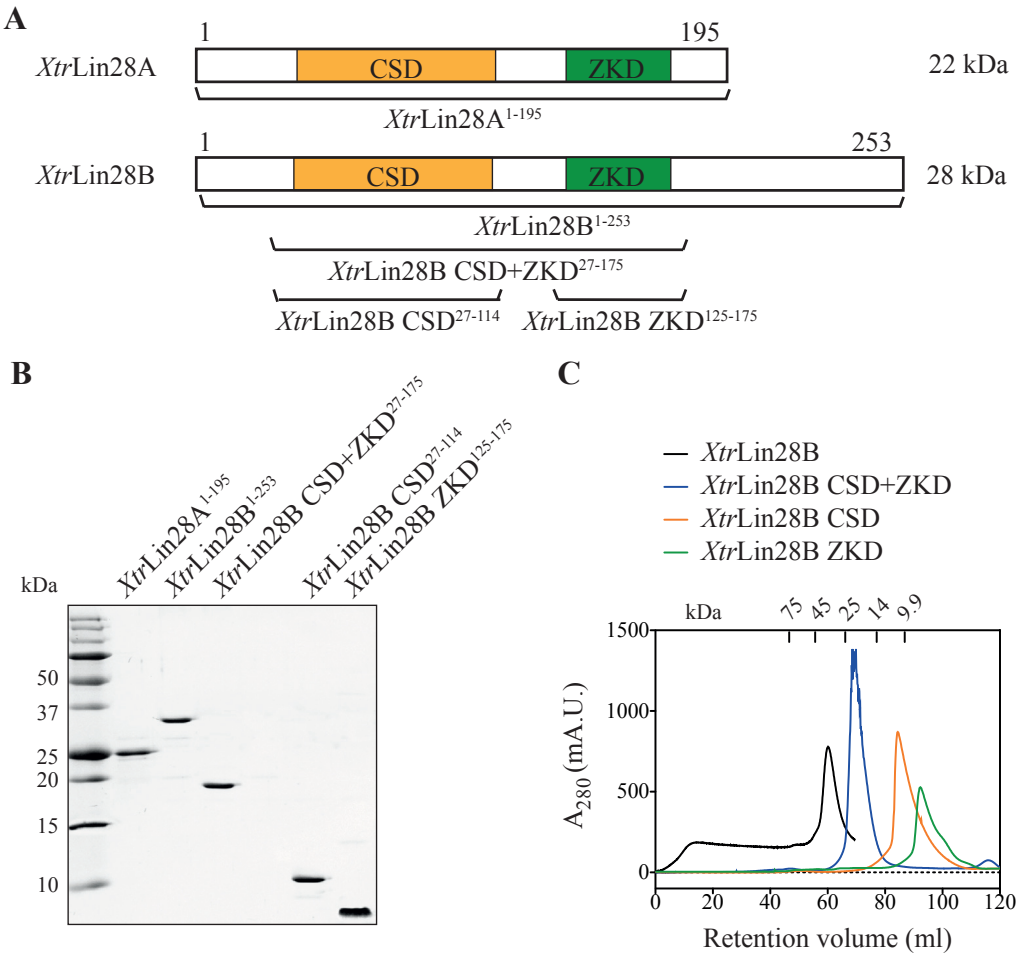


Figure 14: Preparative size-exclusion chromatograms of *Xtr*Lin28 and its isolated domains. (**A**) Domain structure of *Xtr*Lin28A and *Xtr*Lin28B with the molecular weights and amino acids positions indicated. The truncated Lin28B constructs that were predominantly used in this study are indicated below the corresponding domain representation. (**B**) 19% SDS-PAGE of purified *Xtr*Lin28 proteins. (**C**) Superdex 75 SEC of *Xtr*Lin28B, *Xtr*Lin28B CSD+ZKD, *Xtr*Lin28B CSD, *Xtr*Lin28B ZKD.

3.1.3 *Determination of secondary-structure content and folding state*

To demonstrate the folding state and secondary-structure content of WT and truncated *Xtr*Lin28B constructs CD spectra were recorded as described in Section 2.4.4. Only *Xtr*Lin28B CSD exhibited a spectrum typical for a well folded protein that resembles those of β-sheet proteins with a maximum at 197 nm (Figure 15 A, Table 3). This is consistent with previous observations of CSD-containing proteins, which are typically composed of five antiparallel β-sheets [176].

Table 3: Secondary-structure analysis of *Xtr*Lin28B circular dichroism data.

Construct	α-helix (%)	β-sheet (%)	β-turn (%)	random coil (%)	sum (%)
*Xtr*Lin28B WT	4	25	15	54	98
*Xtr*Lin28B CSD+ZKD	4	31	20	44	99
*Xtr*Lin28B CSD	6	41	22	30	99
*Xtr*Lin28B ZKD	3	7	17	73	100

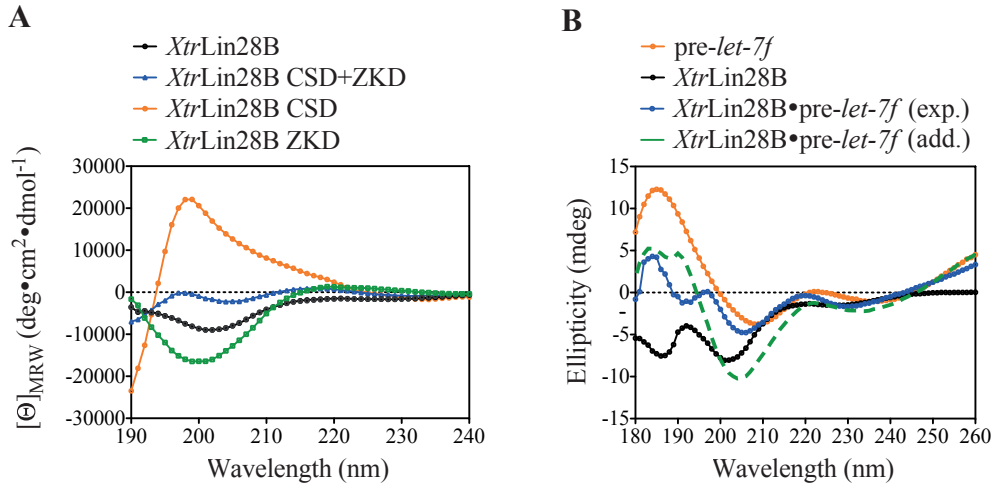


Figure 15: Circular dichroism spectroscopy of Lin28 proteins. (**A**) CD spectra of *Xtr*Lin28B, *Xtr*Lin28B CSD+ZKD, *Xtr*Lin28B CSD and *Xtr*Lin28B ZKD. All spectra were recorded at 20 °C in 0.1 cm pathlength cuvettes using 0.1 mg/ml protein in 10 mM Na-phosphate buffer (pH 7.4), 100 mM NaF. (**B**) Superposition of the CD spectra of 6 µM *Xtr*Lin28B•pre-*let-7f* complex, pre-*let-7f* and *Xtr*Lin28B, respectively. The difference between the experimentally recorded complex spectra (exp.) and the additive spectra of *Xtr*Lin28B and pre-*let-7f* (add.) indicates a change in structure in either protein or RNA.

*Xtr*Lin28B ZKD on the other hand revealed minimal discrete secondary structure and consequently possessed a high random coil content. Previous NMR studies on apo Lin28 ZKD revealed that this domain is quite flexible and contains little defined secondary-structure elements in the absence of bound RNA (PDB 2CQF). The CD spectra of *Xtr*Lin28B WT and *Xtr*Lin28B CSD+ZKD represent a mixture of CSD- and ZKD-spectra, and these proteins are therefore only partially folded in the absence of RNA. The CD spectrum of *Xtr*Lin28B•pre-*let-7f* complex suggests a structural change in either protein or RNA upon binding, as it differs considerably from the additive spectra of *Xtr*Lin28B and pre-*let-7f* (Figure 15 B). Altogether, *Xtr*Lin28B seems to be a highly flexible protein with the CSD making up the only well defined domain with defined secondary structure in absence of RNA. Furthermore, miRNA binding was associated with a structural change in either protein or RNA.

3.1.4 *MicroRNA binding*

Previous studies demonstrated that Lin28 binds to pri- and pre-*let-7* miRNA and inhibits its biogenesis [157, 130, 128]. To quantify the binding affinity and gain information about stoichiometry and binding sites, ITC experiments with two different *in vitro* transcribed pre-*let-7f* miRNAs were performed (see Figure 16). *Xtr*Lin28B bound full-length pre-*let-7f* with an equilibrium dissociation constant (K_D) of 640 ± 80 nM. Interestingly, the stoichiometry ($n = 0.60 \pm 0.01$) indicates that up to two proteins can bind to one pre-*let-7f* molecule. Further ITC experiments with the corresponding Dicer cleavage products of pre-*let-7f* showed that the terminal loop or pre-element (preE) of pre-*let-7f* is sufficient for binding. These constructs showed a similar dissociation constant ($K_D = 860 \pm 160$ nM) compared to full-length pre-*let-7f*, while the stoichiometry indicated a 1:1 binding for preE-*let-7f* ($n = 0.87 \pm 0.09$). Thus, the ds 22 nt long stem of pre-*let-7f* might offer an additional binding site for Lin28. However, as the mature 22 nt *let-7f* duplex RNA itself showed no substantial binding to *Xtr*Lin28B (data not shown) we conclude that preE is essential for binding.

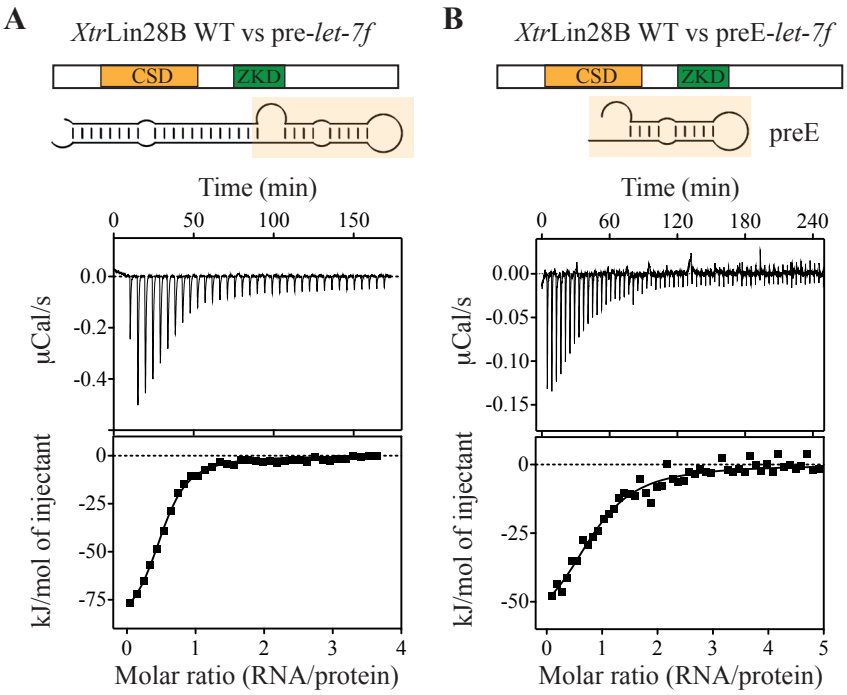


Figure 16: The pre-element (preE) of pre-*let-7f* is sufficient for Lin28 binding. (**A, B**) ITC binding experiments of *Xtr*Lin28B with *in vitro* transcribed pre-*let-7f* and preE-*let-7f*. The following values were obtained from fittings using MICROCAL ORIGIN. *Xtr*Lin28B•pre-*let-7f*: $K_D = 0.64 \pm 0.08$ µM, $n = 0.60 \pm 0.01$; *Xtr*Lin28B•preE-*let-7f*: $K_D = 0.86 \pm 0.16$ µM, $n = 0.87 \pm 0.09$.

3.1.5 *Determination of binding stoichiometry*

To further elucidate the binding stoichiometry of *Xtr*Lin28B•pre-*let-7f*, apo and pre-*let-7f* bound *Xtr*Lin28B were analyzed using right-angle light scattering

(RALS) in combination with analytical size-exclusion chromatography (see Section 2.4.3). In this assay apo *Xtr*Lin28B eluted predominantely as a monomer with a tendancy towards higher oligomeric species at the leading edge of the peak (Figure 17, Table 4). The *Xtr*Lin28B•pre-*let-7f* complex showed a 1:1 to 2:1 (protein:RNA) equilibrium and thus confirms the previously observed stoichiometry obtained by ITC experiments (see Section 3.1.4).

Taken together these data imply, that pre-*let-7f* offers up to two binding sites for *Xtr*Lin28B, with the pre-element (preE) being the major binding site. Furthermore, binding is associated with a structural change in protein and or RNA.

Table 4: Right-angle light scattering data

	*Xtr*Lin28B	*Xtr*Lin28B•pre-*let-7f*
Peak retention volume (ml)	10.70	10.06
Theoretical Mw (kDa)	28.1	50.1 (1:1 complex)
		78.2 (2:1 complex)
Mw (kDa)	36.1	65.9

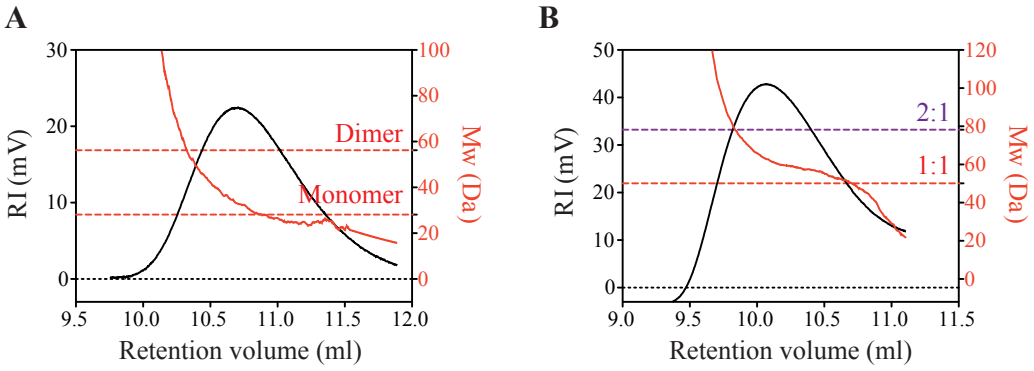


Figure 17: Right-angle light scattering analysis of (**A**) apo *Xtr*Lin28B and (**B**) pre-*let-7f* bound *Xtr*Lin28B. The refractive index (RI, black) and molecular weight (Mw, red) are plotted against the retention volume. Dashed lines indicate the theoretical molecular weights of monomeric/dimeric *Xtr*Lin28B or 1:1/2:1 *Xtr*Lin28B:*pre-let-7f* complexes, respectively.

3.2 RNA-BINDING SPECIFICITIES OF THE ISOLATED LIN28 DOMAINS

3.2.1 *The Lin28 ZKD specifically recognizes a conserved GGAG motif*

To obtain insights into nucleic acid binding specificities of Lin28, the isolated RNA-binding domains of Lin28 were biochemically examined along with the full-length protein. It was previously reported, that Lin28 variants with mutations in the ZKD were dramatically impaired in pre-*let*-7 binding [157, 132]. Further EMSA experiments and *in vitro* uridylation assays showed that a conserved GGAG motif within preE of pre-*let*-7 is critical for Lin28 binding and Lin28-mediated uridylation by TUT4 [132].

To confirm that the Lin28 ZKD mediates this binding preference, ITC experiments with the isolated ZKD and small ssRNA oligonucleotides were performed. *Xtr*Lin28B ZKD bound tightly to GGAG and GGUG-containing oligonucleotides with K_D values of 45 nM or 32 nM, respectively (Figure 18 A, Table 5). A single G-to-A mutation at the first or second position within the GGAG motif impaired binding only slightly (rZ3, rZ4), while mutating both Gs led to a more than 250-fold decrease in binding affinity (rZ5). A mutation of the fourth position strongly impaired binding of the ZKD (rZ6, rZ7). This is consistent with previously observed NMR structures of a homologous ZKD derived of the HIV-1 NC protein, in which each Zn-knuckle mediates a specific contact to the second and forth G of the GGAG/GGUG motif, respectively (50,51).

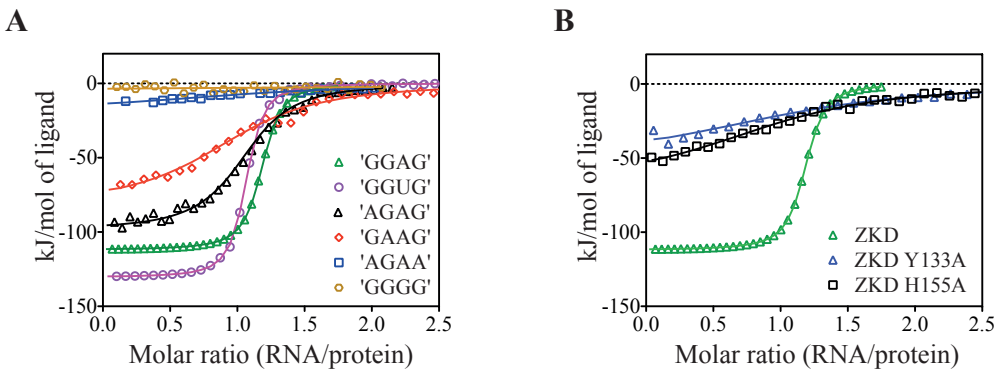


Figure 18: The Lin28 ZKD domain specifically recognizes GGAG or GGAG like motifs. (**A**) Integrated ITC data of *Xtr*Lin28B ZKD binding to ssRNA octamers. All used oligonucleotides contained two A's flanking the indicated motif. The highest K_D's were obtained for oligonucleotides harboring a GGAG or GGUG motif, while polyadenosines or polyuridines showed no binding. (**B**) Integrated ITC data of selected *Xtr*Lin28B ZKD variants binding to 5'-AA GGAG AA-3'. Y133A in the first Zn-knuckle and H155A in the second Zn-knuckle are important for recognizing the GGAG motif. Equilibrium dissociation constants and additionally calculated thermodynamic parameters for each binding reaction are listed in Table 5

.

On the protein level, key residues for a specific interaction within the GGAG motif are Tyr133 in the first and His155 in the second Zn-knuckle, because mutating both residues to A caused a dramatic loss of binding affinity (Figure 18 B). Both residues were shown to interact with each other and at least one G

of the GGAG motif, thereby establishing a kinked conformation in the RNA backbone [227, 228].

Remarkably, no binding was observable for polyadenosine or polyuridines (rA8, rU7). Furthermore, *Xtr*Lin28B ZKD could not bind to a GGGG-containing oligonucleotide, indicating that G itself cannot act as a spacing residue (rZ8). Taken together, the data show that two Gs spaced by at least one nucleotide other than G are specifically recognized by the Lin28 ZKD.

Table 5: Equilibrium dissociation constants of *Xtr*Lin28B ZKD binding to RNA oligonucleotides[a].

ZKD			
Ligand	Sequence	K_D (nM)	Stoichiometry
rZ1	AA GGAG AA	45 ± 4	1.17 ± 0.03
rZ2	AA GGUG AA	32 ± 2	1.04 ± 0.02
rZ3	AA AGAG AA	200 ± 30	1.06 ± 0.02
rZ4	AA GAAG AA	600 ± 170	1.06 ± 0.02
rZ5	AA AAAG AA	11900 ± 2200	0.97 ± 0.07
rZ6	AA GGAA AA	-	-
rZ7	AA AGAA AA	6800 ± 2100	0.83 ± 0.04
rZ8	AA GGGG AA	-	-
rA_8	AA AAAA AA	-	-
rU_7	UU UUUU U	-	-
ZKD variants			
Mutation	Sequence	K_D (nM)	Stoichiometry
Y133A	AA GGAG AA	2200 ± 700	1.14 ± 0.13
H155A	AA GGAG AA	2600 ± 600	1.03 ± 0.07

[a] All experiments were performed at 15 °C in 20 mM Tris/HCl, pH 8.0 and 60 mM KCl.
(-) not determined due to insufficient affinity.

3.2.2 *The Lin28 CSD preferentially binds to pyrimidine-rich sequences with high affinity but limited specificity*

3.2.2.1 *DNA binding*

Previous studies demonstrated that bacterial cold-shock proteins (Csps) and eukaryotic CSDs displayed a preference for pyrimidine-rich ssDNA/ssRNA oligonucleotides with K_D values in the low nanomolar range [180, 182, 229, 181]. To test whether *Xtr*Lin28B CSD has a similar sequence selectivity and affinity, the binding preferences of *Xtr*Lin28B CSD to ssDNA and ssRNA oligonucleotides were tested systematically using fluorescence titration experiments. In

this experiment the fluorescence of a single Trp (Trp39) is used as a probe to measure the binding constants.

First, the question was addressed how many nucleotides are bound by *Xtr*-Lin28B CSD. Therefore, oligothymidines in the range from five to nine nucleotides were used. The binding affinity constantly increased from 5- to 8-mers with dissociation constants ranging from 680 nM to 32 nM at a 1:1 stoichiometric ratio. No further substantial increase in binding strength could be obtained for longer ssDNA oligonucleotides, indicating that *Xtr*Lin28B CSD preferentially binds to hepta- or octanucleotides (Figure 19, Table 6).

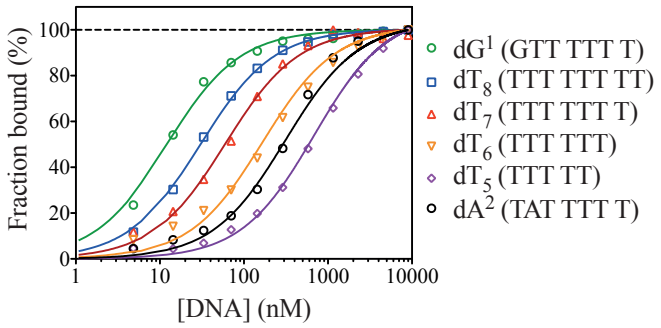


Figure 19: Equilibrium binding data of a subset of ssDNA oligonucleotides derived from fluorescence titration experiments with *Xtr*Lin28B CSD. All binding data were fit assuming a 1:1 stoichiometry. The resulted equilibrium dissociation constants of all binding reactions are listed in Table 6.

Next, the oligonucleotide sequence was altered progressively using heptathymidine (dT_7) as a reference. The binding analysis revealed that *Xtr*Lin28B CSD has the highest affinity towards oligonucleotides containing a G at the first position (dG^1, Table 6). This nucleotide exchange led to a fivefold decrease in the K_D. Furthermore, at positions 2 and 5 a T is preferred over A or C (dT_7 versus $dA^{2/5}$ and $dC^{2/5}$). Otherwise, mutations at positions 3, 4, 6 and 7 caused only subtle changes in the binding affinities with K_D values ranging from 39 nM to 147 nM, indicating that *Xtr*Lin28B CSD has only limited sequence specificity. However, as polypurine heptanucleotides could not bind to *Xtr*Lin28B CSD, the protein seems to have a slight preference for pyrimidine-rich ssDNA (dT_7 versus dG^7, dA^7).

To independently verify these K_D values and the assumed 1:1 stoichiometry, ITC measurements with ssDNA oligonucleotides were performed. All oligonucleotides revealed a 1:1 stoichiometry with binding affinities similar to those observed by fluorescence titration experiments. Taken together, *Xtr*Lin28B CSD binds seven to eight nucleotides with limited sequence specificity except for a preferential binding of G at position 1 and T at position 2 (dG^1, dG^{1b}, Figure 20).

Table 6: Equilibrium dissociation constants of *Xtr*Lin28B CSD•ssDNA complexes derived from fluorescence titration experiments assuming a 1:1 stoichiometry[a].

CSD•ssDNA					
Ligand	Sequence	K_D (nM)	Ligand	Sequence	K_D (nM)
dT_5	TTT TT	680 ± 40	dG^1	G TTT TTT	12 ± 1
dT_6	TTT TTT	169 ± 30	dG^2	T GTT TTT	39 ± 4
dT_7	TTT TTT T	59 ± 5	dG^3	T TGT TTT	139 ± 8
dT_8	TTT TTT TT	32 ± 1	dG^4	T TTG TTT	119 ± 16
dT_9	TTT TTT TTT	44 ± 3	dG^5	T TTT GTT	45 ± 4
			dG^6	T TTT TGT	81 ± 10
dA^1	A TTT TTT	84 ± 5	dG^7	T TTT TTG	80 ± 10
dA^2	T ATT TTT	311 ± 24	dG_7	G GGG GGG	-
dA^3	T TAT TTT	137 ± 10			
dA^4	T TTA TTT	124 ± 12	dC^1	C TTT TTT	45 ± 4
dA^5	T TTT ATT	176 ± 8	dC^2	T CTT TTT	137 ± 22
dA^6	T TTT TAT	147 ± 15	dC^3	T TCT TTT	71 ± 3
dA^7	T TTT TTA	113 ± 15	dC^4	T TTC TTT	73 ± 6
dA_7	A AAA AAA	-	dC^5	T TTT CTT	154 ± 26
			dC^6	T TTT TCT	87 ± 14

[a] All experiments were performed at 15 °C in 20 mM Tris/HCl, pH 8.0 and 60 mM KCl.
(-) not determined due to insufficient affinity.

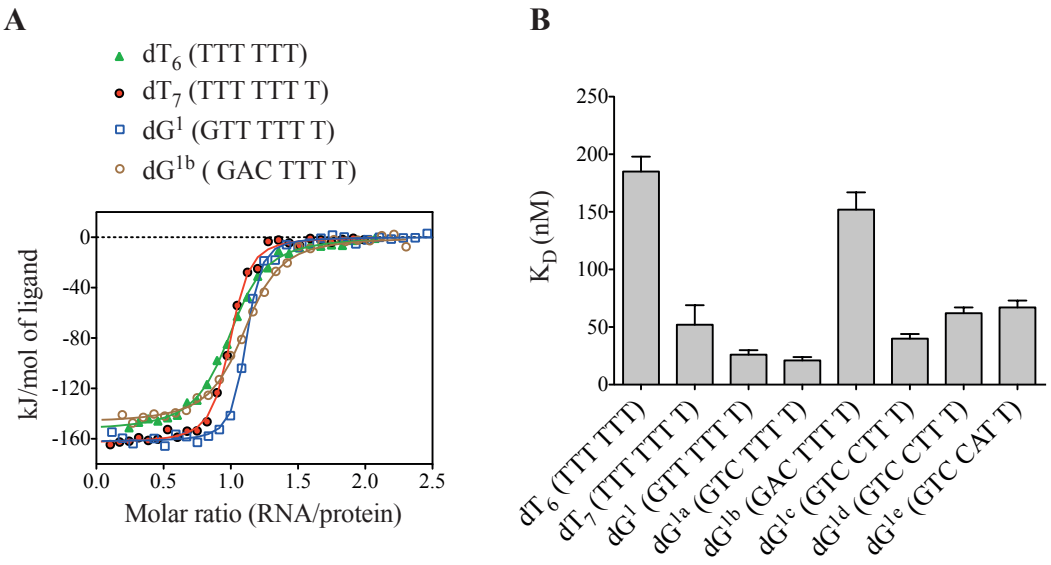


Figure 20: Equilibrium binding data of ssDNA oligonucleotides derived from ITC experiments with *Xtr*Lin28B CSD. (**A**) Representative plots of four ITC experiments. Dissociation constants are shown in (**B**).

3.2.2.2 *RNA binding*

On the RNA level, binding of the oligouridines rU_6 and rU_7 was significantly impaired with an associated 10 to 15-fold increase in K_D with respect to hexa- and heptathymidine (compare Table 6 and Table 7). The missing 5'-methyl group and thus the weaker hydrophobicity of the bases might be the reason for the observed difference. However, U-to-G exchanges at position 1, led to similar K_D values as for the corresponding DNA oligonucleotides (rG^1, Figure 21, Table 7), showing that the differences between rU_6/rU_7 and dT_6/dT_7 can be traced back to this particular binding subsite. Purine-rich oligonucleotides showed no binding at all. This indicates that the CSD requires at least a few pyrimidines for binding ($r(A)_8$ and rGA_6; Figure 21 and Table 7).

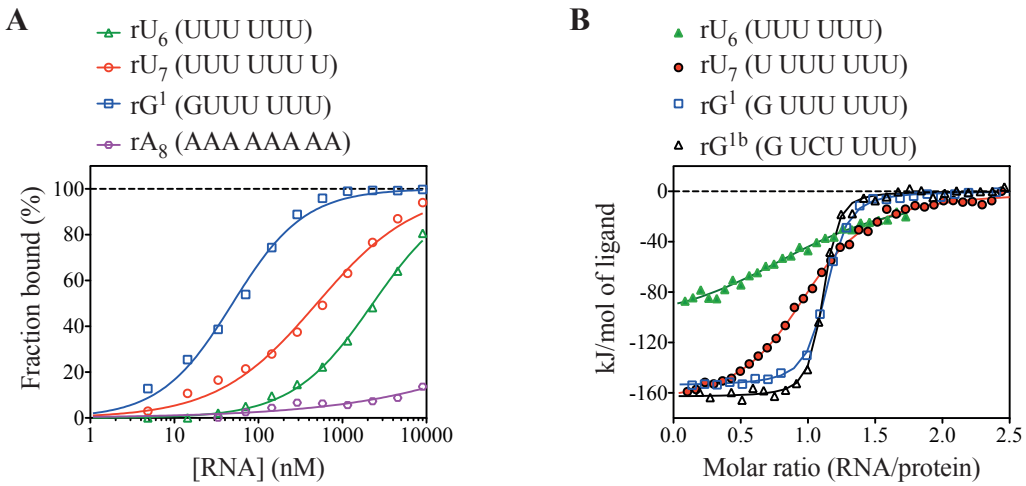


Figure 21: Equilibrium binding data of ssRNA oligonucleotides derived from (**A**) fluorescence titration and (**B**) isothermal titration calorimetry experiments with *Xtr*Lin28B CSD. The resulting K_D values are listed in Table 7.

Table 7: Equilibrium dissociation constants of *Xtr*Lin28B CSD•ssRNA complexes derived from fluorescence titration[a] and isothermal titration calorimetry[b] experiments.

Trp quencher assay			ITC		
Ligand	Sequence	K_D (nM)	Ligand	Sequence	K_D (nM)
rU_6	UUU UUU	2300 ± 80	rU_6	UUU UUU	2760 ± 790
rU_7	UUU UUU U	496 ± 38	rU_7	UUU UUU U	530 ± 65
rG^1	GUU UUU U	49 ± 3	rG^1	GUU UUU U	35 ± 2
rA_8	AAA AAA AA	-	rG^{1b}	GUC UUU U	27 ± 3
			rGA_6	GAA AAA A	-

[a] Fluorescence titration experiments were performed at 20 °C in 20 mM Tris/HCl, pH 7.5, 60 mM KCl, 40 mM N-acetyltryptophanamide.
[b] ITC experiments were performed at 15 °C in 20 mM Tris/HCl, pH 8.0 and 60 mM KCl.
(-) not determined due to insufficient affinity.

To narrow the binding site of *Xtr*Lin28B CSD within its biological target pre-*let-7f*, the minimal binding site of Lin28 (the pre-element) was divided into three RNA heptamers and analyzed with respect to their binding affinity (Figure 22 A, B, Table 8). All oligonucleotides could bind to the CSD even though only r*let-7f* ^{a} showed a binding constant in the low nanomolar range (K_D = 46 nM or 36 nM depending on the method). Interestingly, this RNA sequence is derived from a ss terminal loop of preE-*let-7f* (Figure 22 B) and thus is accessible for binding without structural rearrangements. Moreover, it contains the relevant G at position 1, U at positions 2 and pyrimidine at position 5. Upon mutation of these nucleotides to A, the binding of *Xtr*Lin28B CSD is dramatically impaired (r*let-7f* ^{a3}, Figure 22 D).

Next, the optimal CSD•RNA interaction surface length was analyzed using preE-*let-7f* ^{a} as a template. While an additional G at the 5′ end slightly impaired binding, an extra C at the 3′ end (position 8) improved the binding affinity considerably (r*let-7f* ^{a}, r*let-7f* ^{a4}, r*let-7f* ^{a5}). Adding both nucleotides (r*let-7f* ^{a6}) led to a similar K_D as seen for r*let-7f* ^{a5} (Figure 22 D, Table 8).

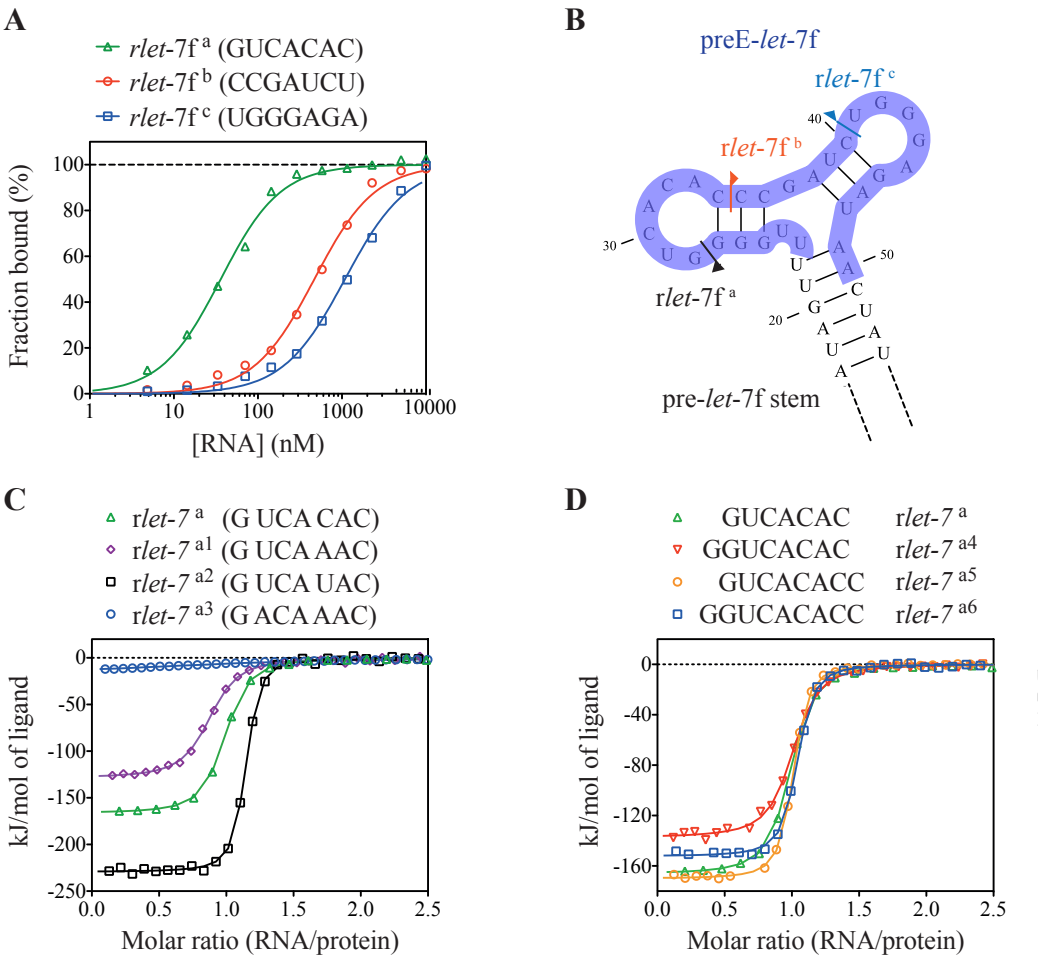


Figure 22: Equilibrium binding data of preE-*let-7* derived ssRNA oligonucleotides with *Xtr*Lin28B CSD. (**A**) Fluorescence titration experiments of *Xtr*Lin28B CSD performed with ssRNA-oligonucleotides derived from preE-*let-7f*. (**B**) Secondary-structure prediction of pre-*let-7f* using CLC-Genomics Workbench 3.6.5. (**C, D**) ITC experiments with *Xtr*Lin28B and pre-*let-7f* ^{a}-derived oligonucleotides.

In summary, *Xtr*Lin28B CSD preferentially binds ss pyrimidine-rich oligonucleotides with up to eight bases. However, it only has moderate sequence specificity, as sequence-specific binding is only mediated at positions 1, 2 and, partially, 5.

Table 8: Equilibrium dissociation constants of preE-*let-7* derived ssRNA oligonucleotides with *Xtr*Lin28B CSD. (**A**) .

Trp quencher assay [a,b]			
Ligand	Sequence	K_D (nM)	Stoichiometry
r*let-7f*[a]	GUC ACA C	36 ± 1.8	-
r*let-7f*$_b$	CCG AUC U	469 ± 23	-
r*let-7f*$_c$	UGG GAG A	1090 ± 60	-
ITC [c]			
Ligand	Sequence	K_D (nM)	Stoichiometry
r*let-7f* [a]	G UCA CAC	46 ± 3	1.01 ± 0.04
r*let-7f* [a1]	G UCA UAC	13 ± 2	1.10 ± 0.04
r*let-7f* [a2]	G UCA AAC	54 ± 9	0.85 ± 0.09
r*let-7f* [a3]	G ACA AAC	1360 ± 600	1.03 ± 0.17
r*let-7f* [a4]	GG UCA CAC	66 ± 9	0.98 ± 0.07
r*let-7f* [a5]	G UCA CAC C	18 ± 2	1.04 ± 0.03
r*let-7f* [a6]	GG UCA CAC C	21 ± 4	1.00 ± 0.06

[a] Fluorescence titration experiments were performed at 20 °C in 20 mM Tris/HCl, pH 7.5, 60 mM KCl, 40 mM N-acetyltryptophanamide.
[b] Binding data were fit assuming a 1:1 stoichiometry.
[c] ITC experiments were performed at 15 °C in 20 mM Tris/HCl, pH 8.0 and 60 mM KCl.

3.3 STRUCTURE OF LIN28 COLD-SHOCK DOMAINS

3.3.1 *Structure determination of XtrLin28B and hLin28B CSD*

To investigate the structural basis for the observed sequence preferences, crystallization trials with human, mouse, *C. elegans* and *Xenopus tropicalis* Lin28 variants were performed as described in Section 2.6.1. First, initial hexagonal crystals of *Xtr*Lin28B$^{27\text{-}114}$ CSD were observed after 2 d at 20 °C using 2.8 M sodium or 4 M potassium acetate as a precipitant. In subsequent fine screens 0.1 M HEPES, pH 7.0, 2.5 M sodium acetate was determined to be the best reservoir solution for crystallization yielding hexagonal *Xtr*Lin28B CSD crystals of 120 µm x 80 µm x 80 µm. Crystals of human Lin28B$^{24\text{-}111}$ CSD were obtained at 4 °C using 2 M ammonium sulfate, 0.2 M sodium chloride and 0.1 M MES, pH 6.5 as reservoir buffer (Figure 23).

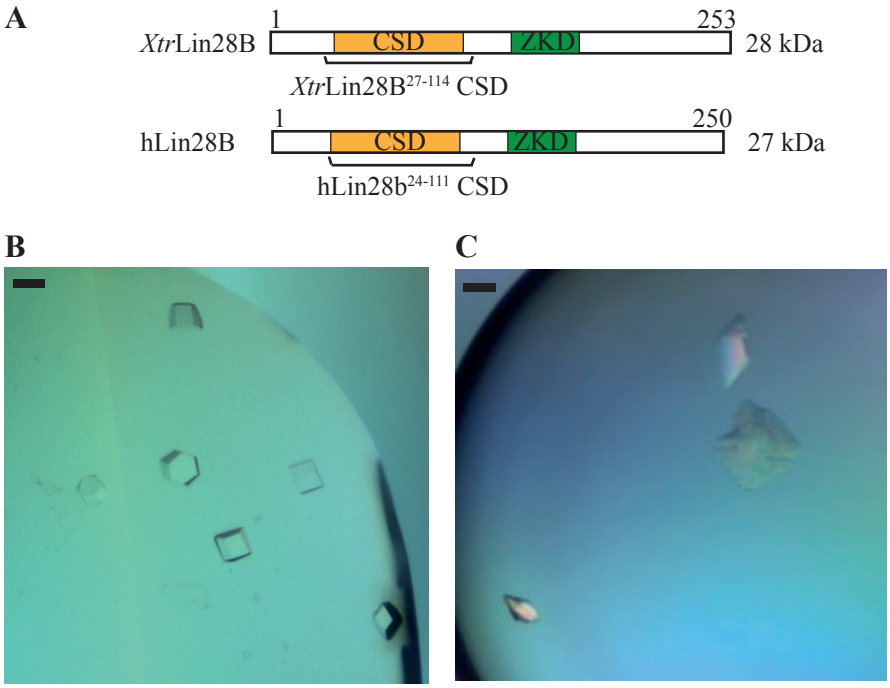


Figure 23: Crystals of *Xtr*Lin28B$^{27\text{-}114}$ CSD and hLin28B$^{24\text{-}111}$ CSD. (**A**) Domain organization *Xtr*Lin28B and hLin28B. The crystallized constructs are indicated. (**B**) Hexagonal crystals of *Xtr*Lin28B grew at 20 °C using 20 mM HEPES, pH 7.0, 2.5 M sodium acetate as reservoir buffer. (**C**) hLin28B CSD crystals were obtained at 4 °C using 2 M $(NH_4)_2SO_4$, 0.2 M NaCl and 0.1 M MES, pH 6.5 as reservoir buffer. The black bars correspond to 100 µm.

X-ray data were collected as described in Section 2.6.3. The best crystal of *Xtr*Lin28B CSD diffracted to 1.06 Å and belonged to the trigonal space group $P3_121$ with cell dimensions of a = b = 52.6 Å, c = 137.9 Å and $\alpha = \beta = 90°$, $\gamma = 120°$. Crystals of hLin28B diffracted to 1.95 Å and belonged to the orthorhombic space group $P2_12_12_1$ with cell dimensions of a = 46.2 Å, b = 62.5 Å, c = 77.3 Å and $\alpha = \beta = \gamma = 90°$. Data collection statistics are summarized in Table 9.

Table 9: Data collection statistics for *Xtr*Lin28B$^{27\text{-}114}$ CSD and hLin28B$^{24\text{-}111}$ CSD.

Data collection	*Xtr*Lin28B CSD	hLin28B CSD
Beamline	BESSY II 14.1	BESSY II 14.1
Wavelength (Å)	0.91841	0.91841
Temperature (K)	100	100
Space group	$P3_121$	$P2_12_12_1$
Cell dimensions		
a (Å)	52.6	46.2
b (Å)	52.6	62.5
c (Å)	137.9	77.3
α (°)	90	90
β (°)	90	90
γ (°)	120	90
V_M (Å^3/Da) †	2.76	2.88
Resolution (Å)	32.3 (1.12) - 1.06	48.59 (1.98) - 1.95
No. of observed reflections	685770 (90497)	68196 (4095)
No. of unique reflections	100170 (15821)	16277 (774)
R_{meas} (%)	6.1 (67.1)	8.0 (23.5)
$< I/\sigma(I) >$	17.3 (2.8)	9.5 (4.1)
Completeness (%)	99.5 (98.2)	96.6 (94.4)

† According to [230].
‡ Numbers in parentheses apply to the highest resolution shell.

Initial phases were obtained by molecular replacement using Phaser and *Salmonella typhimurium* cold-shock protein (3I2Z) or *Xtr*Lin28B CSD as search model (see Section 2.6.4). *Xtr*Lin28B CSD contained four molecules, whereas hLin28B CSD contained two molecules per asymmetric unit. After initial model building using ARP/wARP [231] and several rounds of iterative model building and refinements including anisotropic B-factors for *Xtr*Lin28B CSD and TLS groups for Lin28B CSD (see Section 2.6.5), final R_{work}/R_{free}-factors of 12.0/13.9% and 18.0/23.4% were obtained, respectively. For both models no Ramachandran outlieres were detected. More than 98% of all residues were in the favored region of the Ramachandran plot. The models were validated using the Molprobity server [224] and deposited in the Protein Data Bank (PDBs 3ULJ, 4A4I). A summary of the refinement statistics is given in Table 10.

Table 10: Refinement statistics for *Xtr*Lin28B$^{27\text{-}114}$ CSD and hLin28B$^{24\text{-}111}$ CSD.

Refinement	*Xtr*Lin28B CSD	hLin28B CSD
R_{work}† / R_{free}‡ (%)	12.0 / 13.9	18.0 / 23.4
No. of atoms		
Protein	1378	1308
Hetero atoms	38	11
Solvent atoms	272	208
RMSD		
Bond lengths (Å)	0.013	0.020
Bond angles (°)	1.573	1.796
Ramachandran statistics		
Most favored (%)	99.4	98.1
Allowed (%)	0.4	1.9
Outliers (%)	0.0	0.0

† $R_{work} = \frac{\sum ||F_{calc}| - |F_{obs}||}{\sum |F_{obs}|}$.

‡ R_{free} was calculated with the 5% of the data excluded from the refinement.

3.3.2 *Structure analysis of Lin28 CSDs*

Both hLin28B CSD and *Xtr*Lin28B CSD are remarkably similar with a RMSD for the backbone C_α atoms of only 0.2 Å (Figure 24 A, B). The monomeric proteins reveal a CSD-typical oligosaccharide/oligonucleotide-binding fold [232, 176], which consists of a β-barrel composed of five antiparallel β-strands. The β-barrel can be further divided into two subdomains which are separated by a short 3_{10} helix and an extended loop between β-strands 3 and 4. Interestingly, the overall architecture of both proteins is quite similar to those of other CSDs, however, it has a higher structural similarity to bacterial Csps than to eukaryotic CSD-containing proteins like the human YB-1 (Y-box binding factor 1, Max *et al., in preparation*).

Consistent with this, both Lin28 CSDs resemble their bacterial homologs in having a strong amphipathic character. One side of the β-barrel shows an acidic surface that is unfavorable for nucleic acid binding. The opposite surface forms an hydrophobic platform interspersed with highly conserved, surface exposed aromatic residues that are surrounded by basic and polar residues (Figure 24 C). The latter platform can associate with ss nucleic acids and comprises the highly conserved ribonucleoprotein motif 1 (RNP1, MGFGFISM) and RNP2 (DVFVH-F/Q) (see Figure 11 and 24). Most of these residues are known to interact with nucleic acid ligands via hydrophobic and hydrogen bond contacts and are therefore highly conserved both in sequence and structure [182, 176]. In summary, Lin28B CSDs provide a pre-formed nucleic acid binding platform that can interact with ss nucleic acids via base stacking by its exposed aromatic residues.

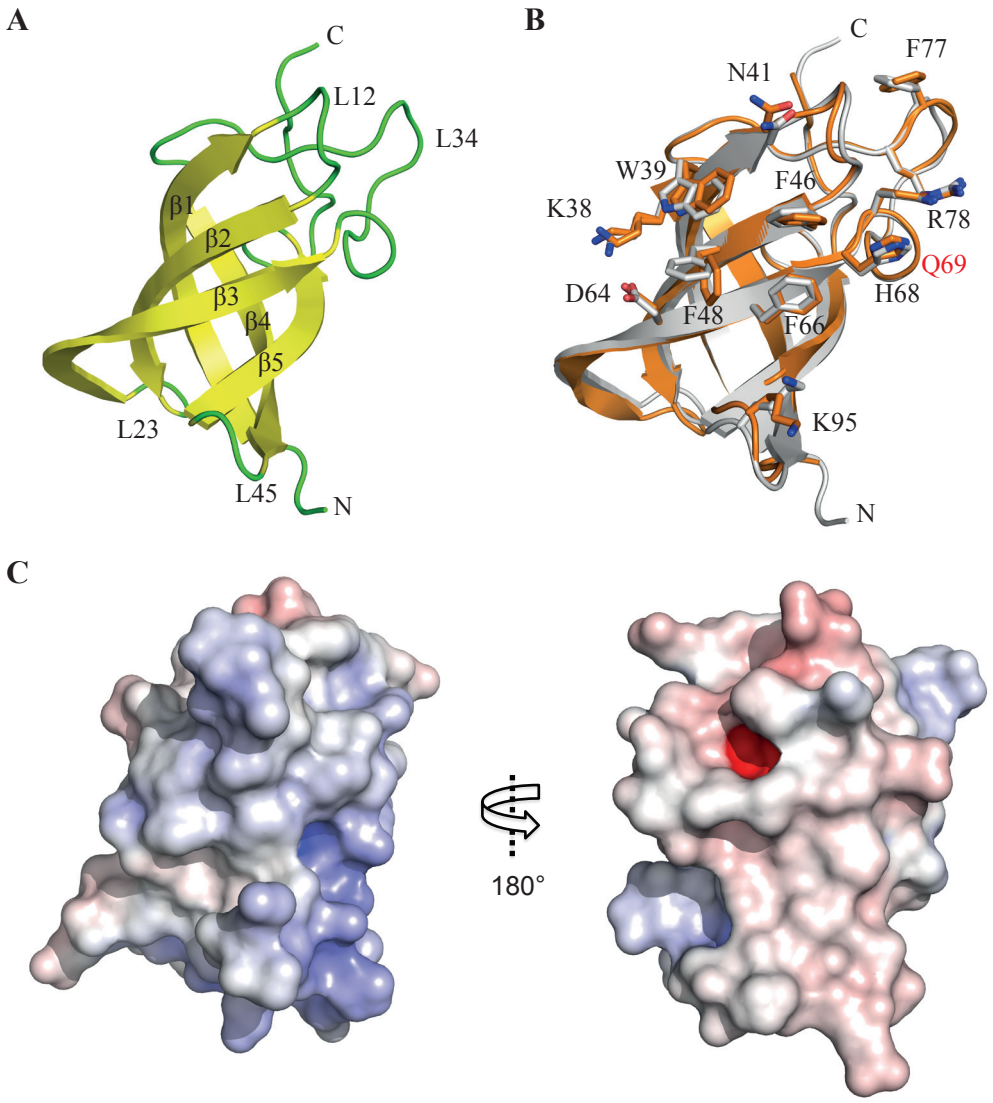


Figure 24: Crystal structures of *Xtr*Lin28B CSD and hLin28B CSD. **(A)** Cartoon representation of *Xtr*Lin28B CSD. The β-barrel-like structure can be divided into two subdomains, which are separated by a short 3_{10} helix and an extended loop between β-strands 3 and 4 (L34). **(B)** Superimposition of hLin28B CSD and *Xtr*Lin28B. Both structures are highly conserved and show a pre-formed nucleic acid binding platform with exposed aromatic and basic residues at one side of the protein. **(C)** The electrostatic surface potential of *Xtr*Lin28B CSD was calculated with APBS [233] for pH 7 with a range from -10 (red) to +10 kT (blue). The Lin28B CSD has a strong amphipathic character with one side being acidic, while the other side harbors the hydrophobic nucleic acid binding platform that is interspersed with basic amino acids.

Compared to their bacterial homologs, Lin28B CSDs display three structural differences. First, β-strand 2 and the following loop are extended by an eight-amino acid insertion. Second, two surface-exposed Lys within β-strand 4 create a strong basic surface next to the hydrophobic platform. Finally, a conserved Phe that was shown to be involved in base stacking interactions with nucleic acid bases in bacterial Csps (34,37,57), is replaced by Gln69 (see Figure 24 B and 11).

3.4 STRUCTURE OF LIN28 CSD IN COMPLEX WITH HEXA- AND HEPTATHYMIDINE

3.4.1 *Structure determinaton of XtrLin28B CSD•dT_6 and XtrLin28B CSD•dT_7*

Next, co-crystallization trials of *Xtr*Lin28B CSD with bound oligonucleotides were performed. To obtain homogeneous protein•nucleic acid complexes, the CSD was incubated with a 1.2 molar excess of oligonucleotide and subjected to SEC (see Section 2.6.2). The presence of a bound oligonucleotide was observable by the A_{260}/A_{280} ratio of the peak fractions. Futhermore, all complexes eluted with 1:1 stoichiometry as judged from their peak retention volume (Table 11 and Figure 25).

Table 11: Size-exclusion chromatography of *Xtr*Lin28B CSD•nucleic acid complexes.

	Peak retention volume (ml)	SEC derived Mw (kDa)	A_{260}/A_{280}
*Xtr*Lin28B CSD	83.98	9.20	0.65
*Xtr*Lin28B CSD•rU_6	83.34	9.55	2.08
*Xtr*Lin28B CSD•dT_6	82.22	10.21	1.47
*Xtr*Lin28B CSD•dT_7	79.41	12.05	1.44

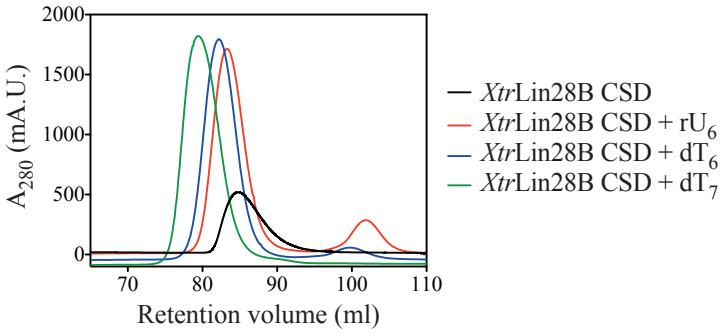


Figure 25: Size-exclusion chromatograms of *Xtr*Lin28B CSD•nucleic acid complexes. The peak retention volumes, SEC-derived molecular weight (Mw) and A_{260}/A_{280} ratio of the peak fractions are listed.

Initial crystal needles and small crystal plates of *Xtr*Lin28B CSD in complex with hexa- and heptathymidine could be obtained using PEG 3350 and PEG 4000 as precipitant. After optimization, clusters of large crystal plates grew at 4 °C within 4 d to a size of up to 250 µm x 60 µm x 20 µm (Figure 26). Single plates were broken off and used for data collection. The best crystal of *Xtr*Lin28B•dT_6 diffracted to 1.75 Å and belonged to spacegroup $P2_12_12_1$ with cell dimensions of a = 51.8 Å, b = 81.9 Å, c = 99.9 Å and $\alpha = \beta = \gamma = 90°$. Crystals of *Xtr*Lin28B•dT_7 diffracted X-rays to a maximal resolution of 1.92 Å and belonged to the same space group as the hexathymidine complex with almost the same cell dimensions (a = 51.6 Å, b = 77.7 Å, c = 100.4 Å and $\alpha = \beta = \gamma = 90°$. A summary of the data collection statistics is given in Table 12.

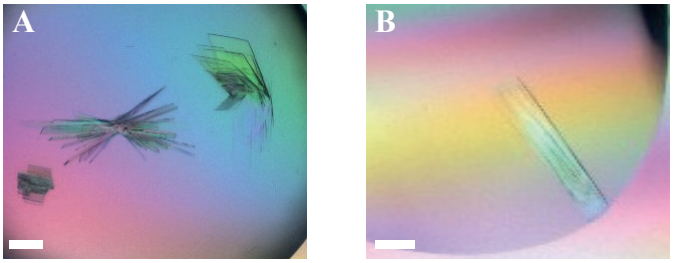


Figure 26: Crystals of hexa- and heptathymidine bound *Xtr*Lin28B CSD. (**A**) Crystals of *Xtr*Lin28B CSD in complex with dT_6 were obtained using 15% (w/v) PEG 3350, 0.1 mM Bis-Tris pH 5.5, 0.1 M sodium thiocyanate as reservoir buffer. (**B**) Crystals of *Xtr*Lin28B CSD•dT_7 grew at 4 °C using 17% (w/v) PEG 3350, 0.2 M sodium thiocyanat as reservoir buffer. The white bar corresponds to 100 µm.

Table 12: Data collection statistics for *Xtr*Lin28B CSD•dT_6 and *Xtr*Lin28B CSD•dT_7.

Data collection	*Xtr*Lin28B CSD•dT_6	*Xtr*Lin28B CSD•dT_7
Beamline	BESSY II 14.2	BESSY II 14.2
Wavelength (Å)	0.91841	0.91841
Temperature (K)	100	100
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a (Å)	51.8	51.6
b (Å)	81.9	77.7
c (Å)	99.9	100.4
α (°)	90	90
β (°)	90	90
γ (°)	90	90
V_M (Å^3/Da) †	2.26	2.09
Resolution (Å)	32.91 (1.85) - 1.75	45.90 (2.04) - 1.92
No. of observed reflections	118417 (15437)	93214 (14521)
No. of unique reflections	39188 (6140)	30854 (4484)
R_{meas} (%)	7.8 (39.0)	8.9 (43.3)
$< I/\sigma(I) >$	13.2 (3.3)	11.5 (3.3)
Completeness (%)	89.4 (87.8)	97.9 (96.9)

† According to [230].
‡ Numbers in parentheses apply to the highest resolution shell.

The structures of both complexes were solved by molecular replacement using the model of apo *Xtr*Lin28B CSD as a template (see Section 2.6.4). The DNA-oligonucleotides were placed manually into the electron density using COOT [219]. After multiple refinement cycles (see Section 2.6.5), final R_{work}/R_{free}-factors of 17.1/19.8% (*Xtr*Lin28B CSD•dT_6) and 20.7/24.5% (*Xtr*Lin28B CSD•dT_6)

could be obtained. The atomic model was validated and deposited in the Protein Data Bank (PDBs 4A75, 4A76). More than 99% of the residues are in the favored or allowed region in the Ramachandran plot. The refinement statistics are summarized in Table 13.

Table 13: Refinement statistics for *Xtr*Lin28B CSD•dT_6 and *Xtr*Lin28B CSD•dT_7.

Refinement	*Xtr*Lin28B CSD•dT_6	*Xtr*Lin28B CSD•dT_7
R_{work}† / R_{free}‡ (%)	17.1 / 19.8	20.7 / 24.5
No. of atoms		
Protein atoms	2743	2695
Nucleic acids atoms	400	435
Hetero atoms	6	0
Solvent atoms	506	345
RMSD		
Bond lengths (Å)	0.012	0.012
Bond angles (°)	1.477	1.400
Ramachandran statistics		
Most favored (%)	98.3	99.4
Allowed (%)	1.7	0.3
Outliers (%)	0.0	0.3

† $R_{work} = \frac{\sum ||F_{calc}| - |F_{obs}||}{\sum |F_{obs}|}$.

‡ R_{free} was calculated with the 5% of the data excluded from the refinement.

3.4.2 *Structure analysis of ssDNA bound Lin28 CSDs*

To simplify the discussion, we define dT_6 nucleotides as T-2 to T-7, and dT_7 nucleotides as T-1 to T-7. The crystal lattices of the dT_6- and dT_7-bound structures are similar, both having four protein•oligonucleotide complexes in the asymmetric unit. While for two of the bound DNA ligands all nucleotides are clearly resolved, the electron densities of the other two DNA molecules reveal only four nucleotides (T-4* to T-7*). The reason for this is that T-3 of the completely resolved oligonucleotides contacts another protein and thus occupies binding subsite 2 within another protein molecule of the asymmetric unit (Figure 27).

Hexa- and heptathymidine superimpose well (RMSD = 0.2 Å) and bind in an identical curved conformation to the hydrophobic platform of *Xtr*Lin28B CSD with their 5′ end located in the loop region between β-strand 3 and 4, then proceeding over a groove formed by residues from β-strands 5 and 3, crossing β-strand 2 until making a final kink over β-strand 1 (Figure 28 A). In the dT_7-bound structure an additional nucleotide (T-1) is visible at the 5′ end stacking on top of Phe77 (Figure 28 B). The bases point to the protein surface and interact mainly via stacking with the surface exposed Phe77, Phe66, Phe48, and

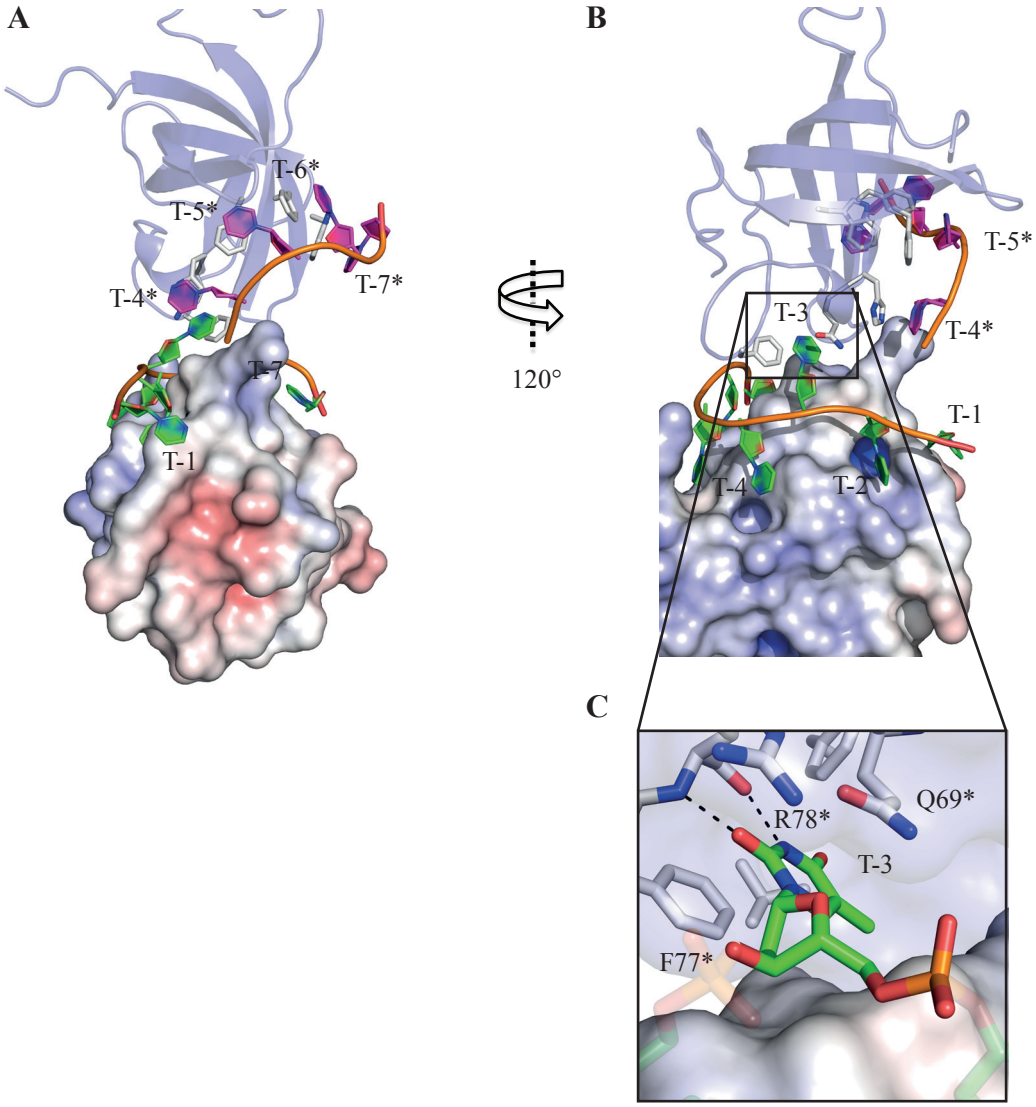


Figure 27: Binding of heptathymidine to *Xtr*Lin28B CSD in the crystal. (**A, B**) For two of the four protein molecules in the asymmetric unit, the nucleotide binding sites are fully occupied by a single heptathymidine. The two other protein molecules reveal density for only four thymidines (T-4* to T-7*). The reason for this is that T-3 of one oligonucleotide occupies the hydrophobic binding pocket formed by Phe77, Gln69, and Val42 that otherwise accommodates T-2. (**C**) Close-up view of T-3 binding site. Binding is mainly mediated through stacking interactions with Phe77 and hydrogen bonds with Arg78. Residues marked with asterisks belong to another molecule of the asymmetric unit.

Trp39 as well as His68 (Figure 28 C). A substitution of these individual side chains by Ala resulted in a dramatic decrease in binding affinity with respect to the preE derived rlet-f ^{a5} oligonucleotide (see Figure 29). Moreover, a replacement of the aromatic residues impaired binding more strongly than the His68Ala mutation. This underlines the dominance of base stacking interactions in the RNA•protein contacts. The observed binding site is already preformed in the apo protein, and therefore only subtle changes are recognizable upon nucleic acid binding. From the four conserved aromatic residues that are directly

involved in base stacking interactions, only the side chain of Phe48 performs a larger conformational change upon nucleotide binding (see Figure 28 D).

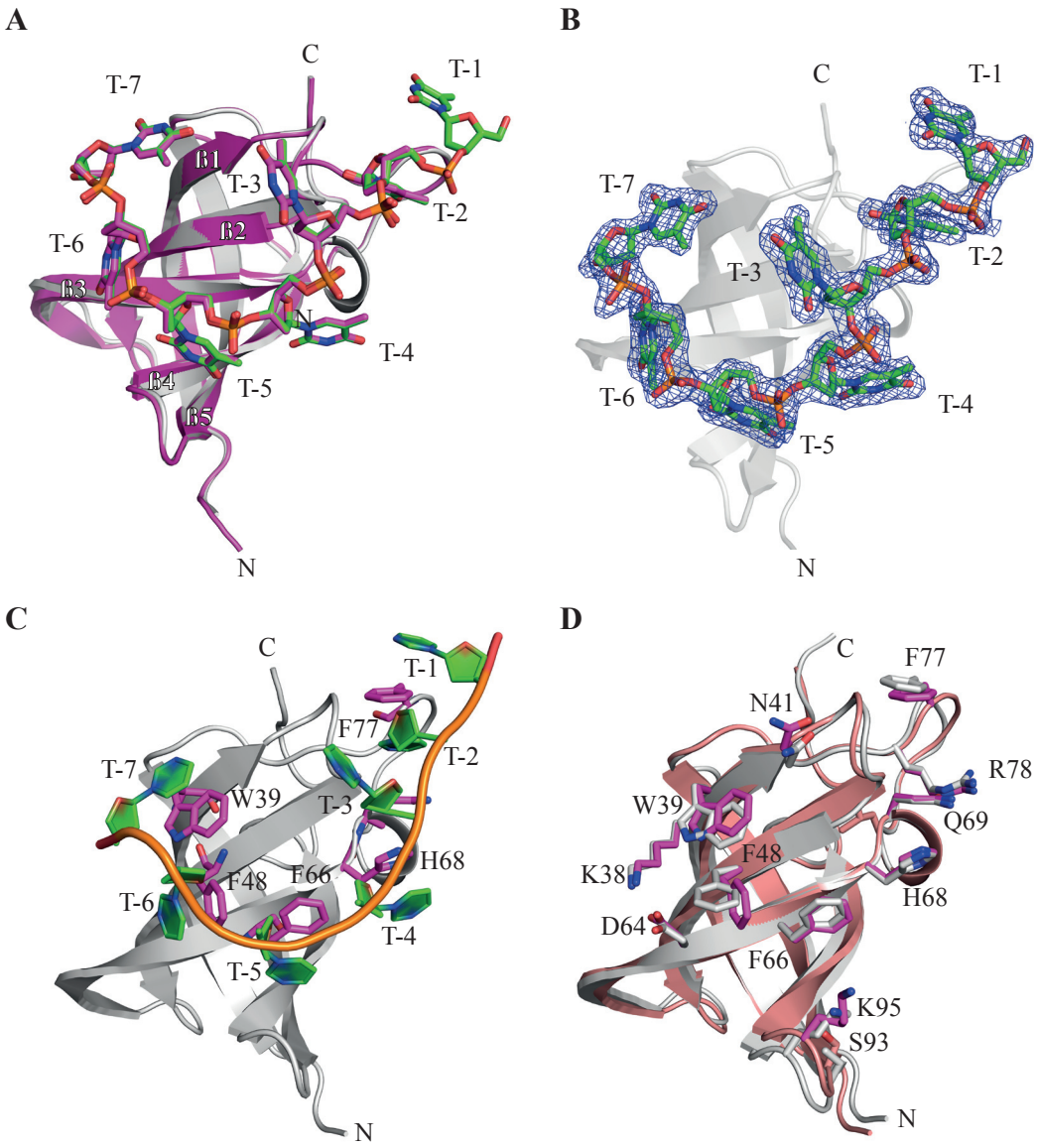


Figure 28: Binding of dT_6 and dT_7 by *Xtr*Lin28B CSD is dominated by base stacking interactions. (**A**) Superimposition of the *Xtr*Lin28B CSD•dT_6 (magenta) and *Xtr*Lin28B CSD•dT_7 crystal structures (gray, protein; green, nucleosides; orange: phosphate). The crystal lattices and binding modes are identical for both structures except for an additional T (T-1) at the 5′ end in the dT_7-bound structure. (**B**) The $2F_o$-F_c electron density for heptathymidine is colored in blue and contoured at 1σ. (**C**) Binding of heptathymidine to *Xtr*Lin28B CSD is dominated by base stacking interactions with four conserved aromatic residues (F77, F66, F48, W39) and one His (His68). (**D**) Superimposition of apo (gray) and dT_7 (rose/magenta) bound *Xtr*Lin28B CSD. The nucleic acid binding site is already pre-formed in *Xtr*Lin28B CSD. Only Phe48 undergoes a larger conformational shift upon nucleic acid binding.

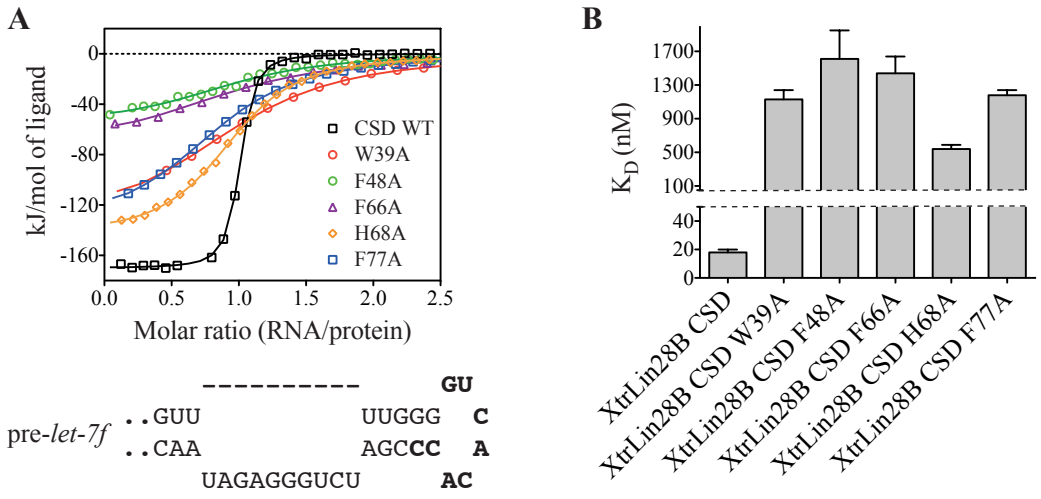


Figure 29: The evolutionary conserved, exposed aromatic residues of *Xtr*Lin28B CSD enable high-affinity binding to ssRNA/ssDNAs. (**A**) ITC experiments of the indicated *Xtr*Lin28B CSD variants and the preE-*let-7* derived ssRNA octamer pre-*let-7f* [a4] and (**B**) the obtained equilibrium dissociation constants.

In comparison to the dT_6-bound structure, the dT_7•*Xtr*Lin28B CSD complex shows an additional nucleotide (T-1) at the 5′ end that stacks on Phe77. Interestingly, nucleotide binding to this site was not observed in previous crystal structures of CSDs with bound oligonucleotides, probably due to the lack of specific hydrogen-bonded contacts with the nucleotide. However, formation of a three-membered stack between T-1, Phe77 and T-2 (Figure 30 A) is consistent with the observed increase in binding affinity from 6- to 7-mers. This continuous stack also explains the favored G at binding subsite 1, as the purine base dramatically improves the stacking interactions with Phe77 and T-2 as judged by modeling. In case of a G, the base is likely further recognized by specific hydrogen bonds, as an A at position 1 impaired binding.

T-2 is bound in a hydrophobic pocket mainly formed by Phe77, Gln69, Val42, Phe40 and the main chain of Arg78. Apart from these hydrophobic interactions, the T-2 base is specifically bound by three hydrogen bonds to the backbone of Arg78 and, via a bridging water molecule, to the backbone of Gly76 (Figure 30 A). A similar sequence-specific interaction mode was reported previously for bacterial Csps [180] and thus seems to explain the conserved preferential binding of U/T at this site. Further base-specific contacts are made by the backbones of Phe97 and Ser93 through groups of T-5 that are normally involved in Watson-Crick TA-base pairing (Figure 30 3B, C). At binding subsite 6, Lys38 and Trp39 contact O^2 of T-6, while Asp64 is connected to HN^3 of the same base and forms an additional salt-bridge with Lys38 (Figure 30 C). While T-4 and T-7 are mainly bound through stacking interactions with the bases, T-3 is not specifically bound but contacts another protein molecule and thus occupies binding subsite 2 within another monomer of the asymmetric unit (see Figure 27 C).

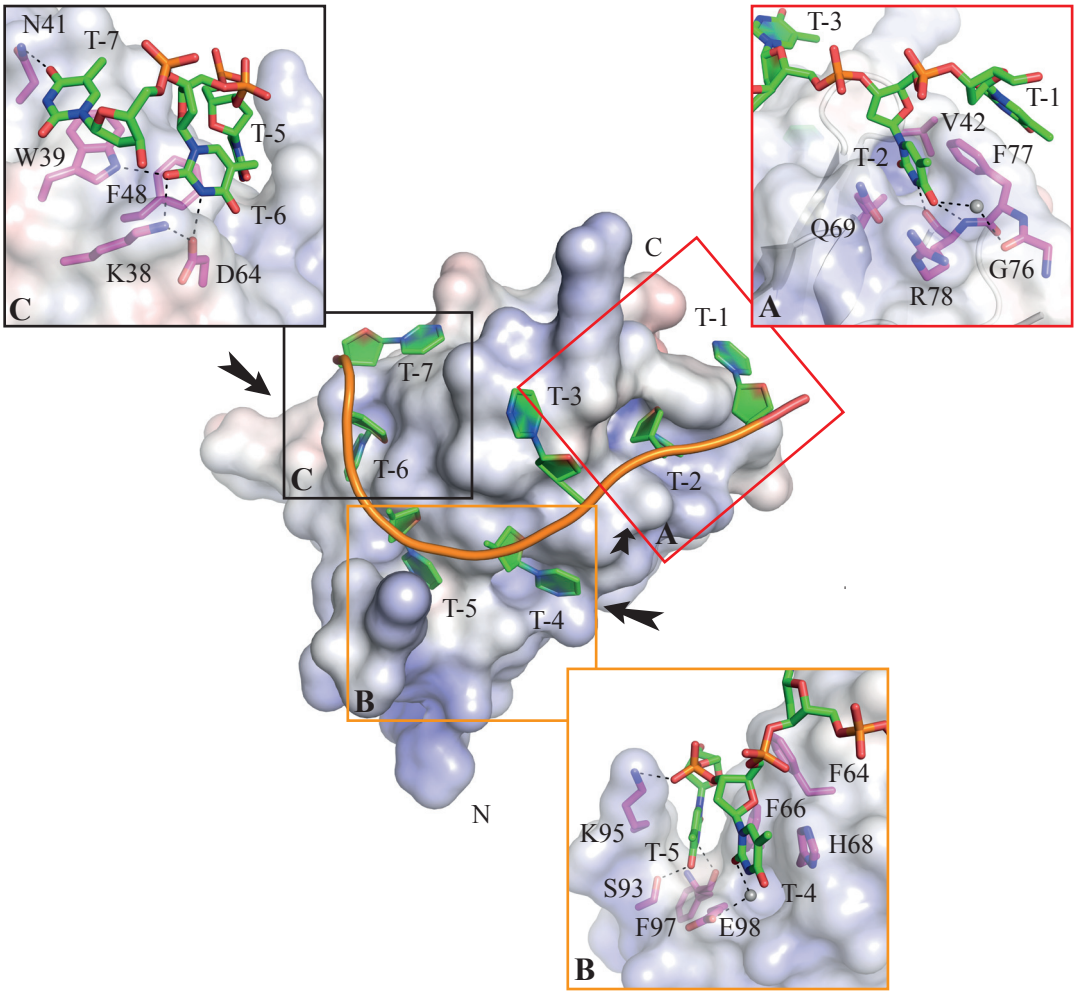


Figure 30: Molecular interactions that promote binding of dT_7 to *Xtr*Lin28B CSD. The electrostatic surface potential of *Xtr*Lin28B CSD upon binding to dT_7 was calculated with APBS for pH 7 with a range from -10 (red) to +10 kT (blue). The bound oligonucleotide is shown in cartoon representation. Water molecules are shown as gray spheres. (**A**) Detailed view of T-2 bound to the hydrophobic pocket at binding subsite 2. Besides hydrophobic contacts including a three-membered stack with Phe77 and T-1, T/U-specific contacts are mediated via the backbone atoms of Arg78 and Phe77. (**B**) Detailed view of T-5 bound in the hydrophobic cleft created by Lys95 and Phe66. Ser93 and Phe97 form hydrogen bonds with O^4 and HN^3 of T-5. Binding of T-4 is only mediated via stacking interactions with His68 and a water molecule-mediated hydrogen bond. (**C**) Detailed view of binding subsites 6 and 7. The thymine base of T-6 is fixed by hydrogen bonds to the side chains of Lys38 and Asp64 that additionally form a stable salt bridge. The bases of T-6 and T-7 are mainly bound via stacking interactions with Phe48 and Trp39, respectively.

3.5 STRUCTURE OF THE LIN28 CSD IN COMPLEX WITH HEXAURIDINE

3.5.1 *Structure determination*

To examine Lin28 binding to natural RNA ligands, further co-crystallization trials with various RNA sequences were performed. Despite extensive efforts, only crystals of *Xtr*Lin28B CSD in complex with hexauridine could be obtained. Crystals grew at 4 °C using 20% (w/v) PEG 3350, 0.2 M potassium or sodium thiocyanate as reservoir buffer (Figure 31). The best crystal of *Xtr*Lin28B•rU_6 diffracted to 1.48 Å and belonged to space group $P2_12_12$ with cell dimensions of a = 96.0 Å, b = 61.2 Å, c = 63.2 Å and $\alpha = \beta = \gamma = 90°$.

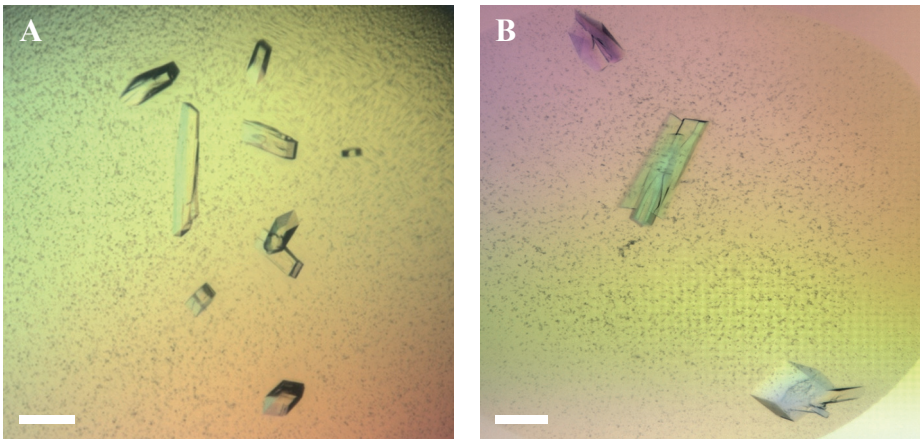


Figure 31: Crystals of *Xtr*Lin28B CSD in complex with hexauridine. Crystals grew at 4 °C using 20% (w/v) PEG 3350, 0.2 M K-thiocyanate (**A**) or Na-thiocyanate (**B**). The white bar corresponds to 100 µM.

Good molecular replacement solutions for rU_6 bound *Xtr*Lin28B CSD were obtained assuming four protein molecules in the asymmetric unit. After an initial model building using ARP/wARP [231], a clear $F_o - F_c$ electron difference density for one RNA-oligonucleotide was revealed. The atomic model for the RNA-hexamer was placed manually into the density using COOT [219] before the entire model was refined in multiple iterative refinement cycles to final R_{work}/R_{free}-factors of 16.9/18.8%. The structure was finally validated and deposited in the protein data bank (pdb-entry: 4ALP). The data collection and refinement statistics are summarized in Table 14.

Table 14: Data collection and refinement statistics for *Xtr*Lin28B CSD•rU_6.[§]

Data collection	*Xtr*Lin28B CSD•rU_6
Beamline	BESSY II 14.1
Wavelength (Å)	0.91841
Temperature (K)	100
Space group	$P2_12_12$
Cell dimensions	
a (Å)	96.0
b (Å)	61.2
c (Å)	63.2
α (°)	90
β (°)	90
γ (°)	90
V_M (Å^3/Da) [*]	2.23
Resolution (Å)	32.4 (1.57) - 1.48
No. of observed reflections	211222 (33252)
No. of unique reflections	60327 (9479)
R_{meas} (%)	6.1 (35.3)
I/σ(I)	15.7 (3.9)
Completeness (%)	96.2 (94.9)
Refinement	
R_{work}[†] / R_{free}[‡] (%)	16.9 / 18.8
No. of atoms	
Protein atoms	2676
Nucleic acids atoms	115
Hetero atoms	12
Solvent atoms	665
RMSD	
Bond lengths (Å)	0.013
Bond angles (°)	1.56
Ramachandran statistics	
Most favored (%)	98.3
Allowed (%)	1.7
Outliers (%)	0.0

[*] According to [230].
[§] Numbers in parentheses apply for the highest resolution shell.
[†] $R_{work} = \frac{\sum ||F_{calc}| - |F_{obs}||}{\sum |F_{obs}|}$.
[‡] R_{free} was calculated with 5% of the data excluded from the refinement.

3.5.2 *Structure analysis of hexauridine bound Lin28 CSD*

Interestingly, the binding mode of hexauridine is different from that observed in the ssDNA-bound complexes. One ssRNA molecule binds to four different protein molecules, two belonging to the same asymmetric unit (Figure 32). The uridine is bound in the T/U-specific hydrophobic pocket mainly formed by Phe77, Gln69 and Val42. However, the uracil base adopts a different orientation and is bound through hydrogen bonds to O^2 and HN^3 via the backbone of Arg78. The unique binding geometry observed in this structure is probably a crystallization artifact, because a 4:1 stoichiometry between Lin28 and RNA was not observed in any of the solution binding experiments which are completely in line with the binding mode observed in the dT_6- and dT_7-bound *Xtr*Lin28B CSD structures (compare Section 3.2.2 and Section 3.4).

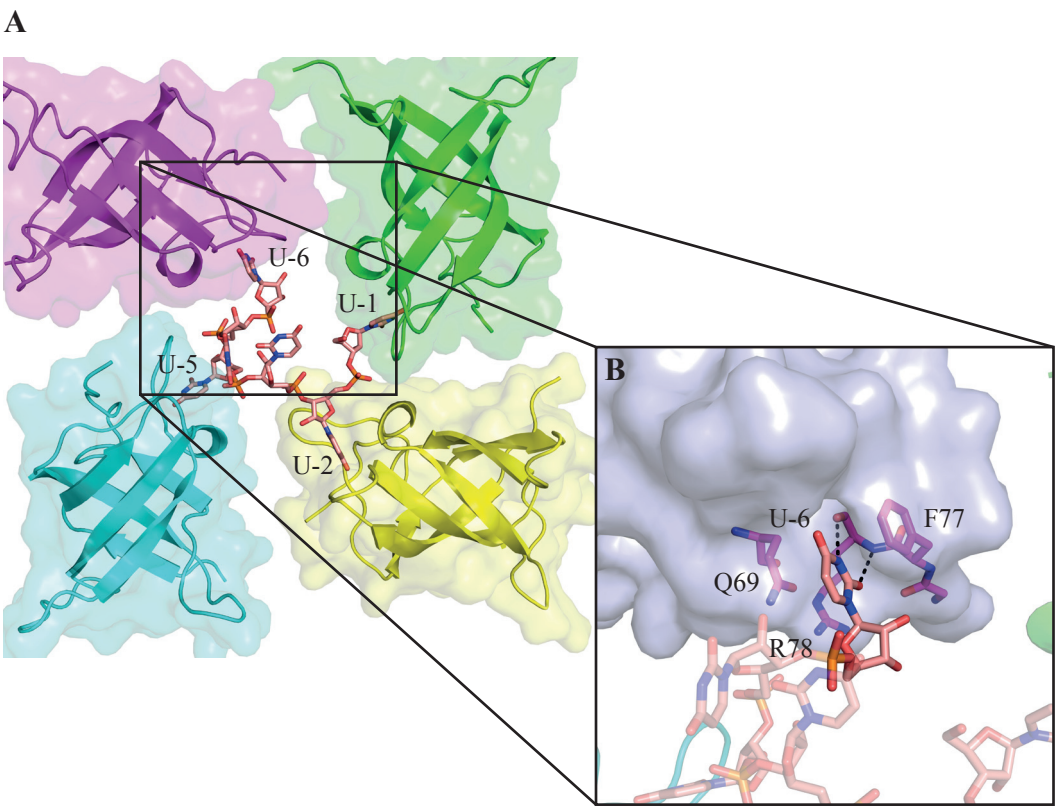


Figure 32: Crystal packing of *Xtr*Lin28B CSD in complex with hexauridine. (**A, B**) In the crystal, one hexauridine binds to four different protein molecules. In all cases, the uridine is bound in the T/U-specific hydrophobic pocket formed by Phe77, Gln69 and Val42 at binding site 2. However, the orientation of the bound uracil is different to the one observed for thymidine in the dT_6/dT_7 bound structures. The backbone of Arg78 forms hydrogen bonds with HN^3 and O^2 instead of O^4 and HN^3 as seen in the *Xtr*Lin28B CSD•dT_7-structure.

3.6 DETERMINANTS OF LIN28•PRE-*let-7* INTERACTIONS

3.6.1 *Both Lin28 RNA-binding domains are required for binding to pre-let-7 and inhibiting its processing by Dicer*

To understand how the observed binding preferences of the individual Lin28 RBDs result in a specific recognition of the natural substrate, EMSAs and *in vitro* processing assays with pre-*let-7f* and Dicer were performed. First, it was determined, which parts of *Xtr*Lin28B are necessary to bind to pre-*let-7* and efficiently inhibit its processing by Dicer.

Using EMSA, the previously obtained K_D value for *Xtr*Lin28B binding to pre-*let-7f* by ITC was confirmed (see Section 3.1.4). A quantification of three EMSA experiments revealed a K_D value of 1.6 ± 0.1 µM. Interestingly, binding of pre-*let-7f* is highly cooperative, as the binding data could best be fit to a Hill equation (see Section 2.5.4) with a Hill coefficient higher than 2. A variant encompassing only the two RBDs (*Xtr*Lin28B$^{27\text{-}175}$ CSD+ZKD) bound with a similar affinity and cooperativity to pre-*let-7f* as WT *Xtr*Lin28B (Figure 33). Unlike the isolated domains, *Xtr*Lin28B CSD+ZKD was able to inhibit the processing of pre-*let-7f* by Dicer to a similar extent as the full-length protein (Figure 34). Moreover, the degree of inhibition was strongly dependent on the Lin28 concentration and thus reflected the competitive nature between Dicer and Lin28 with respect to the accessibility to the Dicer cleavage site within pre-*let-7*.

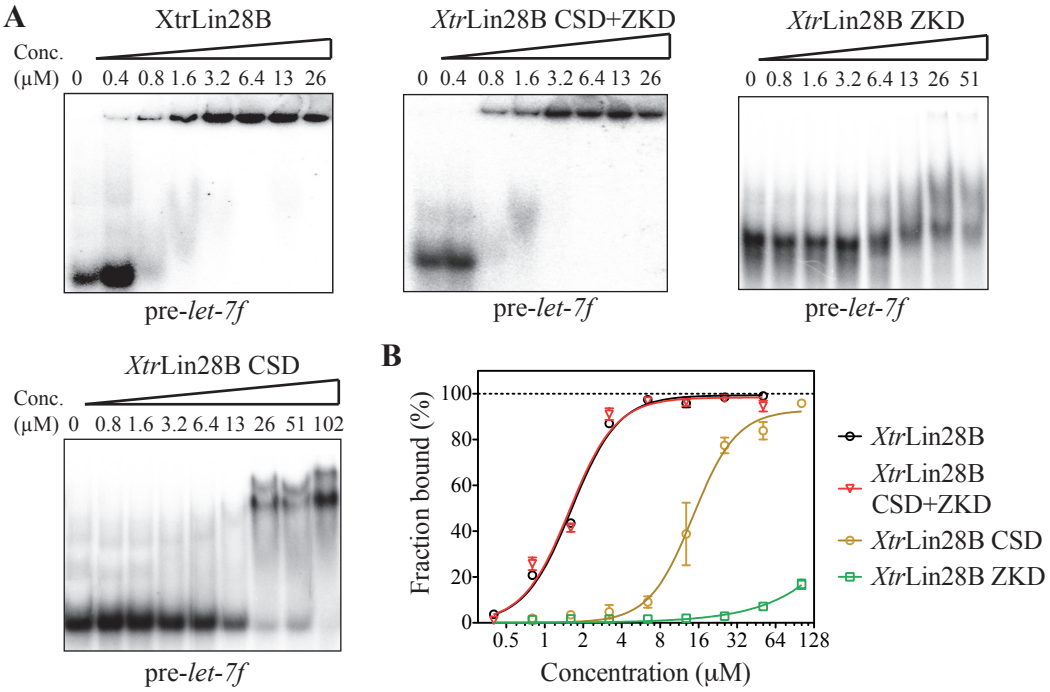


Figure 33: The isolated Lin28 CSD, but not the ZKD, can bind to pre-*let-7*.
(**A**) EMSAs with the indicated protein variants and 1 nM ^{32}P labeled pre-*let-7f*. (**B**) Aggregate binding data derived from three independent EMSA experiments. All binding data were fit with a Hill equation and yielded the following parameters. *Xtr*Lin28B: $K_D = 1.6 \pm 0.1$ µM , $h = 2.3 \pm 0.1$; *Xtr*Lin28B CSD+ZKD: $K_D = 1.6 \pm 0.1$ µM , $h = 2.4 \pm 0.3$; *Xtr*Lin28B CSD: $K_D = 14.5 \pm 1.2$ µM , $h = 2.5 \pm 0.5$.

Next, binding of the isolated RNA-binding domains (*Xtr*Lin28B CSD and *Xtr*Lin28 ZKD) to pre-*let-7f* was analyzed. Surprisingly, little binding was observable for the ZKD (Figure 33). Thus the data imply that the bases of the conserved GGAG motif are not freely accessible for the isolated ZKD, since the presence of the GGAG motif is essential for proper binding of the full-length protein (Figure 35).

By contrast, the isolated CSD bound pre-*let-7f* with an ~ 9-fold lower affinity than the WT protein (Figure 33). Moreover, up to two differently migrating complexes were visible, indicating that the CSD can bind to more than one site. Similar to WT *Xtr*Lin28B and *Xtr*Lin28B CSD+ZKD, binding of the isolated CSD to pre-*let-7f* is highly cooperative (hill coefficient $h = 2.5 \pm 0.5$). This supports the previous observations, that Lin28 might have multiple binding sites within pre-*let-7* miRNA and initial binding of Lin28 CSD is associated with a structural change within other parts of Lin28 or pre-*let-7*. However, despite its ability to bind pre-*let-7f*, *Xtr*Lin28B CSD could not inhibit Dicer mediated processing even at high concentrations (Figure 34).

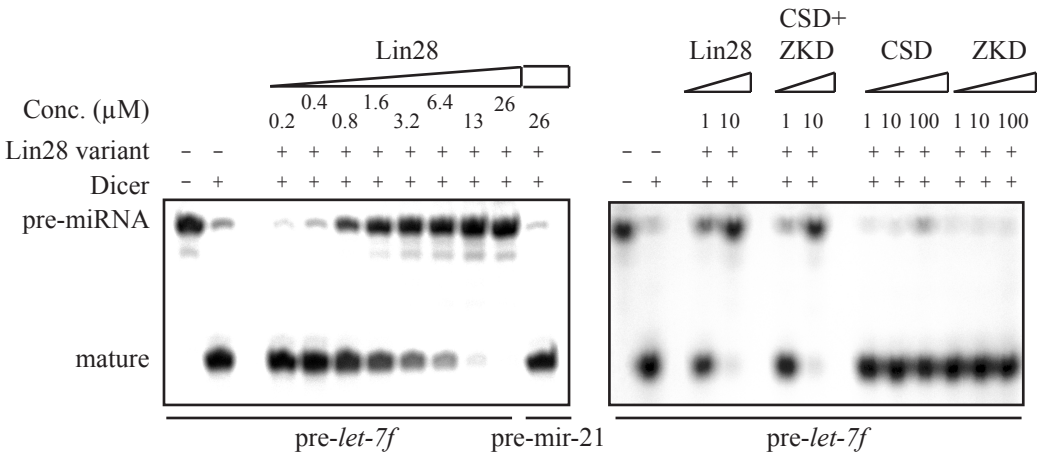
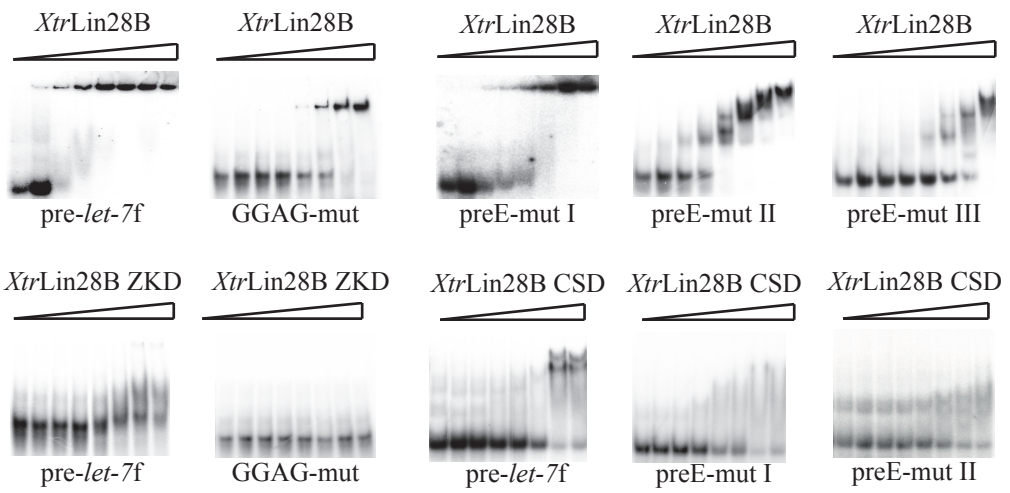


Figure 34: Both Lin28 RNA-binding domains are required for inhibiting pre-*let-7* processing by Dicer. *In vitro* pre-miRNA processing reaction on ^{32}P 5′ end labeled pre-*let-7f*. The indicated concentrations of *Xtr*Lin28B, *Xtr*Lin28B CSD+ZKD, *Xtr*Lin28B CSD and *Xtr*Lin28B ZKD were added to 1 nM pre-*let-7f* in the presence or absence of human Dicer. The samples were resolved by 10% (w/v) denaturing PAGE and visualized by autoradiography.

Considering the binding preference of *Xtr*Lin28B CSD for ss pyrimidine-rich sequences, two potential CSD-binding sites within preE-*let-7f* were mutated: one in the terminal hairpin loop, harboring the r*let-7f* ^{a5} sequence (preE-mut I), and one in a bulge close to the upper stem region and adjacent to the Dicer cleavage site (preE-mut II). Both mutations impair CSD binding, as little complex band was visible in both cases (Figure 35). Consistent with this, these mutations cause a comparable loss of binding affinity in context with the WT protein, even though only a combination of both mutations (preE-mut III) had a similar effect as the GGAG mutation. Very recently, Nam and colleagues [227] showed that mLin28 CSD indeed binds to sequences within pre-*let-7* corresponding to preE-mut I. However, as preE-mut II also affected binding considerably, it is

likely that additional elements within preE have an influence on binding of Lin28 CSD and full-length Lin28.

A



B

Protein	RNA	Score
*Xtr*Lin28B	*pre-let-7f*	+++++
*Xtr*Lin28B ZKD	*pre-let-7f*	-
*Xtr*Lin28B ZKD	*pre-let-7f GGAG-mut*	-
*Xtr*Lin28B	*pre-let-7f GGAG-mut*	++
*Xtr*Lin28B CSD	*pre-let-7f*	+
*Xtr*Lin28B CSD	*pre-let-7f preE-mut I*	-
*Xtr*Lin28B CSD	*pre-let-7f preE-mut I*	-
*Xtr*Lin28B	*pre-let-7f preE-mut I*	+++
*Xtr*Lin28B	*pre-let-7f preE-mut II*	+++
*Xtr*Lin28B	*pre-let-7f preE-mut III*	++

C

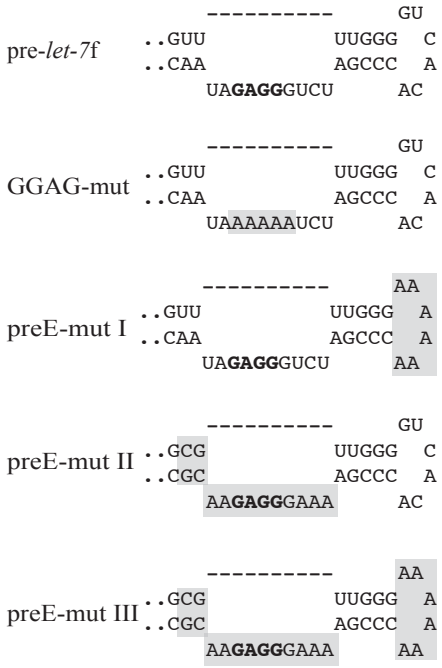


Figure 35: Determinants of Lin28:pre-*let-7f* interactions. (**A**) EMSAs with pre-*let-7f* as a probe, mixed with increasing concentrations of protein (for *Xtr*Lin28B and *Xtr*Lin28B CSD+ZKD: 0, 0.4, 0.8, 1.6, 3.2, 6.4, 13, 26 µM; for *Xtr*Lin28B CSD and ZKD: 0, 0.8, 1.6, 3.2, 6.4, 13, 26, 51 µM). All EMSAs were performed using 1 nM [α-^{32}P]-ATP labeled RNA. (**B**) Results from EMSA experiments shown in (A). The binding affinity was scored according to the following dissociation constant ranges: +++++, 0.8-1.6 µM, ++++ 1.6-3.2 µM, +++ 3.2-6.4 µM, ++ 6.4-12.8 µM, + 12.8-25.6 µM, - > 25.6 µM. (**C**) The terminal loop sequence and secondary structure of the pre-*let-7f* mutants as predicted by Mfold [211]. Regions that were mutated are shaded in gray.

3.6.2 *The Lin28 CSD remodels pre-let-7 miRNA*

The observation that the isolated CSD, but not the ZKD, binds to pre-*let-7f* suggests a multi-step binding reaction, in which the CSD first binds and imposes a structural change within preE, thereby facilitating subsequent binding of the ZKD to the conserved GGAG motif. Consistent with this hypothesis, the conserved GGAG motif is often not single-stranded according to multiple secondary-structure predictions programs and therefore a structural rearrangement of pre-*let-7* would be required for ZKD binding.

To confirm this hypothesis, an RNA remodeling assay with a truncated pre-*let-7g* was performed, which contains the entire pre-element and five bases of the ds upper stem region (defined as pre-*let-7g**). The blunt end of pre-*let-7g** was tagged with a fluorescence quencher pair to monitor potential RNA unwinding of the ds upper stem region upon protein binding (Figure 36 A). Both full-length *Xtr*Lin28 isoforms as well as *Xtr*Lin28B CSD show a clear increase in fluorescence and thus an RNA remodeling activity, while for the ZKD only a slight increase in fluorescence was observable (Figure 36 B). The increase in fluorescence is again cooperative for both the full-length proteins and the CSD and reflects their different binding affinities to pre-*let-7g* (see Figure 37). As observed before for pre-*let-7f*, WT *Xtr*Lin28B bound in the high nanomolar range to pre-*let-7g* ($K_D = 0.7 \pm 0.3$ µM), while the isolated CSD showed an almost 26-fold lower affinity ($K_D = 18.8 \pm 3.8$ µM).

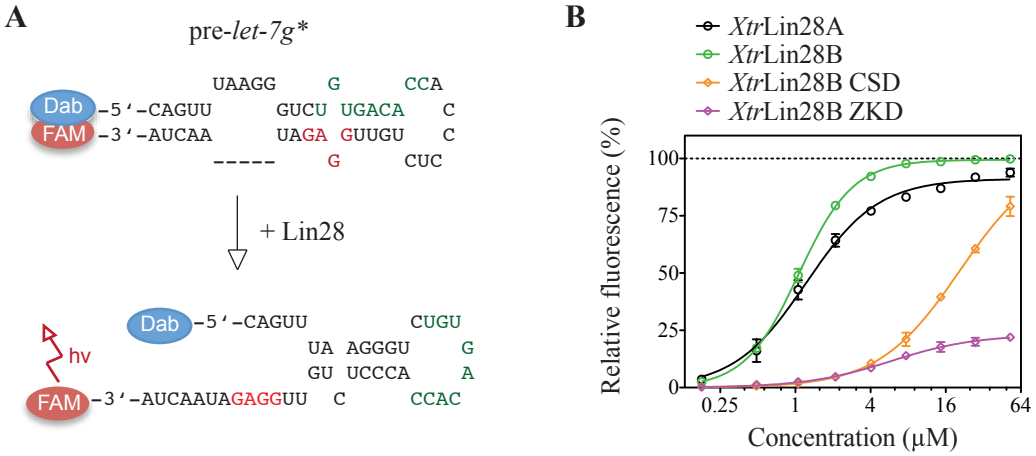


Figure 36: The Lin28 CSD remodels pre-*let-7* miRNA. (**A**) RNA remodeling assay. The indicated proteins were incubated with a truncated pre-*let-7g* that contains the entire pre-element and five base pairs of the upper stem region. The sequence and secondary structure of the used RNA (pre-*let-7g**) is indicated, the suspected ZKD and CSD binding sites are marked in red and green, respectively. The 5' end was modified with the quencher dabcyl (Dab), the adjacent 3' end harbored a fluorescein label (FAM). (**B**) The increase of FAM fluorescence was plotted as a function of titrated Lin28 protein. All experiments were performed using a Cary-Eclipse fluorescence spectrometer at 20 °C in 20 mM Tris (pH 7.5), 60 mM KCl, 10 mM $MgCl_2$ and 1 mM DTT.

The ability of the isolated Lin28 CSD to remodel RNA implies that mutations in the CSD impairing the observed RNA interactions should lead to a stronger

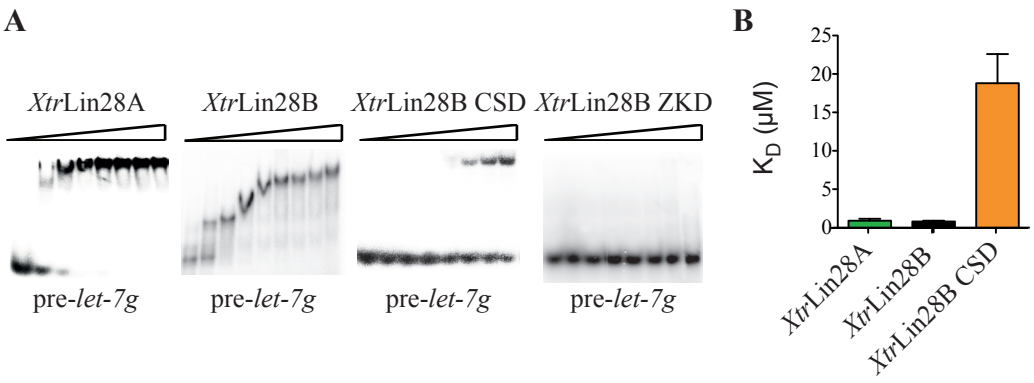


Figure 37: Binding of *Xtr*Lin28B to pre-*let-7g*. (**A**) EMSAs with pre-*let-7g* as a probe, mixed with increasing concentrations of protein (for *Xtr*Lin28B and *Xtr*Lin28B CSD+ZKD: 0, 0.4, 0.8, 1.6, 3.2, 6.4, 13, 26 µM; for *Xtr*Lin28B CSD and ZKD: 0, 0.8, 1.6, 3.2, 6.4, 13, 26, 51 µM). (**B**) Equilibrium dissociation constant from three independent EMSA experiments as shown in (A).

loss of remodeling activity than mutations in the Lin28 ZKD. Indeed, this is the case for all examined CSD mutations even though the K_D values of the analyzed variants with respect to pre-*let-7g* vary only slightly (Figure 38, Figure 39 A, B). Upon closer inspection the Lin28 mutations can be clustered into three classes. First, the ZKD variants *Xtr*Lin28B Tyr133Ala and His155Ala had a similar remodeling activity as the WT protein at low concentrations, but only reached up to 70% of the WT's activity at saturation. Second, variations in the CSD affecting binding subsites 5, 6 and 7 (*Xtr*Lin28B Phe66Ala, Phe48Ala and Trp39Ala) still showed modest remodeling activity that almost reached the level of the ZKD variants at high concentrations. Finally, *Xtr*Lin28B His68Ala (binding subsite 4), Phe77Ala and the double variant *Xtr*Lin28B Phe77Ala/Arg78Gly (binding subsite 1,2) displayed little remodeling activity.

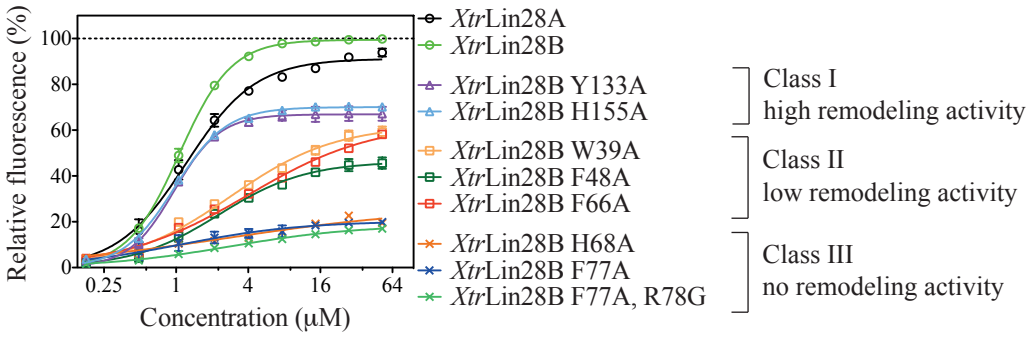


Figure 38: Conserved histidine and phenylalanine residues are indispensable for Lin28's remodeling activity. RNA remodeling assay with pre-*let-7g** (see Figure 36) and the indicated Lin28 variants were performed as described in Section 2.5.7. The increase of fluorescence was plotted as a function of titrated Lin28 protein.

These observations are highly interesting, as the recently observed CSD binding site within pre-*let-7g* [227] is not accessible under the predicted secondary structure by Mfold [211]. Consequently, these variants might bind in an alternative mode to pre-*let-7g* since in the corresponding EMSAs, all analyzed variants

still shifted pre-*let-7g* completely at high concentrations (Figure 39 A, B). Nevertheless, all variants were impaired in binding to pre-*let-7g* and consequently led to an increased pre-*let-7g* cleavage by Dicer (Figure 39 C).

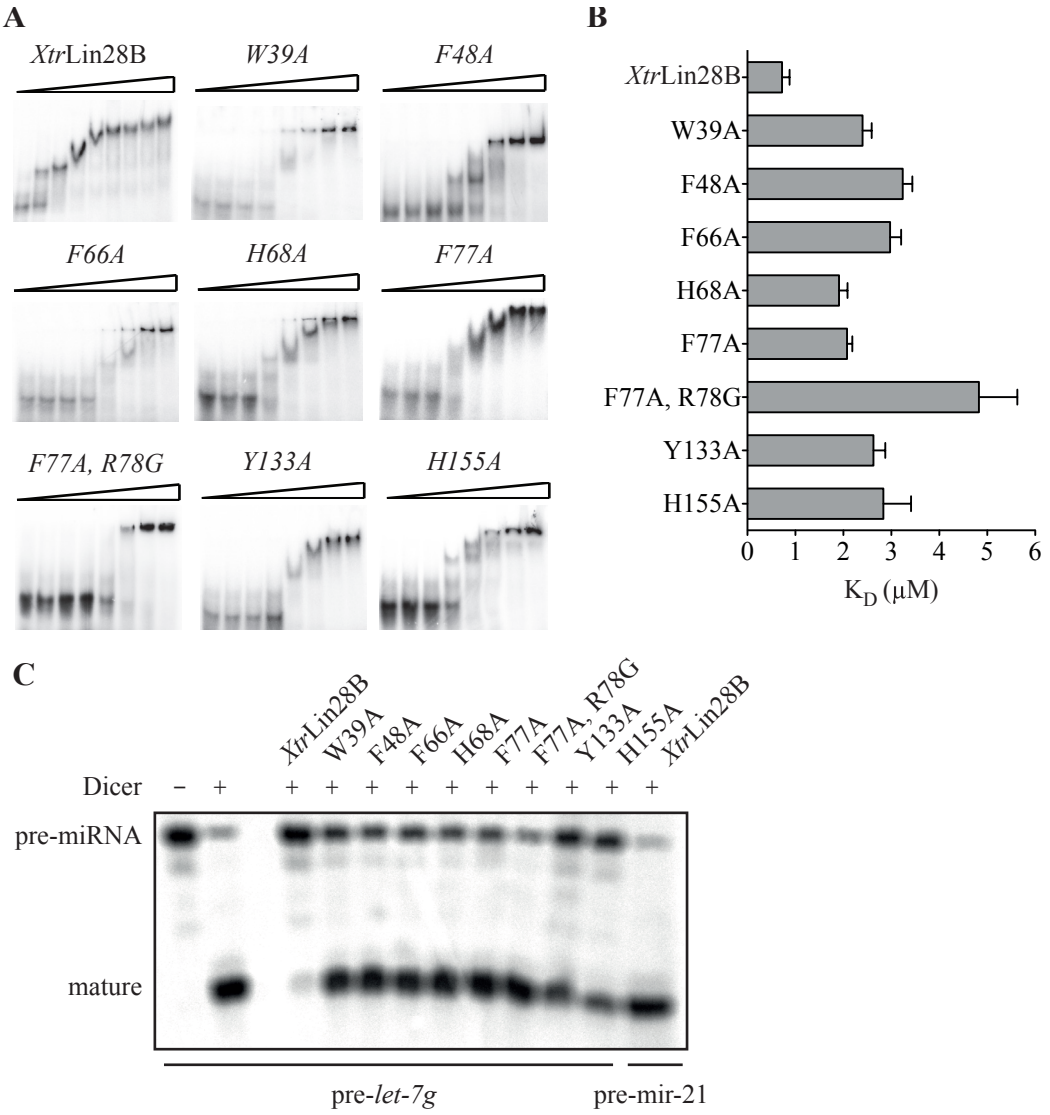


Figure 39: Binding affinity of Lin28 variants for pre-*let-7g* and their ability to inhibit pre-*let-7g* processing by Dicer. (**A**) EMSAs with pre-*let-7g* as a probe, mixed with increasing concentrations of *Xtr*Lin28B variants (0, 0.4, 0.8, 1.6, 3.2, 6.4, 13, 26 μM). The samples were resolved by 10% (w/v) native PAGE and visualized by autoradiography. (**B**) Equilibrium dissociation constant along with their standard deviations from three independent EMSA experiments as shown in (A). (**C**) *In vitro* pre-miRNA processing reaction on pre-*let-7g*. 10 μM of the indicated *Xtr*Lin28B variant were added to ^{32}P 5′ end-labeled pre-*let-7g* in the presence or absence of human Dicer. The samples were resolved by 10% (w/v) denaturing PAGE and visualized by autoradiography.

3.6.3 *Binding of Lin28 to pre-let-7 is a multi-step process*

To gain further insight into the observed remodeling reaction, remodeling kinetics of *Xtr*Lin28B•pre-*let-7g** were measured using stopped-flow mixing experiments. Mixing of 15 µM *Xtr*Lin28B with 100 nM pre-*let-7g** resulted in an association reaction with two phases. A major fast reaction making up 85% of total amplitude with a time constant of τ_{fast} = 20 ms is followed by a minor slow phase (τ_{slow} = 210 ms, Figure 40 A). Next, the concentration dependencies of the reaction rates were observed using constant pre-*let-7g** and increasing *Xtr*Lin28B concentrations (Figure 40 B). Interestingly, only the fast rate increased linearly with increasing protein concentrations with a slope of 1.15×10^6 $M^{-1}s^{-1}$ and a y-intercept of 35.24 ± 1.8 s^{-1} (Figure 40 C). As the second rate is concentration independent (Figure 40 D), the binding and remodeling reaction can interpreted by a two-step mechanism (Equation 7).

$$\text{Lin28} + \text{RNA} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} [\text{Lin28} \bullet \text{RNA}]^* \xrightarrow{k_2} \text{Lin28} \bullet \text{RNA} \tag{7}$$

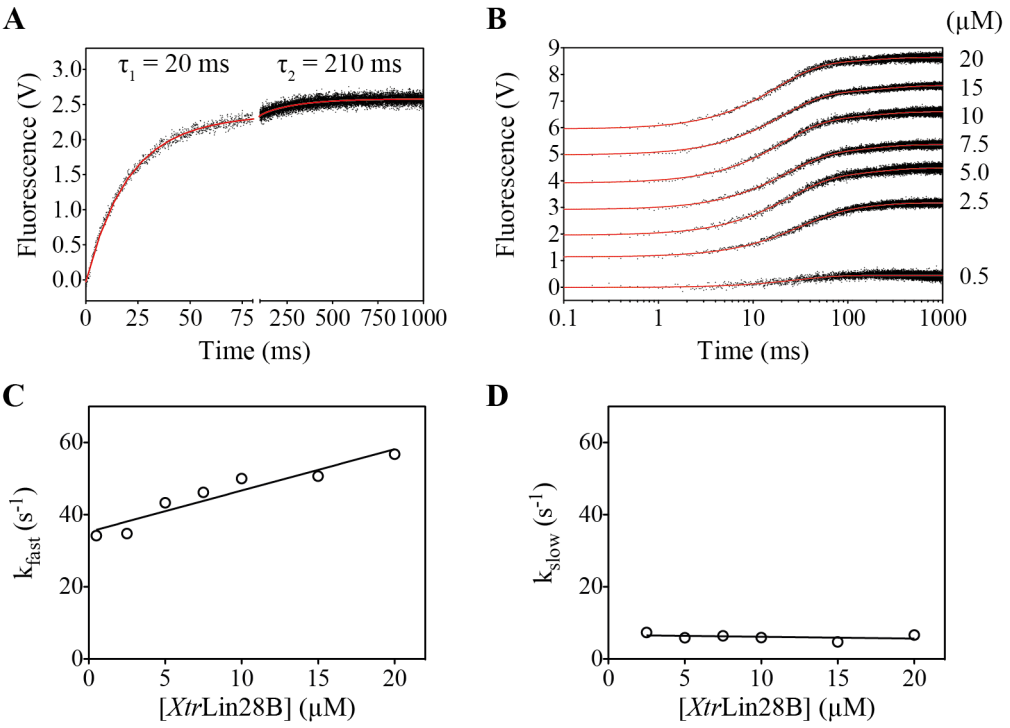


Figure 40: Binding of pre-*let-7g** by *Xtr*Lin28B can be interpreted by a two-step mechanism. (**A**) After rapid mixing of 100 nM pre-*let-7g** and 15 µM *Xtr*Lin28B, the change of FAM-fluorescence was monitored for 1 s. Kinetic traces were fit to a bi-exponential equation with the following time constants: τ_1 = 20 ms, τ_2 = 210 ms (**B**) The kinetic traces of pre-*let-7g** remodeling were measured at increasing *Xtr*Lin28B concentrations. Individual curves were translated along the y-axis. (**C**) The fast rates (k_1) increased linearly with increasing [*Xtr*Lin28B] and revealed a slope (= k_1) of $1.15 \pm 0.17 \times 10^6$ $M^{-1}s^{-1}$ and a y-intercept (= k_{-1}) of 35.24 ± 1.8 s^{-1}. (**D**) The slower rate (k_{slow} = k_2) was concentration independent with a mean value of 6.6 ± 0.73 s^{-1}.

Next, stopped-flow mixing experiments with *Xtr*Lin28B variants of each class and the isolated CSD under saturation conditions were performed (Figure 41). The time course of the *Xtr*Lin28B Tyr133Ala remodeling activity could be fit to either a mono-exponential equation (τ = 28 ms) or a bi-exponential equation (τ_{fast} = 26 ms, τ_{slow} = 510 ms) with a very weak second time constant making up only 3.7% of the total amplitude (Figure 42 D). This indicates that the second rate observed in the WT can be traced back to the binding of the ZKD. Considering the flexible linker between both RBDs and the high structural difference between apo Lin28 ZKD and RNA-bound Lin28 ZKD [227, 228], the slow second time constant may reflect the necessary structural rearrangement in these regions to facilitate binding. The isolated CSD showed a one-phase association reaction with a similar time constant as observed for *Xtr*Lin28B Tyr133Ala (τ = 29 ms), even though the amplitude was much lower (Figure 42 B). For *Xtr*Lin28B Trp39Ala, after a fast association reaction (τ_{fast} = 15 ms) that was completed in 80 ms, a slow dissociation reaction was observed (τ_{slow} = 110 ms) (Figure 42 C). The decrease in fluorescence might reflect impaired binding of this construct and thus a higher dissociation rate compared to WT protein, as the fast remodeling rate was not impaired. For *Xtr*Lin28B Phe77Ala the remodeling activity was almost completely impaired and thus the data could not be fit (Figure 42 E). This supports our previous observations that Phe77 and His68 are crucial for the remodeling reaction.

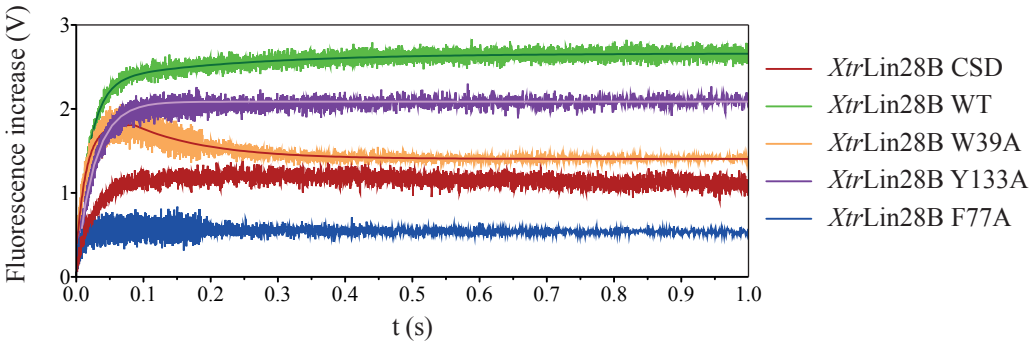


Figure 41: Time courses of pre-*let-7g** binding and remodeling by different Lin28 variants. After rapid mixing of 100 µM pre-*let-7g** with 15 µM (final concentrations) of the indicated *Xtr*Lin28B variant, changes of FAM-fluorescence were monitored for 1 s using a Chirascan stopped-flow instrument. Traces of at least 12 replicates were fit to a mono- (Y133A, CSD) or bi-exponential curve (*Xtr*Lin28B), respectively (solid lines). For W39A, the time course of the first 80 ms were fit to a single-exponential association curve, while traces from 90 ms to 1 s were fit to a one-phase decay curve. Data about the individuals fits of each curve including residuals are depicted in (Figure 42).

Taken together, the data clearly demonstrate that the *Xtr*Lin28B CSD imposes a structural change within preE of pre-*let-7* that facilitates a subsequent binding of the ZKD. However, as the isolated CSD is unable to block the processing of pre-*let-7g* alone, the ZKD probably serves as an anchor for directional binding and thus ensures a constant opening of the Dicer cleavage site.

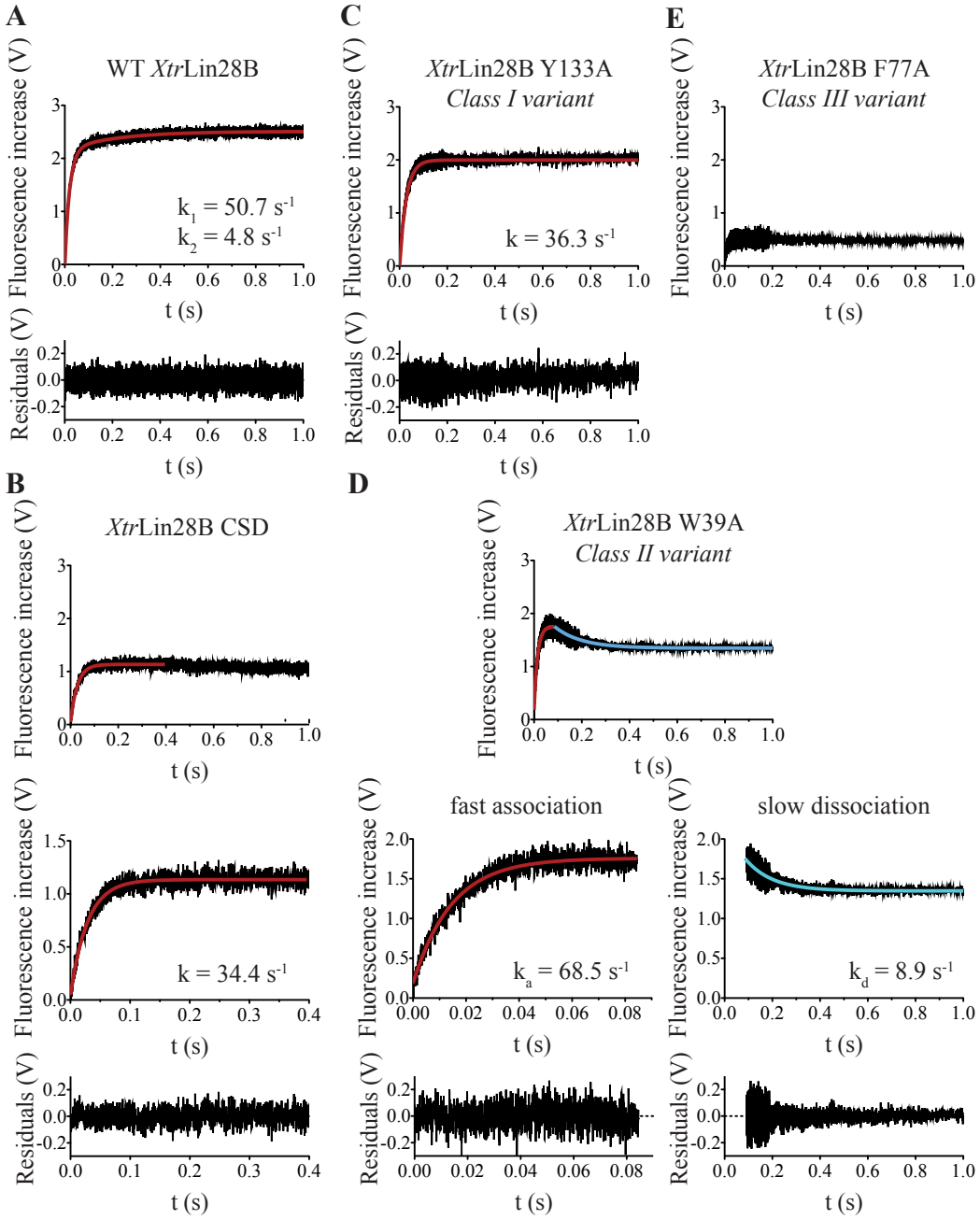


Figure 42: Pre-steady state kinetics of pre-*let-7g** binding and remodeling by different *Xtr*Lin28B variants. Time courses were recorded after rapid mixing of 15 µM of the indicated *Xtr*Lin28B variant and 100 nM pre-*let-7g** (final concentrations) and either fit to a mono- or bi-exponential binding equation. The following parameter were obtained. (**A**) *Xtr*Lin28B: two phases; A_{fast} = -2.1 V, k_{fast} = 50.7 s^{-1}, A_{slow} = -0.4 V, k_{slow} = 4.8 s^{-1}. (**B**) *Xtr*Lin28B CSD: one phase; A = -1.1 V, k = 34.4 s^{-1}. (**C**) *Xtr*Lin28B Y133A: one phase; A = -2.0 V, k = 36.3 s^{-1}. (**D**) *Xtr*Lin28B W39A: two phases; data of the first 80 ms were used to fit the fast association reaction, data from 90 ms to 1 s were used to fit the slower dissociation reaction. A_{fast} = -1.8 V, k_{fast} = 68.5 s^{-1}, A_{slow} = 0.8 V, k_{slow} = 8.9 s^{-1}. (**E**) *Xtr*Lin28B F77A: no fit possible. Residuals for each fit are shown below each plot.

3.7 LIN28 PROMOTED URIDYLATION OF PRE-*let-7* MIRNA

In addition to directly inhibiting pre-*let-7* maturation, Lin28 is known to promote 3′ end uridylation of pre-*let-7* leading to degradation of the miRNA by cellular nucleases [131]. Recent studies identified the terminal uridyl transferase TUT4 (also called Zcchc11 in humans) and the non-canonical poly(A) polymerase PUP2 in *C. elegans* as enzymes that uridylate pre-*let-7* in a Lin28-dependent fashion [132, 163]. To elucidate the molecular basis of Lin28 mediated uridylation of pre-*let-7* via TUT4/PUP2, the protein determinants for these interactions were examined.

3.7.1 *PUP2•Lin28 interactions and uridylation of pre-let-7*

3.7.1.1 *PUP2 purification*

To analyze PUP2•*Cel*Lin28 interactions *in vitro*, recombinantely expressed PUP2 was purified. Therefore, PUP2 was PCR-amplified from cDNA of *C. elegans* and cloned into the bacterial expression vector pDEST15. After sequence verification, plasmids were transformed into *E. coli* Rosetta 2 (DE3) and the gene was expressed as described in (Section 2.3.2). For purification, *E. coli* lysate containing GST-PUP2 was loaded onto a 5 ml GSH-Sepharose, washed extensively with GST wash buffer and cleaved on-column by incubation with 5 mg TEV protease overnight. The presence of 5 mM $MgCl_2$ and 1 mM ATP in the wash buffer was necessary to remove *E. coli* chaperones like Hsp60 and Hsp70 that otherwise co-eluted with PUP2. After SEC using a Superdex 200 column that was connected to a 5 ml Ni-NTA column to remove hexahistidine-tagged TEV protease, almost homogeneous PUP2 was obtained (Figure 43).

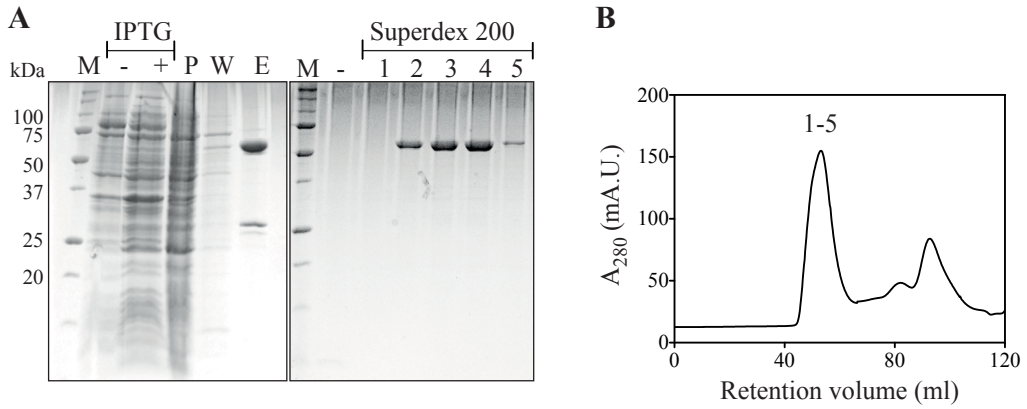


Figure 43: Purification of PUP2. (**A**) 15% SDS-PAGE of various samples taken during the purification of PUP2. -/+ IPTG: whole-cell bacterial lysates before and after induction; P: insoluble protein, W: washing fractions after application of soluble protein on GSH-affinity column, E: eluate after on-column cleavage with 5 mg TEV protease overnight, M: Molecular weight standard, 1-5: Superdex 200 fractions of the indicated peak in (B). (**B**) Superdex 200 SEC chromatogram. The indicated fractions were pooled and concentrated for further experiments.

3.7.1.2 *PUP2 mediate uridylation of pre-let-7*

Purified PUP2 appeared folded and exhibited the signature of a mixed α-helical and β-sheet containing protein as judged by CD spectroscopy (Figure 44 A, Table 15). Furthermore, PUP2 was enzymatically active and could uridylate the 3' end of *Cel*pre-*let-7 in vitro* even in the absence of Lin28. However, upon addition of recombinant *Cel*Lin28 to the reaction mix, PUP2-mediated uridylation of *Cel*pre-*let-7* was strongly increased, indicating that Lin28 promotes uridylation of *Cel*pre-*let-7* (Figure 44 B).

Table 15: Secondary-structure analysis of PUP2 CD data.

	α-helix (%)	β-sheet (%)	β-turn (%)	random coil (%)	sum (%)
PUP2	27	25	19	28	99

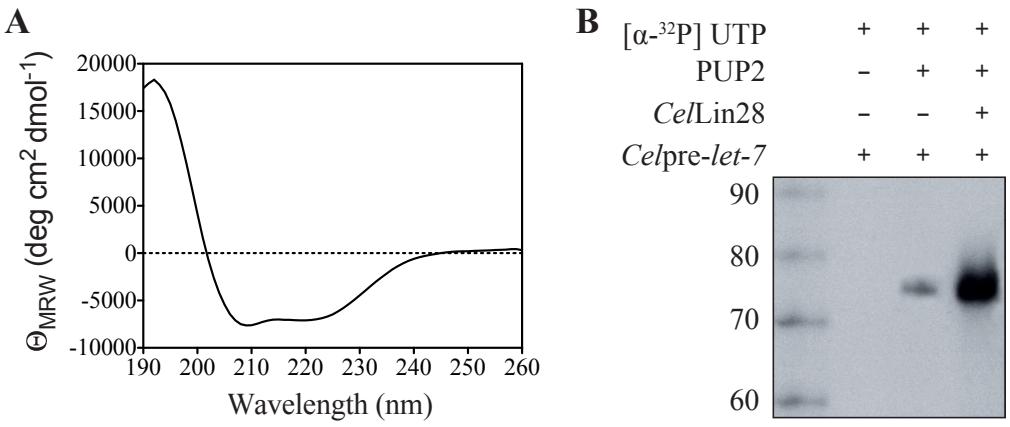


Figure 44: Lin28 promotes PUP2-mediated uridylation of pre-*let-7* miRNA. (**A**) Circular dichroism spectra of recombinantly purified PUP2. CD data were used to predict the secondary-structure content shown in Table 15. (**B**) *In vitro* uridylation of *Cel*pre-*let-7* by PUP2. *In vitro* transcribed *Cel*pre-*let-7* miRNA was incubated with PUP2 and [α—^{32}P] UTP in the presence and absence of *Cel*Lin28 as described in Section 2.5.9. All samples were resolved by 10% (w/v) TBE-urea PAGE and visualized by autoradiography.

3.7.1.3 *PUP2•Lin28 interactions*

Despite its stimulating effect on pre-*let-7* uridylation, no direct interaction could be measured between *Cel*Lin28 and PUP2 using ITC (Figure 45 A). However, as no *Cel*pre-*let-7* miRNA was present in the ITC experiment, additional LUMIER assays were performed as described in Section 2.7.3. Surprisingly, no interaction was observable between *Cel*Lin28 and PUP2 (Figure 45 B) indicating that PUP2 interacts only transiently with *Cel*Lin28•pre-*let-7*.

Taken together, *Cel*Lin28 promotes PUP2 mediated uridylation of pre-*let-7*. However, no stable and direct interaction between *Cel*Lin28 and PUP2 was observable indicating that both proteins interact transiently.

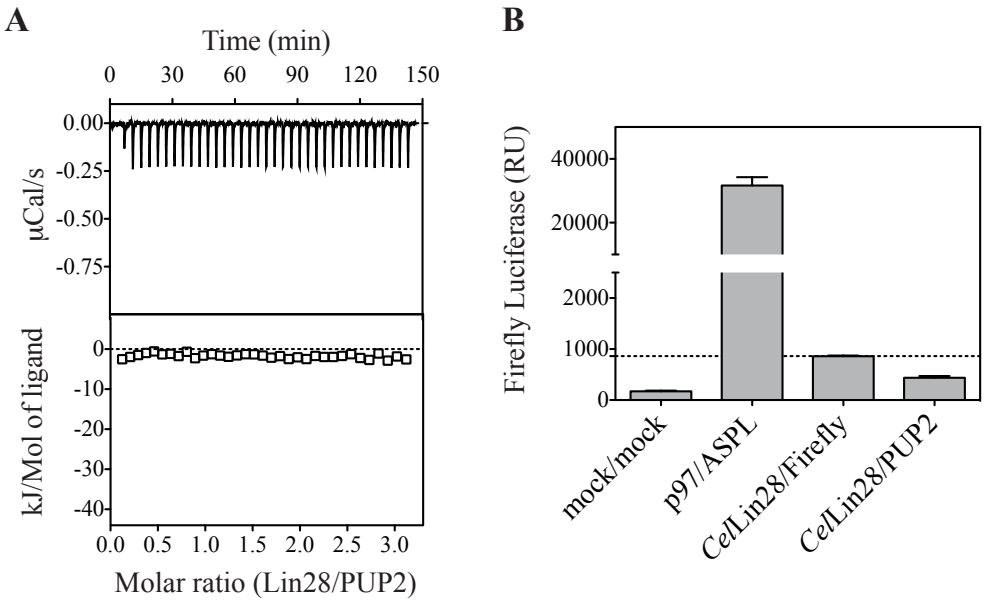


Figure 45: Lin28 and PUP2 show no stable interaction. (**A**) ITC experiment performed with 20 µM PUP2 in the cell and 300 µM *Cel*Lin28 in the syringe at 15 °C. (**B**) LUMIER assay. Plasmids encoding PA-RL-tagged *Cel*Lin28, and V5-FL-tagged PUP2 were transfected into HEK293 cells. After cell lysis, PA-RL-*Cel*Lin28 was immunoprecipitated using anti-Protein A antibodies. After washing, the presence of PA-RL-*Cel*Lin28 and V5-FL-PUP2 was selectively analyzed. PA-RL-p97/V5-FL-ASPL served as positive, PA-RL-*Cel*Lin28/V5-FL as negative control.

3.7.2 *TUT4•hLin28 interactions and uridylation of pre-let-7*

3.7.2.1 *The TUT4 CCHC Zn-knuckles II and III are essential for hLin28-dependent uridylation or pre-let-7*

In mammals the terminal-uridyl transferase TUT4 (also called Zcchc11) is known to mediate the uridylation of pre-*let-7* miRNA in a hLin28-dependent fashion [131, 132]. To investigate the protein and RNA determinants for this process, multiple protein variants of both hLin28 and TUT4 were created and analzyed with respect to their uridylation and interaction capability *in vitro* and *in vivo* (Figure 46).

As His- and GST-tagged TUT4 proteins were not soluble after recombinant expression in *E. coli*, expression plasmids encoding Flag-TUT4 were transfected into HEK293 cells and immunopurified using anti-Flag beads, eluted with Flag peptide and verified by Western blot (Figure 47 A). Immunopurified Flag-TUT4 was enzymatically active and showed a strong uridylation of pre-*let-7g* only in the presence of recombinantly purified hLin28 *in vitro*. A deletion of the C-terminal 264 residues of TUT4 (TUT4 ΔC) had little effect on the enzymatic activity. In contrast, an additional deletion of CCHC Zn-knuckles II and III (also referred to as TUT4 ZKD) completely impaired pre-*let-7* uridylation. A mutation of the Zn-chelating residues, C1298A and C1362A (TUT4 ZKD-mut) in context of the full-length protein had a similar effect as the deletion of the

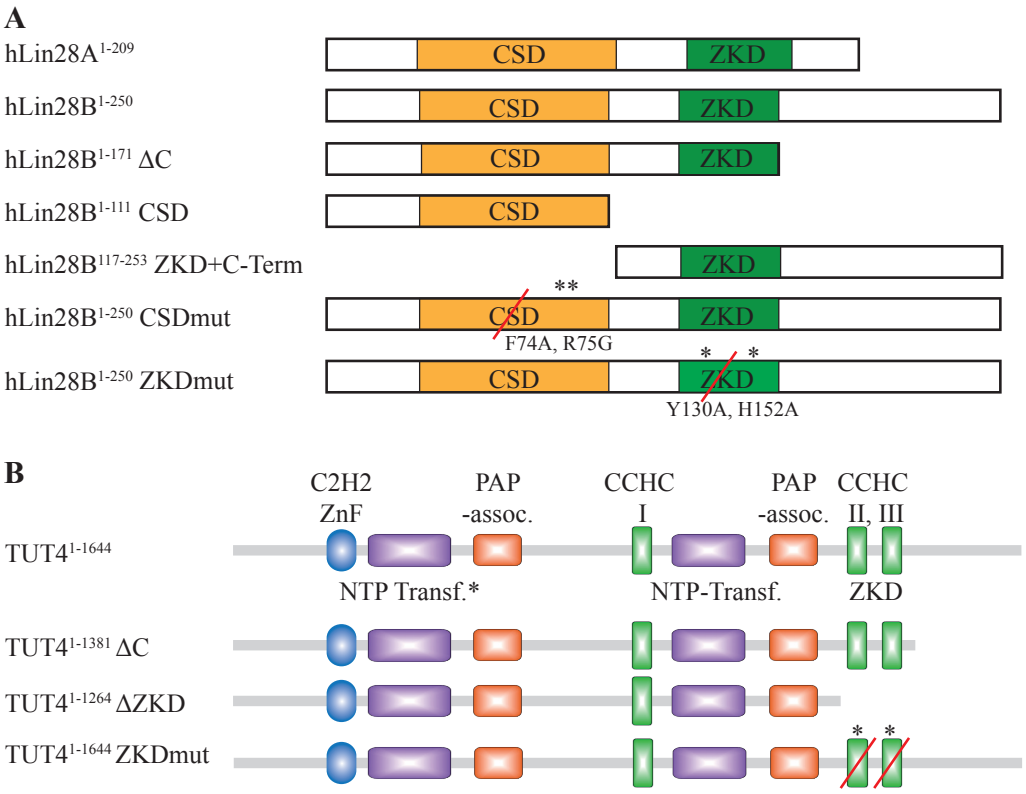


Figure 46: Schematic representation of human hLin28 (**A**) and TUT4 (**B**) constructs used for *in vitro* uridylation and interaction studies. The domain architecture including catalytically active domains, amino-acid ranges and point variations are indicated. NTP Transf.*: non-functional nucleotidyl-transferase domain, in which an essential Asp is mutated in the active site; PAP assoc.: Poly(A) polymerase associated domain; CCHC I, II, III: CCHC-type Zn-knuckles - CCHC Znk II and III are in close proximity to each other and probably have a similar function as the ZKD in Lin28.

TUT4 ZKD (TUT4 ΔZKD) indicating that these CCHC Zn-knuckles are essential for pre-*let-7g* uridylation (Figure 47 B).

To check whether hLin28 recruits TUT4 to pre-*let-7g* via direct protein-protein interactions and thereby promotes pre-*let-7* uridylation, co-immunoprecipitation experiments were performed. Plasmids encoding Flag-tagged TUT4 and V5-tagged hLin28B were transfected into HEK293 cells and immunoprecipitated after cell lysis using anti-Flag antibodies. WT TUT4 and TUT4 ΔC could precipitate hLin28B even though the precipitation was much stronger upon addition of *in vitro* transcribed pre-*let-7g* to the extract. The deletion variant TUT4 ΔZKD as well as TUT4 ZKD-mut could not mediate the interaction (Figure 47 C). Thus, TUT4 requires a functional ZKD (CCHC Zn-knuckles II and III) and pre-*let-7* to interact with hLin28.

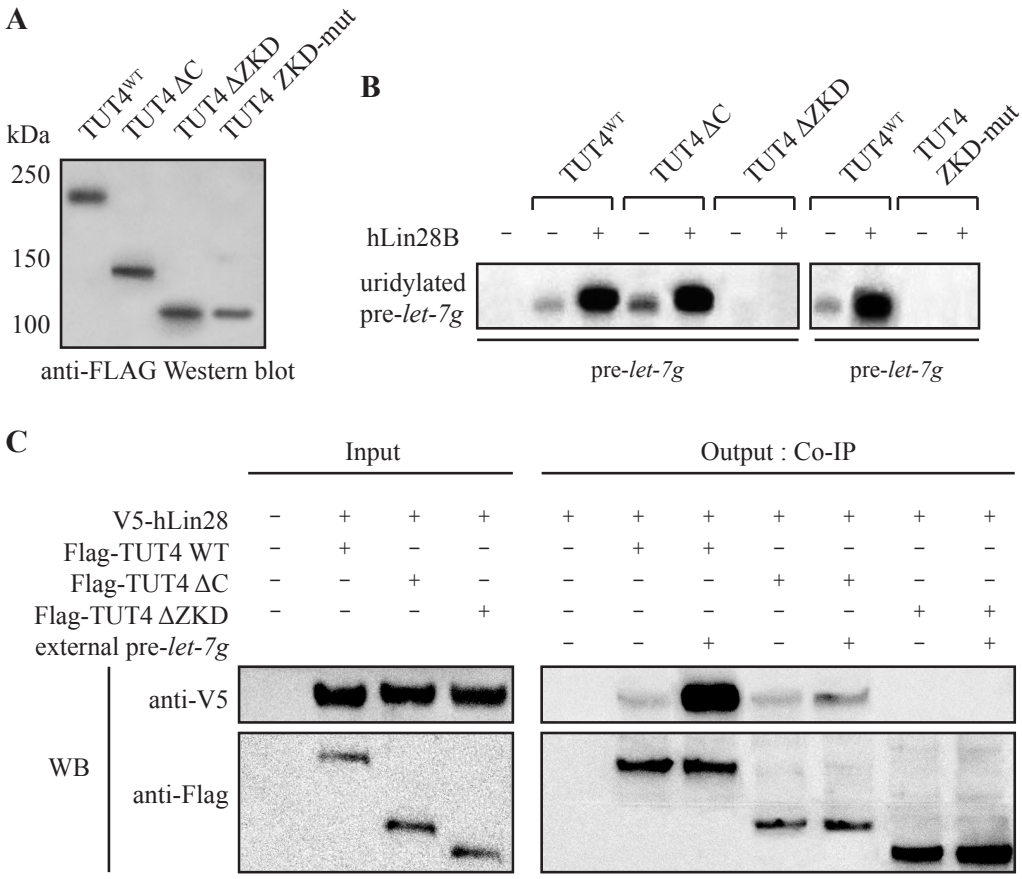


Figure 47: The TUT4 CCHC Zn-knuckles II and III are essential for hLin28B-dependent uridylation of pre-*let-7*. (**A**) Anti-Flag western blots of immunoprecipitated Flag-TUT4 variants after expression in HEK293 cells (**B**) *In vitro* uridylation assay with *in vitro* transcribed pre-*let-7g* and immunopurified TUT4 variants. (**C**) Co-immunoprecipitations of hLin28B and TUT4 in the presence and absence of pre-*let-7g*. V5-hLin28B and Flag-TUT4 were co-expressed in HEK293 cells (Input). After cell lysis, hLin28B and TUT4 were co-immunoprecipitated using anti-Flag agarose beads. To check, whether the interaction depends on the presence of pre-*let-7* miRNA, 1 µg of *in vitro* transcribed pre-*let-7* miRNA was added to individual HEK293 cell extracts prior to co-immunoprecipitation.

3.7.2.2 *hLin28 domain requirements for TUT4 mediated pre-let-7 uridylation*

Next, it was determined which parts of hLin28 are necessary for TUT4 mediated uridylation of pre-*let-7*. Both hLin28 paralogs, hLin28A and hLin28B, strongly stimulated uridylation of pre-*let-7*. Furthermore, uridylation activity was dependent on the conserved GGAG motif within pre-*let-7*, as a GGAG-to-AAAA mutation completely abolished the activity (Figure 48). A deletion of hLin28B's C-terminus impaired uridylation considerably. Moreover, constructs that lost either the CSD (hLin28B ZKD+C-Term) or the ZKD including C-terminus (hLin28B CSD) failed to enhance the enzymatic activity of TUT4. Consistent with this, in presence of the structure-based point variants hLin28B CSD-mut (Phe74Ala, Arg75Gly) and hLin28B ZKD-mut (Tyr130Ala, His152Ala), TUT4 showed a reduced activity even though the level of uridylation was higher

than for the hLin28B deletion variants. Thus the C-terminus and both RBDs are essential to enhance pre-*let-7* uridylation by TUT4.

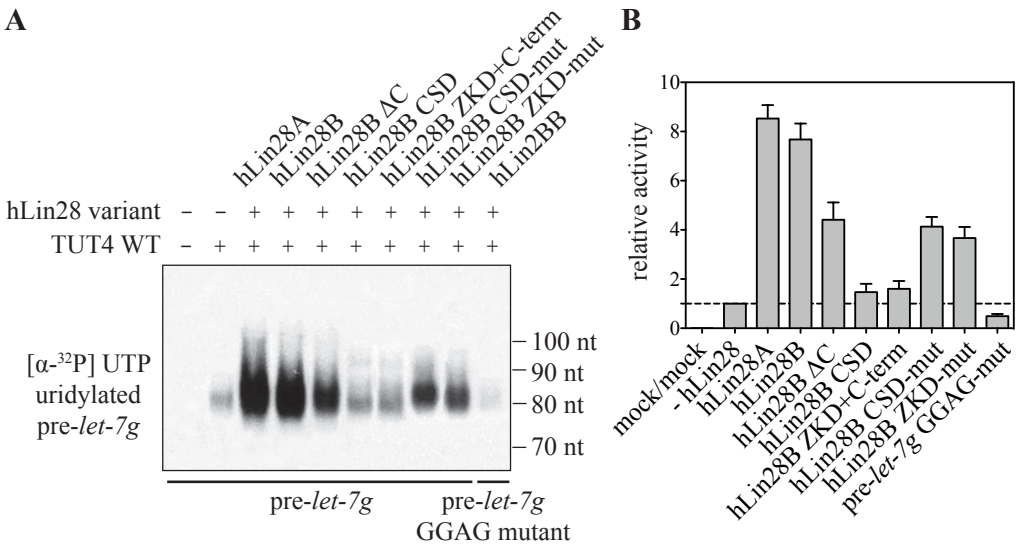


Figure 48: Both hLin28 RNA-binding domains as well as the C-terminus are required for an efficient TUT4 mediated uridylation of pre-*let-7* miRNA. (**A**) *In vitro* uridylation of pre-*let-7g* by TUT4. The uridylation activity of immunopurified Flag-TUT4 was measured using [α^{32}P] UTP and *in vitro* transcribed pre-*let-7g* in the presence and absence of the indicated hLin28 constructs. The reaction products were resolved by 10% (w/v) urea-PAGE and visualized by autoradiography. Band intensities were analyzed from three independent experiments using ImageJ64 und used to generate the relative activities illustrated in (**B**).

Using a LUMIER assay (Section 2.7.3), the hLin28 domain requirements for direct TUT4 interactions were quantitatively analyzed. Therefore PA-RL-tagged hLin28 and V5-FL-tagged TUT4 variants (see Section 3.7.2) were transfected into HEK293 cells. After cell lysis in the presence of RNase inhibitor, PA-RL-hLin28 variants were immunoprecipated using anti-PA antibody, washed and checked for the presence of the corresponding TUT4 variant.

Both hLin28 isoforms show a clear interaction with TUT4 with roughly a sevenfold higher FL signal than the negative control (hLin28A vs FL, Figure 49). A deletion of the C-terminus or the CSD caused a decreased interaction capability (hLin28B ΔC), whereas a loss of the ZKD+C-terminus impaired the interaction completely (hLin28B ZKD+C-Term). This underscores the importance of an intact hLin28B C-terminus and ZKD for TUT4 interaction. Moreover, pre-*let-7* is critical for the interaction, as variations disturbing the RNA binding of hLin28 (hLin28B CSD-mut, hLin28 ZKD-mut) showed roughly a threefold lower interaction compared to WT hLin28. On the TUT4 level, deletion of CCHC fingers II and III almost completely abolished the hLin28•TUT4 interaction. This supports our previous observation and thus raises the question whether TUT4 ZKD itself contacts the conserved GGAG motif of pre-*let-7*. Taken together, the data imply that it requires both hLin28 RBDs including its C-terminus to recruit TUT4 to pre-*let-7* and enhance the 3'-uridylation activity.

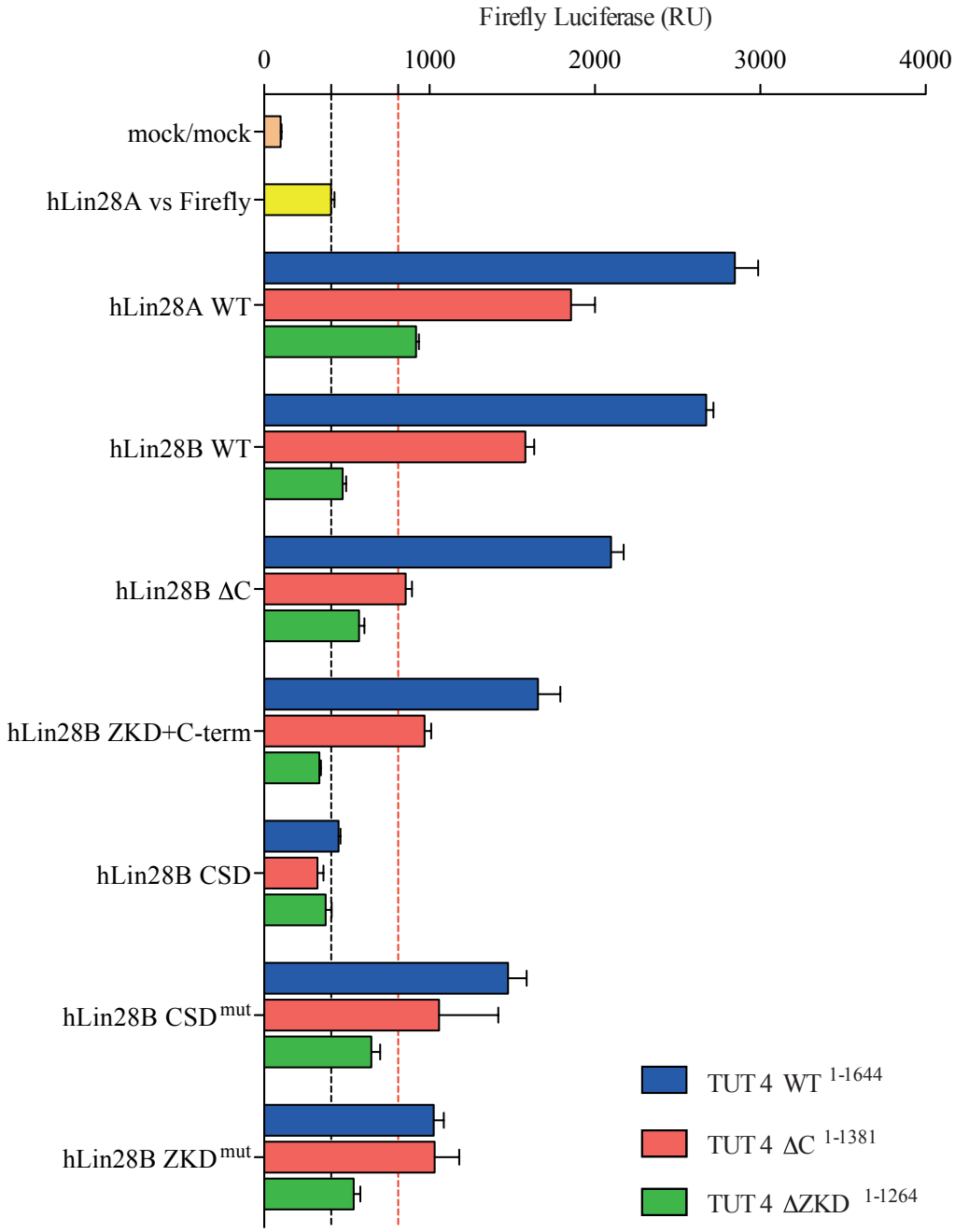


Figure 49: Minimal domain requirements for hLin28•TUT4 interaction. Plasmids encoding PA-RL-tagged hLin28 and V5-FL-tagged TUT4 constructs were transfected into HEK293 cells. After cell lysis, RL-hLin28 constructs were immunoprecipated using anti-PA antibodies. After washing, the presence of PA-RL-hLin28 and V5-FL-TUT4 proteins were selectively analyzed. PA-RL-hLin28/V5-FL served as a negative control (background). The dashed red line indicates a two-fold higher signal than background.

4 DISCUSSION

4.1 THE LIN28 CSD CAN BIND A WIDE RANGE OF SINGLE-STRANDED RNA/DNA SEQUENCES

The structure of *Xtr*Lin28B CSD in complex with ssDNA and ssRNA oligonucleotides, in concert with a systematic binding analysis, has provided valuable information about its function in pre-*let-7* miRNA and mRNA binding. Lin28 CSDs bind with high affinity to ss nucleic acids via a conserved nucleic acid-binding platform mainly formed of exposed aromatic side chains that are surrounded by basic surface residues. This binding platform is already preformed in the apo protein and, consequently, only subtle changes are observed upon nucleic-acid binding. The crystal structure of the *Xtr*Lin28B CSD bound to dT_7 revealed binding of up to seven nucleotides. This observation is in good agreement with the binding data, since the CSD exhibited the highest affinity for 7- to 9-mers. Such binding preferences have also been described for bacterial Csps [180, 182, 229, 181, 234, 178, 179], and recently also for mLin28A in complex with preE-derived RNAs [227]. In the mLin28•RNA structures, a closed stem loop within the RNA, as well as elements of the basic linker between CSD and ZKD, triggered the binding of an additional eighth and ninth nucleotide. However, despite their different sequence and chemical properties, dT_7/dT_6 as well as preE-derived RNAs exhibited a remarkably similar binding mode to Lin28 CSDs. The nucleic acids bind in an almost identical curved conformation to the hydrophobic platform and share the same binding subsites (Figure 50 A).

With regard to sequence specificity, my structural data revealed sequence-specific binding at positions 2 and 6 and partially at position 5. At these binding sites, T/U-specific hydrogen bonds are mediated between *Xtr*Lin28B CSD and each base. Most notably the presence of a salt bridge between Lys38 and Asp64 limits the flexibility and consequently the size of the binding pocket at position 6. This salt bridge is structurally conserved in most CSDs and establishes, in concert with a neighboring Trp (Trp39), specific hydrogen bonds to O^2 and HN^3 of the T/U base. Comparison between bacterial CSD•ssDNA/RNA and Lin28 CSD•ssDNA/RNA structures revealed an almost identical orientation of the bound T/U, thereby underscoring the clear preference for T/U at this binding subsite (Figure 50 B).

In contrast, binding subsites 2 and 5 can also accommodate purines as observed in recent mLin28A•RNA structures [227]. Since Lin28 CSD mediates hardly any contacts with the sugar-phosphate backbone, the DNA/RNA backbone can adopt different conformations to optimize binding. Thus Lin28 CSD can bind a wide range of different ssDNA/ssRNA sequences without disrupting the observed hydrogen bonds. For example, in the mLin28A•preE-*let-7f** structure, the sugar-phosphate backbone moves away from the protein thereby enabling a larger G to bind into the pocket of binding subsite 2 without disrupting hydrogen bonds (Figure 50 C). This flexibility is also observed at the other

binding sites and reflects the dominance of stacking interactions in the binding interface.

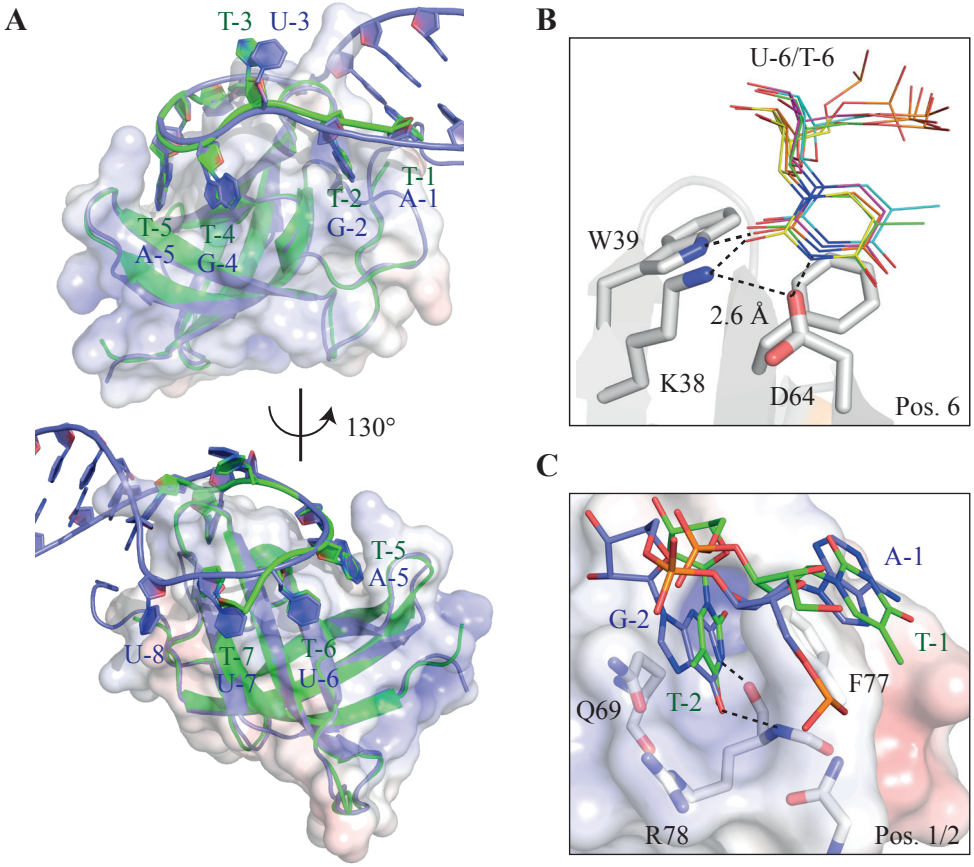


Figure 50: The Lin28 CSD can bind to a wide range of different RNA sequences. (**A**) Superimposition of *Xtr*Lin28B•dT_7 (green) and mLin28A•preE-*let-7f** (blue) (PDB-ID 3TSO). The protein-nucleic acid interaction surface is similar for binding subsites 1 to 7. Binding of the additional nucleotide (U-8) in the mLin28A•preE-*let-7f** structure is triggered by the formation of a closed RNA loop. Base-stacking interactions are dominant in both structures and allow high-affinity binding of a wide range of ssDNA/ssRNA oligonucleotides. (**B**) Superimposition of bound nucleotides at binding subsite 6 derived from various bacterial and Lin28 CSDs in complex with ssDNA/ssRNA. All structures contained a T/U nucleotide at binding subsite 6. The salt bridge between a structurally conserved Lys and Asp (Lys38, Asp64) limits the size of the pocket and establishes specific hydrogen bonds with the T/U base. (**C**) Since few interactions are formed with the sugar phosphate backbone, the bound oligonucleotide can adopt different backbone conformations to optimize the binding with Lin28 CSD. For example, at binding subsite 2 the sugar phosphate backbone of mLin28A•preE-*let-7f** is farther displaced from the protein, thereby enabling binding of G (G-2) instead of T (T-2) without disrupting hydrogen bonds.

Based on the *in vitro* binding experiments, *Xtr*Lin28B CSD resembled its bacterial homologs in exhibiting a slight preference for pyrimidine-rich sequences, even though the overall specificity is rather low. The lower affinity towards ssDNA oligonucleotides compared to bacterial Csps is likely due to the pres-

ence of Gln69 instead of a Phe that would otherwise contribute to base-stacking in bacterial Csps. Except for the polyuridines rU_6 and rU_7, no substantial difference in binding between ssRNA and ssDNA oligonucleotides was detected. In general, high binding affinities for oligonucleotides containing $G^1UNNUNN^7$ or $G^1TNNTNN^7$ motifs were observed. These motifs are slightly different to the recently proposed CSD-binding motif ($N^1GNGAYN^7NN$, [227]) especially with respect to the corresponding protein binding subsites. The discrepancy might be due to the use of small oligonucleotides in the binding experiments that might shift in register. However, based on my structural and biochemical studies, a three-membered stack between the first base, Phe77 and the second base seems to be more favorable than the interaction at the eighth binding site. Moreover, in the *Xtr*Lin28B CSD•rU_6 structure, U was bound exclusively at binding subsite 2, indicating that this particular site has a high affinity towards this base. Therefore, it is likely that oligonucleotides used for solution binding experiments in this study, indeed bound as observed in *Xtr*Lin28B CSD•dT_7.

4.2 THE LIN28 ZKD GOVERNS SPECIFICITY FOR *let-7* PRECURSORS

The broad spectrum of potential CSD binding sequences probably reflects the low sequence conservation within its natural substrate, the preE of pre-*let-7* (see Figure 8). Sequence-specific binding of Lin28 is mediated through the ZKD, since the isolated ZKD displayed a clear preference for GGAG- or GGUG-containing oligonucleotides. These motifs are conserved within preE's of pre-*let-7* (see Figure 8) and thus are responsible for specific binding of Lin28 to *let-7* precursors.

Based on the ITC experiments, the *Xtr*Lin28B ZKD specifically recognizes ssDNA and ssRNA oligonucleotides containing a GGHG motif, with H being any base except G. Oligonucleotides containing four G's in a row are known to from stable quadruplex structures [235], possibly explaining why *Xtr*Lin28B ZKD did not bind these RNAs. Further mutagenesis experiments revealed that at least two G's spaced by one or two bases are necessary for binding. These observations are congruent with NMR solution structures of HIV-1 NC ZKDs, in which the 2^{nd} and 4^{th} G of a GGAG or a GGUG tetraloop was specifically bound in hydrophobic pockets of the second and first Zn-knuckle, respectively [191, 192]. Recent structural studies with mLin28A and hLin28A ZKDs revealed similar interactions with GGAG-containing RNAs (Figure 51). Extensive hydrogen bonding is mediated with all three G's. Only the 1^{st} and 4^{th} nucleotide are bound in a hydrophobic pocket formed by His162/Met170 in the first and His148/Tyr140 in the second Zn-knuckle. His162 and Tyr140 are critical residues that contact each other and stack with the bases, thereby establishing a kinked conformation in the RNA. Consequently, mutation of the corresponding His and Tyr in *Xtr*Lin28B CSD (Tyr133Ala,His155Ala) impaired binding considerably (Tyr133Ala: 48-fold, His155Ala: 57-fold, see Table 5). Since these residues are not proximal in apo hLin28A ZKD (PDB-ID 2CQF), GGAG-binding causes, as indicated by the CD data, a large conformational change in both protein and RNA. Considering the proximity between the GGAG motif and the Dicer

cleavage site in pre-*let*-7, the latter might be important for Lin28's physiological function.

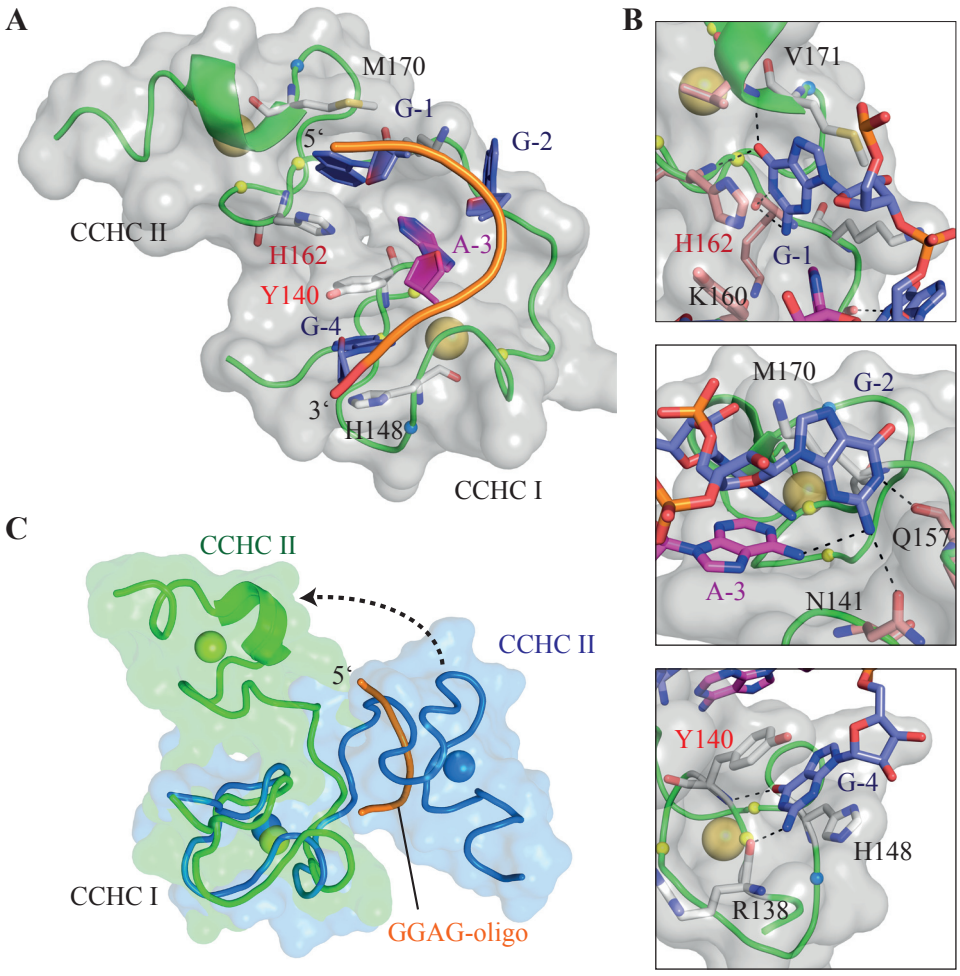


Figure 51: The Lin28 ZKD specifically binds to GGAG-containing ssRNAs. (**A**) Structure of mLin28A ZKD bound to a GGAG motif (derived from PDB-ID 3TSO). mLin28A is represented in green cartoon and the bound GGAG motif in violet (G) and pink (A). Tyr140 of the 1st and His162 of the 2nd Zn-knuckle are key residues for the interaction, since they contact each other and stack with the bases, thereby establishing a kinked conformation in the RNA. (**B**) All three G's of the GGAG motif are specifically recognized via various hydrogen bonds. Furthermore, G-1 and G-4 are bound in a hydrophobic pocket formed by His140, His162, Tyr140 and Met170. A-3 assists in formation of a kinked RNA backbone since it mediates hydrogen bonds with G-2 and in some structures also with G-1. (**C**) Comparison between apo hLin28A ZKD (blue) (PDB-ID 2CQF) and GGAG-bound mLin28A ZKD (green) (PDB-ID 3TSO). The linker between the two CCHC Zn-knuckles undergoes a large conformational shift upon GGAG-binding. Therefore, RNA binding is associated with large structural rearrangements in Lin28's ZKD.

4.3 IMPLICATIONS OF LIN28-MEDIATED PRE-*let-7* BINDING

Although the isolated *Xtr*Lin28B ZKD exhibited strong specificity towards GGAG-containing oligonucleotides, it poorly bound to pre-*let-7f/g* and thus could not inhibit pre-*let-7* processing by Dicer. Furthermore, mutation of the GGAG motif to AAAA did not disrupt binding of WT *Xtr*Lin28B completely, but descreased the binding affinity by a factor of ~10. My data demonstrated that both Lin28 RBDs tethered together via a basic linker are the minimal requirements for efficient pre-*let-7* binding and Dicer blocking. This observation is in agreement with recent studies, though the binding constants of Lin28•pre-*let-7* differed considerably. While two groups reported K_D's in the low and subnanomolar range [151, 228], most studies determined binding constants in the high nanomolar to low micromolar range [157, 158, 227, 236]. The different experimental procedures and binding conditions such as EMSA versus ELISA-based assays, protocol for hairpin formation, salt and Mg^{2+} concentrations might explain the observed discrepancies. For example, Loughlin *et al.* reported a strong salt dependency of Lin28•pre-*let-7* binding constants [228], whereas Piskounova *et al.* observed a more than 1000-fold decrease of K_D upon removal of competitor tRNA in EMSAs (compare [157] and [151]). Despite these discrepancies, the indepedent binding data in this study provide strong evidence that *Xtr*Lin28B binds to pre-*let-7f/g* in the high nanomolar (ITC data, conducted without competitor tRNA) to low micromolar range (EMSA data, performed with competitor tRNA). In addition, the IC_{50} of *Xtr*Lin28B-dependent inhibition of pre-*let-7f* processing by Dicer (1.4 ± 0.4 µM, see Figure 34) as well as the apparent K_D derived from RNA remodeling experiments with pre-*let-7g** (*Xtr*Lin28A: 1.2 ± 0.1 µM; *Xtr*Lin28B: 1.07 ± 0.02 µM; see Figure 36) support the observed binding data.

With regard to binding stoichiometry, recent studies observed 1:1 binding between Lin28 CSD+ZKD and pre-*let-7*'s preE using analytical ultracentrifugation or LILBIC (laser-induced liquid bead ion desorption mass spectrometry [227, 236]. The RALS and ITC binding experiments revealed a similar result for WT *Xtr*Lin28B•preE-*let-7f*, even though the SEC peak was not monotonic and showed higher oligomeric species at the leading edge of the peak. For full-length pre-*let-7f*, however, both experimental approaches revealed a 1:1 to 2:1 (protein:RNA) equilibrium, indicating that more than one Lin28 molecule can bind to pre-*let-7f*. The premise that Lin28 can bind to both miRNA duplex and preE was already suggested by Rybak et al. [130]. They showed that competitor RNAs corresponding to the 5′ ds stem of pre-*let-7a* or to its preE impaired Lin28•pre-*let-7* complex formation. Moreover, results from gel-shift assays with mLin28A and pre-*let-7g* have suggested binding of up to three Lin28 molecules at high concentrations [194]. However, the crystal structures of truncated mLin28A with preE-*let-7*-derived RNAs [227] suggest that WT Lin28 binds with a 1:1 stoichiometry to preE-*let-7* and some parts of the ds stem. Additional ssRNA sequences might be generated in the pre-*let-7* stem as a process of binding, thereby explaining the 1:1 to 2:1 stoichiometries in the RALS and ITC experiments.

Interestingly, pre-*let-7f/g* binding and remodeling by WT *Xtr*Lin28B, *Xtr*Lin28B CSD and *Xtr*Lin28B CSD+ZKD were strongly cooperative. Since Hill coeffi-

cients of all these variants are almost identical, we conclude that Lin28 binds to multiple sites of pre-*let-7* and that the CSD is essential for cooperative binding and RNA remodeling. Consistent with this, a mutation in the internal loop of pre-*let-7f* next to the GGAG motif (preE-mut I) had a similar effect on binding as a mutation of the terminal hairpin loop (preE-mut II, see Figure 35). Moreover, *Xtr*Lin28B CSD bound to pre-*let-7g* although the observed CSD binding site is not single-stranded according to secondary-structure predictions. Given the structural diversity of *let-7* family preEs (Figure 52), Lin28 CSD must fulfill two functions: i) it acts as an RNA remodeling factor that locally induces structural changes in preE's, thereby ensuring accessibility of the GGAG motif; ii) it contributes to overall binding affinity of Lin28•pre-*let-7* complexes via base stacking interactions with terminal loop nucleotides, thereby establishing a stable and directional binding of Lin28. The flexible Arg/Lys-rich linker between Lin28s' RBDs [227, 194] can span variable distances between the CSD and the ZKD, thus facilitating binding of Lin28 to structurally diverse pre-elements of *let-7* family members.

4.4 CSDs are versatile tools for local RNA remodeling

A recent study proposed that Lin28 induces a structural change within preE-*let-7g* and thus directly prevents the processing of pre-*let-7g* by Dicer [158]. Using enzymatic footprinting, the authors reported that upon Lin28 binding, a part of the upper stem of pre-*let-7* becomes more susceptible to cleavage by single-strand specific ribonucleases.

In this thesis, I could demonstrate that the structural change within pre-*let-7g* is mediated through the CSD, since only mutations within this domain considerably impaired the remodeling activity. In principle, two mechanism can explain how Lin28 CSD induces structural changes within pre-*let-7* (Figure 53 A).

First, it could alter the equilibrium between two different RNA conformations by selectively binding to one folding state (conformational selection). This mechanism implies that two energetically similar secondary structures exist under physiological conditions and therefore might apply to pre-*let-f1* (Figure 52). Second, the CSD could associate with ss internal and hairpin loops and melt less stable ds stems, thereby remodeling the overall RNA secondary structure (induced fit). This process is highly reversible, unless a downstream factor, such as binding of the Lin28 ZKD to the now accessible GGAG motif, completes the remodeling reaction. Given that pre-*let-7g* does not contain a low-energy secondary structure where the GGAG motif is single-stranded, the induced fit mechanism remains the only option in this case. The high cooperativity of Lin28•pre-*let-7g* binding and the different rates of pre-*let-7g** remodeling support this hypothesis.

The function of CSDs as protein domains that can induce local secondary structural changes within RNA was originally proposed for bacterial Csps in the process of transcription antitermination [237, 185, 238]. In the case of *E. coli* CspE, a melting pathway was suggested, whereby two conserved Phe and one His, which resemble Phe46 and Phe66 (binding subsite 5) and His68 (binding subsite 4) in *Xtr*Lin28B CSD, intercalate between bases of the stem. In

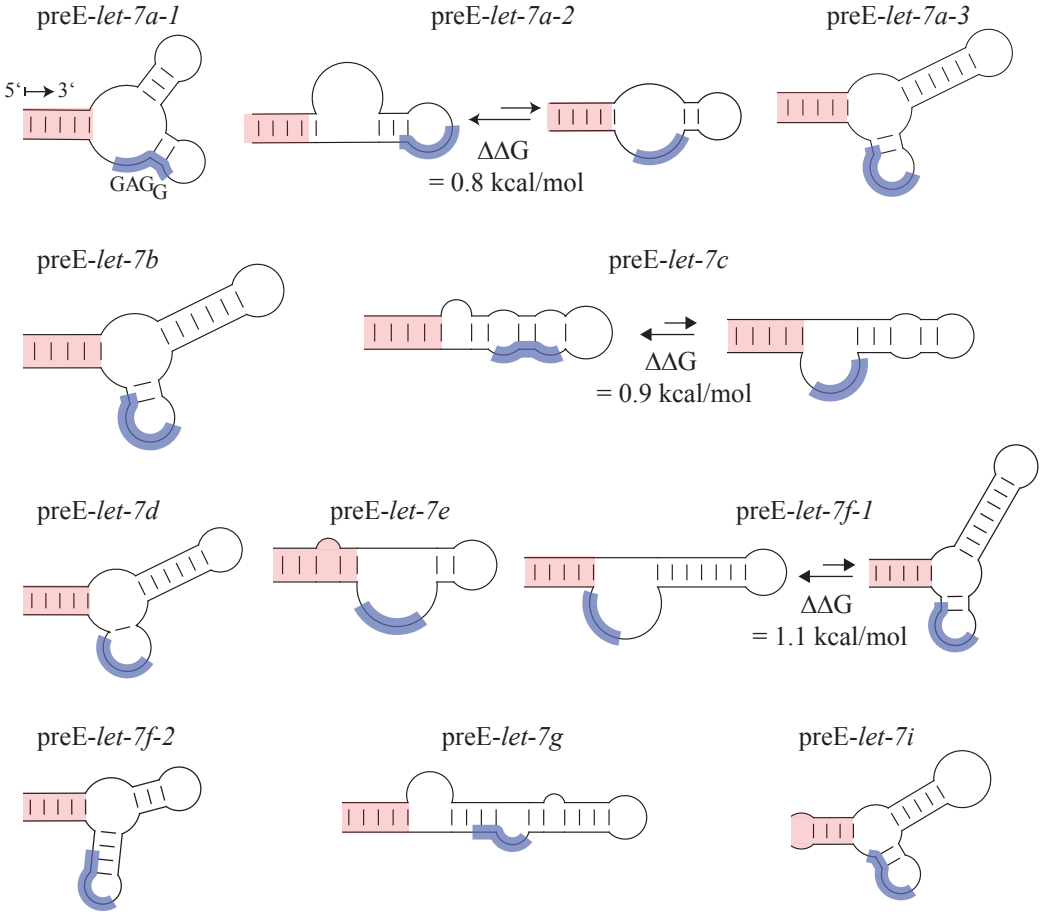


Figure 52: The pre-element of *let-7* family members are structurally diverse. In six out of eleven human *let-7* family members, the conserved GGAG motif (blue) is not freely accessible for ZKD binding according to secondary-structure predictions. In three preE's (pre-*let-7a-2*, pre-*let-7c*, pre-*let-7f-1*) the GGAG motif is inaccessible for GGAG binding in at least one folding state. Secondary-structure predictions of human *let-7* family members (except miR-98 and miR-202) were calculated and drawn by CLC genomics workbench 3.65. All lowest-energy secondary structures of each *let-7* group within a ΔΔG range of 1.5 kcal/mol are depicted. For simplicity, only 5 bp of the miRNA stem are depicted (labeled in red).

agreement with this observation, His68 and Phe77 (binding subsite 1,2) are essential for the remodeling activity of *Xtr*Lin28B, and it is likely that a similar melting pathway exists for Lin28 CSD. Based on the crystal structures and remodeling experiments, I could generate a model explaining how Lin28 can melt dsRNAs (Figure 53 B, C). In a first step *Xtr*Lin28B CSD binds via its hydrophobic platform (Trp39, Phe46, Phe66 and His68) to ss loops close to the ss/dsRNA junction. The plane of His68's imidazole ring is positioned orthogonal to the plane of the first RNA base-pair and disrupts hydrogen bonding between the bases. As a consequence, one base flips into the hydrophobic pocket mainly formed by Phe77, while the other base stacks on His68. This process could be repeated until an optimal interaction with the CSD is mediated, thereby

generating new RNA secondary structures with exposed sequence motifs for downstream factors.

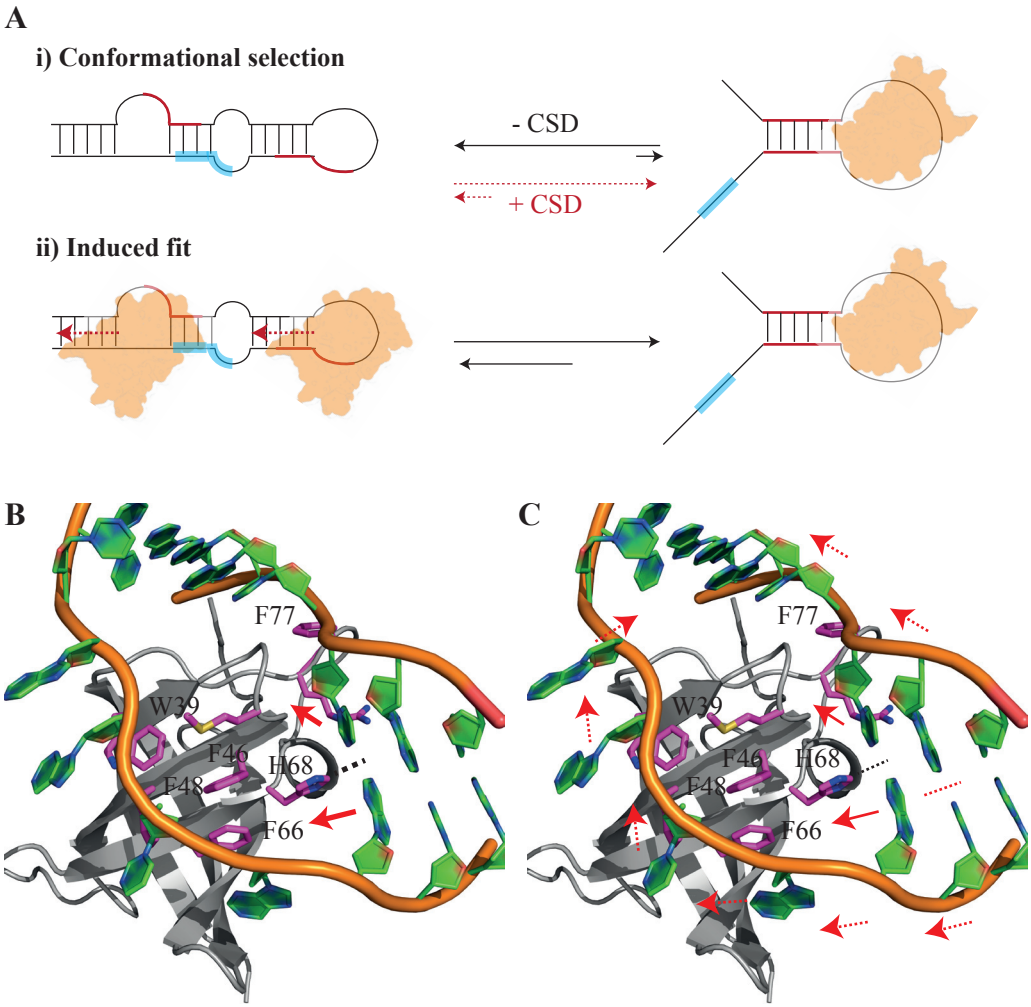


Figure 53: Mechanisms of Lin28 CSD-induced structural changes in RNAs. (**A**) Lin28 CSD can induce a structural change via two mechanisms: (i) Conformational selection: it could selectively bind to a less stable RNA folding state, thereby stabilizing it and shift the equilibrium towards this secondary structure; ii) Induced fit: it could bind to ss/ds RNA junctions within internal or hairpin loops and reversibly melt short ds stems. As a result, more favorable binding sites for the CSD could be generated and binding motifs for downstream factors might become exposed, which drive the process in one direction. (**B, C**) On a molecular level, Lin28 CSD might associate with ss internal or hairpin loops via its hydrophobic platform (Trp39, Phe48, Phe66, Phe46), thereby positioning His68 at the ss/dsRNA junction. The imidazole ring of His68 intercalates between bases of the stem. Then, one base flips into the hydrophobic pocket mainly formed by Phe77 (binding subsite 2), whereas the other stacks on His68. This process could be repeated until optimal interactions with the CSD are achieved.

4.5 MODEL OF LIN28-MEDIATED INHIBITION OF PRE-*let-7* PROCESSING

The kinetic analysis of pre-*let-7g** remodeling revealed two phases. Since variations in *Xtr*Lin28B ZKD (Tyr133Ala, His155Ala) impaired only the second rate, we concluded that the first rate can be traced back to contributions of the CSD. Furthermore, the ZKD and the flexible Arg/Lys-rich linker between the RBDs are unstructured in solution and would require larger structural rearrangements for RNA binding (see Figure 51 C). Based on these data, we can propose a multistep binding model: Lin28 CSD first binds pyrimidine-rich ss sequences within preE and induces a conformational change that facilitates subsequent binding of the ZKD to the conserved GGAG motif (Figure 54). Since binding of *Xtr*Lin28B and *Xtr*Lin28B CSD to pre-*let-7f* is highly cooperative (Figure 33, [194]), it is possible that remodeling is performed in trans by another Lin28 molecule via its CSD. Moreover, Lin28 CSD melts the upper stem region including the Dicer cleavage site [158, 227]. As a consequence of RNA remodeling, the GGAG motif becomes exposed and is therefore accessible for ZKD binding. Since the conserved GGAG motif is adjacent to the Dicer cleavage site, ZKD binding to this motif stabilizes an unwound Dicer cleavage site by bending the RNA, thereby directly inhibiting pre-*let-7* processing by Dicer. This would also explain why both intact RBDs are required for efficient inhibition of pre-*let-7* processing by Dicer, even though the isolated CSD and most of the analyzed variants were able to bind and remodel pre-*let-7g*. Structural changes within the RNA as well as the sequence specific interaction via the ZKD ensure directional binding of Lin28, which may be essential for recruiting TUT4 and promoting polyuridylation.

4.6 IMPORTANCE OF CCHC ZN-KNUCKLES IN PRE-*let-7* POLYRURIDYLATION

Aside from direct binding and inhibition of pre-*let-7* processing, Lin28 also promotes polyuridylation of pre-*let-7* miRNA, thereby irreversibly inhibiting Dicer cleavage and marking it for degradation by cellular nucleases [160]. In the absence of Lin28, TUT4 showed only basal uridylation of pre-*let-7*. A recent study reported that monouridylation of pre-miRNAs with a single nt 3′ overhang triggers processing by Dicer [162]. This activity was demonstrated to be mediated by TUT4 and other cellular TUTases in somatic cells, which normaly lacks Lin28, thus explaining the basal activity of TUT4 in the *in vitro* experiments. Conversely, upon addition of Lin28, uridylation of pre-*let-7* was ~8-fold enhanced and manifested as a shift of the corresponding band in denaturing urea PAGE (see Figure 48). The latter indicates that rather polyuridylation than monouridylation occured. The strong stimulation of polyuridylation activity might be explained through direct recruitment of TUT4 by Lin28, as demonstrated in the co-immunoprecipitation and LUMIER assays. However, the interaction was significantly enhanced after addition of *in vitro* transcribed pre-*let-7*, suggesting that a binary Lin28•pre-*let-7* complex is necessary for efficient recruitment of TUT4.

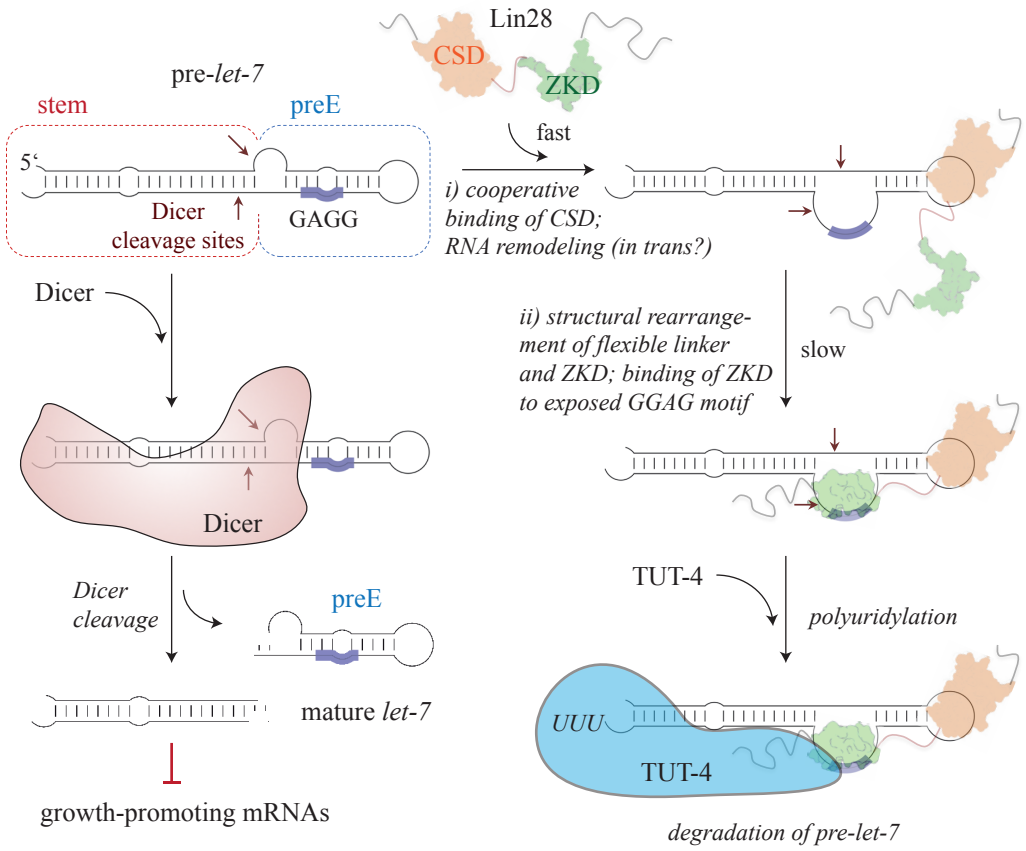


Figure 54: Schematic model of Lin28-mediated binding and inhibition of pre-*let-7* maturation. In the absence of Lin28, Dicer's PAZ domain recognizes the 3'-overhang of pre-*let-7* and cleaves the substrate about 22 nt from the end (Dicer cleavage site, indicated by red arrows). In the presence of Lin28, pre-*let-7* binds to Lin28's preformed nucleic acid-binding platform within the CSD. Binding is mediated by pyrimidine-comprising ss regions within preE-*let-7* and associated with preE-remodeling and melting of the upper stem, including the Dicer cleavage site (fast reaction rate). Given that binding of Lin28 to pre-*let-7* is highly cooperative, remodeling of pre-*let-7* may be facilitated in trans by another Lin28 molecule. Once the conserved GGAG motif is completely accessible, Lin28's ZKD specifically binds the motif and anchors Lin28 in position (slow reaction rate). The second reaction phase is rather slow, as both the ZKD and flexible linker must undergo large conformational changes as witnessed from apo and nucleotide-bound structures. As a consequence, the Dicer cleavage site remains constantly unwound, and Lin28 inhibits cleavage by Dicer. Moreover, the sequence specific interaction via the ZKD ensures directional binding of Lin28 to pre-*let-7*, which may be essential for the recruitment of TUT4 and further polyuridylation of pre-*let-7*.

When identifying the regions of TUT4 that are essential for pre-*let-7* polyuridylation, it was observed that depletion or mutation of CCHC Zn-knuckles II and III completely abolished the uridylation activity. As observed for Lin28, these variants of Zn-knuckles typically recognize single-stranded G's or G-containing dinucleotides [188, 227, 228]. Since TUT4 Zn-knuckles II and III are separated by a 44 nt long linker, they might act more independently from each other and eventually recognize individual G-bulges rather than the GGAG motif. Possi-

ble binding sites within pre-*let-7* could be the G-rich seed sequence or G-bulges often predicted in the ds stem. In agreement with a previous study [132], my experiments indicated that the GGAG motif is essential for TUT4 uridylation. However, this could simply reflect the importance of this motif for directional Lin28 binding. Thus, it might be the Lin28-induced structural change within pre-*let-7*, which facilitates TUT4 binding and triggers processive polyuridylation instead of monouridylation. Consistent with this hypothesis, both Lin28 RBDs were essential for the TUT4•Lin28 interaction and for promoting pre-*let-7* uridylation.

Whereas seven TUTases are present in humans (Figure 55 A), only TUT4 and to a minor extent TUT7 are capable of polyuridylating pre-*let-7* [132, 239]. Given their domain structure, TUT4 and TUT7 share several unique features that are likely essential for pre-*let-7* uridylation. In addition to the catalytic motif, which is common to all human TUTases, TUT4 and TUT7 contain three CCHC Zn-knuckles, a second (functionally inactive) nucleotidyl transferase domain (NTD*) linked to a PAP-associated domain, a putative C2H2 Zn-finger as well as low-complexity sequences at the N- and C-termini. The latter are known to be important for localization to RNA stress granules or P-bodies [240] thus linking the enzymes to the miRNA biogenesis/decay pathway (see Section 1.4.3). A similar function was associated with Lin28's C-terminus, which is also composed of low-complexity sequences. Indeed, depletion of hLin28B's C-terminus disturbed the interaction with TUT4 but still allowed modest stimulation of pre-*let-7* uridylation. Conversely, TUT4's C-terminus had only minor impact on TUT4•hLin28 interaction and pre-*let-7* uridylation, suggesting that only Lin28's C-terminus is important for pre-*let-7* uridylation.

In addition to CCHC Zn-knuckles II and III, a second, non-functional catalytic motif (NTD*/PAP-associated domain) and a putative C2H2 Zn-finger were also demonstrated to be essential for pre-*let-7* uridylation, thereby explaining why only TUT4 and TUT7 can polyuridylate pre-*let-7* [239]. A recent crystal structure of the yeast TUT4 homolog Cid1 revealed that the catalytic motif (NTD + PAP-associated domain) can nonspecifically bind to ~13 nucleotides via basic patches across the enzyme's surface (Figure 55 B) [241]. Therefore, the additional NTP*/PAP-associated domain and C2H2 Zn-Finger might be required for recognizing the longer ds stem of pre-*let-7* and thus are crucial for processive polyuridylation in a Lin28-dependent manner. Since *C. elegans* PUP2 does not possess any of these domains, it might use a distinct mechanism for pre-*let-7* polyuridylation.

4.7 IMPLICATIONS OF LIN28•MRNA BINDING

Apart from inhibiting *let-7* biogenesis, Lin28 also exhibits *let-7* independent functions. Multiple studies showed that Lin28 binds to a variety of mRNAs and modulates their translation [171, 172, 174, 173]. High-throughput sequencing of Lin28-crosslinked mRNAs revealed a strong enrichment of GGAG or GGAG-like motifs (GGAGA, GGAGAU [173] or AAGNGG, AAGAGN, NUGUGN [173, 174]). These observations underscore the importance of Lin28's ZKD for mediating sequence specific binding. Conversely, hardly any significantly en-

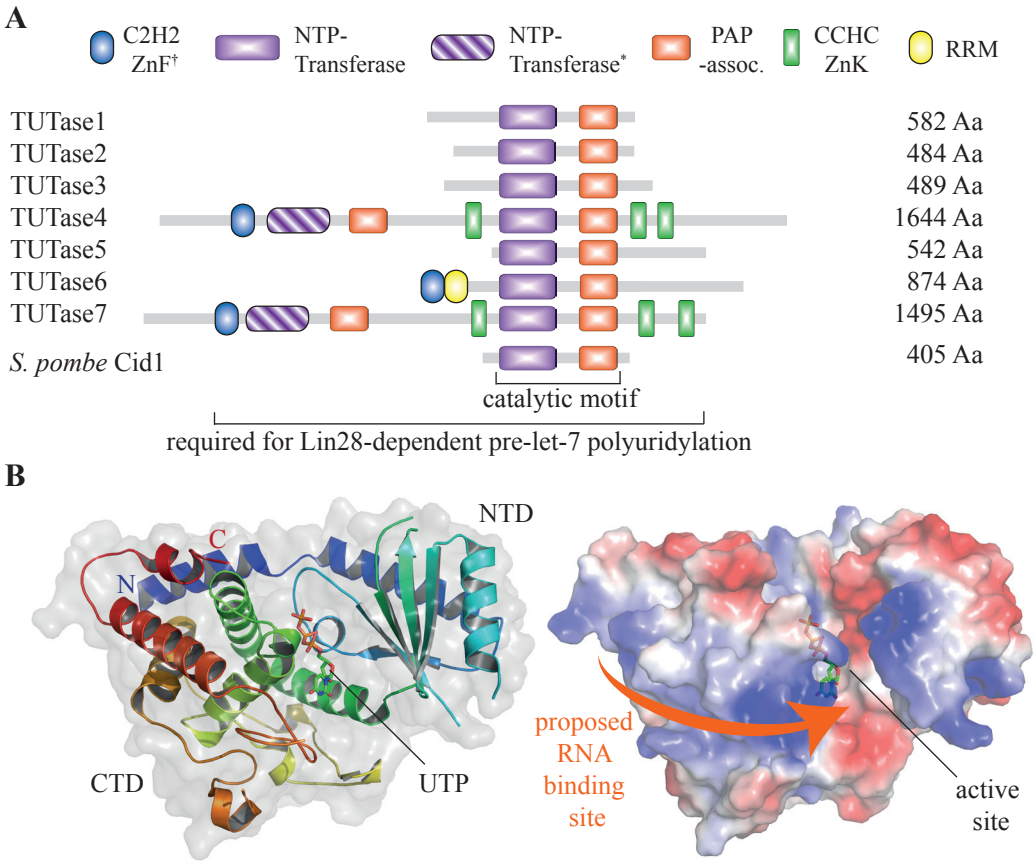


Figure 55: Human TUTases. (**A**) Schematic domain architecture of human TUTases. C2H2 ZnF: putative Zn-finger; NTP: nucleotidyl-transferase domain; NTP*: non-functional nucleotidyl-transferase domain; PAP-assoc.: poly(A) polymerase-associated domain; CCHC ZnK: retroviral-type CCHC Zn-knuckle; RRM: RNA-recognition motif; *S. pombe*: *Schizosaccharomyces pombe*. (**B**) Cartoon and space-filling representations of *S. pombe* Cid1•UTP (PDB-ID 4E80). The structure comprises two globular domains encompassing a large catalytic cleft that binds UTP. Nonspecific RNA binding is supposedly mediated via three basic patches at the protein surface that can harbor up to 13 nucleotides. Additional CCHC Zn-knuckles, a C2H2 Zn-finger and a non-functional NTP/PAP-associated domain are required for Lin28-stimulated pre-*let-7* polyuridylation.

riched motif could be detected upstream or downstream of this cluster, thereby confirming the low sequence specificity of Lin28 CSD. Though most of the observed GGAG-like motifs were detected in hairpin loops, some binding sites were also located in more complex RNA structures and branched hairpins indicating that the CSD might remodel mRNA secondary structure to expose binding motifs for downstream factors.

Interestingly, the downstream effects of Lin28•mRNA binding were quite distinct in recent genome-wide studies and comprised translational stimulation of growth-promoting [171] and alternative splicing factors, [173] as well as translational repression of ER-destined mRNAs [174]. In addition to the localization of binding sites (coding sequence, 3′ UTR), it is probable that Lin28-induced structural changes within mRNAs and/or direct protein-protein interactions af-

fect the translation of target mRNAs. Since *Xtr*Lin28B CSD+ZKD was still able to bind pre-*let-7* with similar affinity as the full-length protein, we conclude that regions outside the RBDs have other functions than RNA binding and may be involved in protein-protein interactions. In agreement with this observation, depletion of hLin28B's C-terminus impaired the interaction with TUT4. Moreover, truncation of Lin28's N- and C-termini abrogated an association with RNA helicase A (RHA), thus diminishing the translation of growth-promoting mRNAs [172]. In addition, so far no one has observed marginal difference in RNA binding between the two Lin28 isoforms. This raises the question whether the existence of two Lin28 isoforms in mammals, which differ mainly in C-terminal length, are important to mediate different protein interactions and thus promote alternate cellular responses.

APPENDIX A

A.1 INSTRUMENTS

Table 16: Instruments used in this study

INSTRUMENT	TYPE	MANUFACTURER
Agarose gel chamber	HG370, HG330	Savant
Blotting device	TE70 PWR	Amersham Biosciences
Camera	DXC-390P	Leica
CD spectrometer	Chirascan	Applied Photophysics
Centrifuges	Biofuge pico	Heracus
	Biofuge stratos	Heracus
	5417R	Eppendorf
	Avanti-J26 XP	Beckman Coulter
Chromatography columns	GSTrap FF 5 ml	GE Healthcare
	Heparin HP 5 ml	GE Healthcare
	Superdex HiLoad S75 16/60	GE Healthcare
	Superdex HiLoad S75 26/60	GE Healthcare
	HisTrap FF 5 ml	GE Healthcare
Chromatography systems (FPLC)	Äkta Explorer	GE Healthcare
	Äkta Prime Plus	GE Healthcare
	Äkta Purifier	GE Healthcare
Concentrators	Vivaspin	Sartorius
	Amicon Ultra	Millipore
Cryoloops	Mounted Cryoloops	Hampton
Crystal observation	Homebase (20°C)	The Automation Partnership (TAP)
	Rock Imager	Formulatrix
	Imager (4°C, 20°C)	
Crystallization plates	Crystal quick	Greiner Bio-One

Table 16 – Continued on next page

Table 16 – Continued from previous page

INSTRUMENT	TYPE	MANUFACTURER
Crystallization robot	Gryphon	Art Robbins Instruments
	Hydra II	Matrix Technologies
Denaturing PAGE	Model S2 Sequencing Apparatus	Biometra
Electroporation cuvettes	0.4 cm	Invitrogen
Electroporation system	Gene Pulser Xcell	Bio-Rad
EMSA-PAGE	PerfectBlue Dual Gel System 20x20 cm	peqlab
Fluidizer	Microfluidizer	Microfluidics
Gel documentation	BA digital Basis Set	Biometra
	LAS 4000	Fujifilm
	UV-table	Biometra
Incubator	MIR-153	SANYO
Isothermal titration calorimeter	VP-ITC	GE Healthcare
Microscope	Wild M3C	Leica
	Wild M420	Leica
Nanodrop sepctrophotometer	ND 1000	peqlab
Peristaltic pump	P-1	Pharmacia
pH-meter	CG840, Lab850	Schott
Phosphoimager	FLA-3000	Fujitsu
Precision scale	MCI Analytical AC1205	Sartryos
RALS system	VE3580 RI Detector	Viscotek
	270 Dual Detector	Viscotek
Scales	MR2002 CP2202S-0CE	Sartorius
SDS-PAGE	Mini 2-D System	Bio-Rad
Shaker Incubator	Innova	New Brunswick Scientific
	HT	Infors

Table 16 – Continued on next page

Table 16 – Continued from previous page

INSTRUMENT	TYPE	MANUFACTURER
Luminescence plate reader	TECAN Infinite M200	Tecan
Thermal block	Thermomixer 5437	Eppendorf
Thermocycler	PTC-200	MJ Research
Ultracentrifuge	Optima L-100	Beckman Coluter
Vortex mixer	7-2020	neoLAB

A.2 ENZYMES AND ANTIBODIES

Table 17: Enzymes, antibodies and serum proteins used in this study

ENZYME	MANUFACTURER
Benzoase	Merck kGaA, Darmstadt, D
Calf intestine phosphatase	NEB, Frankfurt am Main, D
Dicer (human)	Gelantis, San Diego, USA
DNAse I	Roche, Mannheim, D
Gateway LR Clonase II Enzyme mix	Life Technologies, Carlsbad, USA
Lysozyme	Sigma-Aldrich, St.Louis, USA
Pfu Plus! DNA Polymerase	Roboklon, Berlin, D
PreScission Protease (GST-tagged)	GE Healthcare, Piscataway, USA and Daumke lab, Berlin, D
Ribonuclease inhibitor	Roboklon, Berlin, D
RiboLock RNase inhibitor	Thermo Fisher Scientific, Vilnius, Lithuania
Restriction endonucleases	NEB, Frankfurt am Main, D
RNase-free DNase Set	Qiagen, Hilden, D
T4 DNA Ligase	NEB, Frankfurt am Main, D
T7 RNA polymerase	Roboklon, Berlin, D
T7 RNA polymerase (6xHis-tagged)	Heinemann lab, Berlin, D
T4 Polynucleotide Kinase	Life Technologies, Carlsbad, USA
Taq DNA Polymerase	Roboklon, Berlin, D
TEV protease (His-tagged)	Heinemann lab, Berlin, D
Thermostable pyrophosphatase	Roboklon, Berlin, D
ANTIBODY / PROTEIN	**MANUFACTURER**
Anti-Flag (mouse)	Sigma-Aldrich, St.Louis, USA
Anti-His HRP conjugate	Qiagen, Hilden, D (see also Kit)
Anti-mouse HRP	Sigma-Aldrich, St.Louis, USA
HRP conjugate (goat)	
Anti-V5 (mouse)	Sigma-Aldrich, St. Louis, USA
Anti-sheep IgG (rabbit)	Jakson Immunoresearch, Suffolk, UK
Bovine serum albumin (BSA)	Carl Roth GmbH, Karlsruhe, D
Sheep gamma globulin	Jakson Immunoresearch, Suffolk, UK

A.3 KITS

Table 18: Kits used in this study

KIT	MANUFACTURER
Additive Screen	Hampton Research, Aliso Viejo, USA
Bradford protein assay	Bio-Rad, Hercules, USA
Dual Luciferase Assay System	Promega, Fitchburg, USA
Enhanced Chemiluminescence (ECL) Kit	GE Healthcare, Piscataway, USA
Gelfiltration LMW calibration kit	GE Healthcare, Piscataway, USA
illustra MicroSpin G-50 Columns	GE Healthcare, Piscataway, USA
KinaseMax 5′ end labeling kit	Life Technologies, Carlsbad, USA
Mark 12 unstained protein standard	Life Technologies, Carlsbad, USA
Penta-His HRP Conjugate Kit	Qiagen, Hilden, D
Precision Plus prestained protein standard	Bio-Rad, Hercules, USA
Precision Plus unstained protein standard	Bio-Rad, Hercules, USA
QIAGEN Plasmid Maxi Kit	Qiagen, Hilden, D
QIAprep Spin Miniprep Kit	Qiagen, Hilden, D
QIAquick Gel Extraction Kit	Qiagen, Hilden, D
QIAquick PCR Purification Kit	Qiagen, Hilden, D
Recombinant Dicer Enzyme Kit	Genlantis, SanDiego, USA
T7 *in vitro* transcription kit	Roboklon, Berlin, D
The $AmSO_4$ Suite	Qiagen, Hilden, D
The Anions Suite	Qiagen, Hilden, D
The Cations Suite	Qiagen, Hilden, D
The Classics II Suite	Qiagen, Hilden, D
The Classics Lite Suite	Qiagen, Hilden, D
The Classics Suite	Qiagen, Hilden, D
The ComPAS Suite	Qiagen, Hilden, D
The JSCG+ Suite	Qiagen, Hilden, D
The MPD Suite	Qiagen, Hilden, D
The Nucleix Suite	Qiagen, Hilden, D
The PACT Suite	Qiagen, Hilden, D
The PEGs II Suite	Qiagen, Hilden, D
The PEGs Suite	Qiagen, Hilden, D
The pHClear II Suite	Qiagen, Hilden, D
The pHClear Suite	Qiagen, Hilden, D
The ProComplex Suite	Qiagen, Hilden, D

A.4 MEDIA

Table 19: Media

MEDIUM	COMPONENTS
Luria-Bertani (LB)	5 g/l yeast extract
	10 g/l tryptone
	5 g/l NaCl
LB Agar	1.5% (w/v) Agar in LB
SOB medium	20 g/l tryptone
	5 g/l yeast extract
	0.5 g/l NaCl
	2.5 mM KCl
SOC medium	1x SOB medium
	10 mM $MgCl_2$
	10 mM $MgSO_4$
	20 mM glucose
TB medium (4x)	96 g/l yeast extract
	48 g/l tryptone
	16% (w/v) glycerol
TB buffer	0.34 M KH_2PO_4
	1.44 M K_2HPO_4

A.5 BUFFERS

As mentioned previously, for purification of proteins containing Zn-fingers/Zn-knuckles, 10 µM $ZnSO_4$ were added to the corresponding buffers†.

Table 20: Protein purification buffers

BUFFER	COMPONENTS
General buffers	
1x PBS pH 7.4	137 mM NaCl
	2.7 mM KCl
	4.37 mM Na_2HP_4
	1.47 mM KH_2PO_4
Lysis buffer	1x PBS pH 7.4
	0.5 M NaCl
	5% (v/v) glycerol
	1 mM dithiothreitol (DTT)
	~1 µg/ml DNase I
	1 U/ml Benzoase
	1 tablet EDTA-free Complete Protease Inhibitor
	10 µM $ZnSO_4$ (optional†)
TEV cleavage buffer	20 mM Tris pH 8.0
	150 mM NaCl
	1 mM DTT
	10 µM $ZnSO_4$ (optional†)
Hexahistidine-tagged proteins	
Ni-NTA equilibration buffer	20 mM Tris pH 8.0
	500 mM NaCl
	0.5 mM DTT
	10 mM imidazole
	10 µM $ZnSO_4$ (optional†)
Ni-NTA wash buffer	20 mM Tris/HCl, pH 8.0
	500 mM NaCl
	40 mM imidazole
	0.5 mM DTT
	10 µM $ZnSO_4$ (optional†)
Ni-NTA elution buffer	20 mM Tris/HCl, pH 8.0
	500 mM NaCl

Table 20 – Continued on next page

Table 20 – Continued from previous page

BUFFER	COMPONENTS
	250 mM imidazole
	0.5 mM DTT
	10 µM $ZnSO_4$ (optional†)
GST-tagged proteins	
GST equilibration buffer	20 mM HEPES pH 7.5
	300 mM NaCl
	1 mM DTT
	10 µM $ZnSO_4$ (optional†)
GST wash buffer	20 mM HEPES pH 7.5
	500 mM NaCl
	1 mM DTT
	5 mM $MgCl_2$
	0.5 mM ATP
	10 µM $ZnSO_4$ (optional†)
GST on-column clevage buffer	20 mM HEPES pH 7.5
	300 mM NaCl
	1 mM DTT
	10 µM $ZnSO_4$ (optional†)
GST elution buffer	20 mM HEPES pH 7.5
	300 mM NaCl
	1 mM DTT
	20 mM glutathione (reduced)
	10 µM $ZnSO_4$ (optional†)
Cation exchange chromatography (Heparin)	
Heparin loading buffer	20 mM Bis-Tris, pH 6.5
	50 mM NaCl
	1 mM DTT
	10 µM $ZnSO_4$ (optional†)
Heparin elution buffer	20 mM Bis-Tris, pH 6.5
	1 M NaCl
	1 mM DTT
	10 µM $ZnSO_4$ (optional†)

Table 20 – Continued on next page

Table 20 – Continued from previous page

BUFFER	COMPONENTS
SEC	
SEC buffer I	20 mM Tris pH 8.0
	100 mM NaCl
	1 mM DTT
	10 µM $ZnSO_4$ (optional†)
SEC buffer II	20 mM Tris pH 8.0
	100 mM KCl
	1 mM DTT
	10 µM $ZnSO_4$ (optional†)
SEC buffer III	20 mM HEPES pH 7.5
	300 mM NaCl
	1 mM DTT
	10 µM $ZnSO_4$ (optional†)

Table 21: Molecular biology and RNA biochemistry buffers.

BUFFER	COMPONENTS
Agarose gels and denaturing PAGE	
TAE buffer, pH 8.0	40 mM Tris
	1 mM EDTA, pH 8.0
	20 mm acetic acid
1x TBE buffer, pH 8.0	90 mM Tris-borate
	2 mM EDTA
TBE urea buffer	1x TBE buffer
	8 M urea
EMSA	
Tris-glycine buffer, pH 8.3	25 mM Tris
	192 mM L-glycine
EMSA binding buffer	20 mM Tris/HCl, pH 8.0
	60 mM KCl
	10 U Ribonuclease Inhibitor
	1 mM DTT
EMSA loading buffer	1x EMSA binding buffer
	50% (v/v) glycerol
	0.1% (w/v) bromophenol blue
	0.1% (w/v) xylene cyanol
In vitro transcription and uridylation	
Dicer reaction buffer	20 mM Tris/HCl, pH 8.0
	150 mM NaCl
	5 mM $MgCl_2$
	1 mM ATP
	0.5 mM DTT
In vitro transcription buffer	80 mM HEPES, pH 7.5
	3 mM $MgCl_2$
	1 mM spermidine
	0.1 mg/ml BSA
In vitro uridylation buffer	20 mM Tris/HCl pH 7.8
	100 mM KCl
	5% (v/v) glycerol
	3 mM $MgCl_2$

Table 21 – Continued on next page

Table 21 – Continued from previous page

BUFFER	COMPONENTS
SDS-PAGE	
SDS running buffer, 10x	250 mM Tris/HCl
	1% (w/v) SDS
	2.5 M glycine
2x SDS sample buffer	100 mM Tris/HCl pH 6.8
	20% glycerol
	0.01% (w/v) bromophenol blue
	4% (v/v) β-mercaptoethanol
	12.5 mM EDTA
Coomassie staining solution	0.25% (w/v) Coomassie brilliant blue R250
	50% (v/v) ethanol
	10% (v/v) acetic acid
Coomassie destaining solution	50% ethanol
	10% (v/v) acetic acid

Table 22: Buffers for biophysical measurements.

BUFFER	COMPONENTS
CD buffer	20 mM Na-phosphate buffer pH 7,4
	150 mM NaF
Fluorescence quencher assay buffer	1x ITC buffer
	40 µM N-acytyl tryptophanamide
ITC buffer	20 mM Tris/HCl, pH 8.0
	60 mM KCl
RNA remodeling buffer	20 mM Tris/HCl, pH 7.5
	60 mM KCl
	10 mM $MgCl_2$
	1 mM DTT

Table 23: Western blot and LUMIER buffers

BUFFER	COMPONENTS
Western blot and immunodetection	
Blocking buffer (anti-Flag, anti-V5)	1x PBS-T
	5% milk powder
Blotting buffer (pH 8.3)	25 mM Tris
	192 mm glycine
	10% isopropanol
PBS-T	1x PBS
	0.1% (v/v) Tween-20
LUMIER buffers	
Carbonate buffer	70 mM $NaHCO_3$
	30 mM Na_2CO_3
Coating solution	10 µg/ml gamma globulin (sheep)
	(in carbonate buffer)
Capture solution	3.3 µg/ml anti-sheep IgG (rabbit)
	(in carbonate buffer)
HEK293 lysis buffer	0.1% NP40
	50 mM HEPES, pH 7.4
	150 mM NaCl
	1.5 mM $MgCl_2$
	1 mM EDTA
	1 mM DTT
	~1 µg/ml DNase I (RNase-free)
	1 tablet Complete protease inhibitor
	10 U/ml RiboLock RNase inhibitor
TBS-T	10 mM Tris/HCl, pH 7.4
	150 mM NaCl
	0.05% Tween-20
TP solution	1% (w/v) BSA from casein
	(in carbonate buffer)

A.6 PROTEIN AND MIRNA CONSTRUCTS

A.6.1 *Plasmids encoding for protein constructs*

Table 24: List of plasmids encoding for proteins that were prepared in the present work.

VECTOR	PROTEIN	LENGTH	COMMENT
C. elegans Lin28 constructs			
pENTR4	*Cel*Lin28	1-227	WT
pENTR4	*Cel*Lin28	1-184	N-terminus+CSD+ZKD
pENTR4	*Cel*Lin28	44-184	CSD+ZKD
pENTR4	*Cel*Lin28	50-184	CSD+ZKD
pENTR4	*Cel*Lin28	44-191	CSD+ZKD
pENTR4	*Cel*Lin28	44-227	CSD+ZKD+C-terminus
pENTR4	*Cel*Lin28	50-129	CSD
pENTR4	*Cel*Lin28	141-184	ZKD
pENTR4	*Cel*Lin28	50-184	Δ136-138, replaced by AAA
pENTR4	*Cel*Lin28	50-184	Δ136-138
pENTR4	*Cel*Lin28	50-184	Δ129-138, replaced by AAA
pENTR4	*Cel*Lin28	50-184	Δ129-138
pDEST15	*Cel*Lin28	1-227	WT
pDEST15	*Cel*Lin28	1-184	N-terminus+CSD+ZKD
pDEST15	*Cel*Lin28	44-184	CSD+ZKD
pDEST15	*Cel*Lin28	50-184	CSD+ZKD
pDEST15	*Cel*Lin28	44-191	CSD+ZKD
pDEST15	*Cel*Lin28	44-227	CSD+ZKD+C-terminus
pDEST15	*Cel*Lin28	50-129	CSD
pDEST15	*Cel*Lin28	141-184	ZKD
pDEST15	*Cel*Lin28	50-184	Δ136-138, replaced by AAA
pDEST15	*Cel*Lin28	50-184	Δ136-138
pDEST15	*Cel*Lin28	50-184	Δ129-138, replaced by AAA
pDEST15	*Cel*Lin28	50-184	Δ129-138
pDEST17	*Cel*Lin28	1-227	WT
pDEST17	*Cel*Lin28	1-184	N-terminus+CSD+ZKD
pDEST17	*Cel*Lin28	44-184	CSD+ZKD
pDEST17	*Cel*Lin28	50-184	CSD+ZKD
pDEST17	*Cel*Lin28	44-191	CSD+ZKD
pDEST17	*Cel*Lin28	44-227	CSD+ZKD+C-terminus
pDEST17	*Cel*Lin28	50-129	CSD

Table 24 – Continued on next page

Table 24 – Continued from previous page

VECTOR	PROTEIN	LENGTH	COMMENT
pDEST17	*Cel*Lin28	141-184	ZKD
pDEST17	*Cel*Lin28	50-184	Δ136-138, replaced by AAA
pDEST17	*Cel*Lin28	50-184	Δ136-138
pDEST17	*Cel*Lin28	50-184	Δ129-138, replaced by AAA
pDEST17	*Cel*Lin28	50-184	Δ129-138
pDEST-*Renilla*	*Cel*Lin28	1-227	WT
pDEST-*Renilla*	*Cel*Lin28	1-184	N-terminus+CSD+ZKD
pDEST-*Renilla*	*Cel*Lin28	44-184	CSD+ZKD
pDEST-*Renilla*	*Cel*Lin28	50-184	CSD+ZKD
pDEST-*Renilla*	*Cel*Lin28	44-191	CSD+ZKD
pDEST-*Renilla*	*Cel*Lin28	44-227	CSD+ZKD+C-terminus
pDEST-*Renilla*	*Cel*Lin28	50-129	CSD
pDEST-*Renilla*	*Cel*Lin28	141-184	ZKD
C. elegans PUP2 constructs			
pENTR4	*Cel*PUP2	2-509	WT
pENTR4	*Cel*PUP2	44-494	
pENTR4	*Cel*PUP2	59-509	
pENTR4	*Cel*PUP2	59-494	
pENTR4	*Cel*PUP2	59-434	
pENTR4	*Cel*PUP2	2-434	
pENTR4	*Cel*PUP2	2-494	
pENTR4	*Cel*PUP2	111-434	
pENTR4	*Cel*PUP2	111-494	
pDEST15	*Cel*PUP2	2-509	WT
pDEST15	*Cel*PUP2	44-494	
pDEST15	*Cel*PUP2	59-509	
pDEST15	*Cel*PUP2	59-494	
pDEST15	*Cel*PUP2	59-434	
pDEST15	*Cel*PUP2	2-434	
pDEST15	*Cel*PUP2	2-494	
pDEST15	*Cel*PUP2	111-434	
pDEST15	*Cel*PUP2	111-494	
pDEST17	*Cel*PUP2	2-509	WT
pDEST17	*Cel*PUP2	44-494	
pDEST17	*Cel*PUP2	59-509	

Table 24 – Continued on next page

Table 24 – Continued from previous page

VECTOR	PROTEIN	LENGTH	COMMENT
pDEST17	*Cel*PUP2	59-494	
pDEST17	*Cel*PUP2	59-434	
pDEST17	*Cel*PUP2	2-434	
pDEST17	*Cel*PUP2	2-494	
pDEST17	*Cel*PUP2	111-434	
pDEST17	*Cel*PUP2	111-494	
pDEST-*Firefly*	*Cel*PUP2	2-509	WT
pDEST-*Firefly*	*Cel*PUP2	44-494	
pDEST-*Firefly*	*Cel*PUP2	59-509	
pDEST-*Firefly*	*Cel*PUP2	59-494	
pDEST-*Firefly*	*Cel*PUP2	59-434	
pDEST-*Firefly*	*Cel*PUP2	2-434	
pDEST-*Firefly*	*Cel*PUP2	2-494	
pDEST-*Firefly*	*Cel*PUP2	111-434	
pDEST-*Firefly*	*Cel*PUP2	111-494	
human Lin28A constructs			
pQLinkH	hLin28A	1-209	WT
pQLinkH	hLin28A	33-179	CSD+ZKD
pQLinkH	hLin28A	33-185	CSD+ZKD
pQLinkH	hLin28A	39-179	CSD+ZKD
pQLinkH	hLin28A	39-185	CSD+ZKD
pQLinkH	hLin28A	33-193	CSD+ZKD
pQLinkH	hLin28A	33-159	CSD+CCHC I
pQLinkH	hLin28A	33-111	CSD
pQLinkH	hLin28A	33-120	CSD
pQLinkH	hLin28A	33-130	CSD
pQLinkH	hLin28A	33-121	CSD
pQLinkH	hLin28A	34-193	CSD+ZKD
pQLinkH	hLin28A	126-185	ZKD
pQLinkH	hLin28A	126-193	ZKD
pQLinkH	hLin28A	134-185	ZKD
pQLinkH	hLin28A	134-193	ZKD
pENTR4	hLin28A	1-209	WT
pDEST53	hLin28A	1-209	GFP fused WT
pDEST-*Renilla*	hLin28A	1-209	WT

Table 24 – Continued on next page

Table 24 – Continued from previous page

VECTOR	PROTEIN	LENGTH	COMMENT
pcDNA 3.1/nV5	hLin28A	1-209	WT
human Lin28B constructs			
pENTR4	hLin28B	1-250	WT
pDEST17	hLin28B	1-250	WT
pQLinkH	hLin28B	29-168	CSD+ZKD
pQLinkH	hLin28B	29-177	CSD+ZKD
pQLinkH	hLin28B	19-168	CSD+ZKD
pQLinkH	hLin28B	19-177	CSD+ZKD
pQLinkH	hLin28B	29-104	CSD
pQLinkH	hLin28B	29-112	CSD
pQLinkH	hLin28B	24-111	CSD
pENTR4	hLin28B	24-171	CSD+ZKD
pENTR4	hLin28B	24-171	Δ111-126
pENTR4	hLin28B	24-171	Δ111-126, replaced by AAAA
pENTR4	hLin28B	24-171	Δ111-126
pDEST15	hLin28B	24-171	CSD+ZKD
pDEST15	hLin28B	24-171	Δ111-126
pDEST15	hLin28B	24-171	Δ111-126, replaced by AAAA
pDEST15	hLin28B	24-171	Δ111-126
pENTR4	hLin28B	1-171	N-Term+CSD+ZKD
pENTR4	hLin28B	1-111	N-Term+CSD
pENTR4	hLin28B	117-250	ZKD+C-term
pENTR4	hLin28B	117-171	ZKD
pENTR4	hLin28B	1-250	F74A, R75G
pENTR4	hLin28B	1-250	Y130A, H152G
pDEST53	hLin28B	1-250	WT
pDEST53	hLin28B	1-171	N-Term+CSD+ZKD
pDEST53	hLin28B	1-111	N-Term+CSD
pDEST53	hLin28B	117-250	ZKD+C-term
pDEST53	hLin28B	117-171	ZKD
pDEST53	hLin28B	1-250	F74A, R75G
pDEST53	hLin28B	1-250	Y130A, H152G
pDEST-*Renilla*	hLin28B	1-250	WT
pDEST-*Renilla*	hLin28B	1-171	N-Term+CSD+ZKD
pDEST-*Renilla*	hLin28B	1-111	N-Term+CSD

Table 24 – Continued on next page

Table 24 – Continued from previous page

VECTOR	PROTEIN	LENGTH	COMMENT
pDEST-*Renilla*	hLin28B	117-250	ZKD+C-term
pDEST-*Renilla*	hLin28B	117-171	ZKD
pDEST-*Renilla*	hLin28B	1-250	F74A, R75G
pDEST-*Renilla*	hLin28B	1-250	Y130A, H152G
pcDNA 3.1/nV5	hLin28B	1-250	WT
pcDNA 3.1/nV5	hLin28B	1-171	N-Term+CSD+ZKD
pcDNA 3.1/nV5	hLin28B	1-111	N-Term+CSD
pcDNA 3.1/nV5	hLin28B	117-250	ZKD+C-term
pcDNA 3.1/nV5	hLin28B	117-171	ZKD
pcDNA 3.1/nV5	hLin28B	1-250	F74A, R75G
pcDNA 3.1/nV5	hLin28B	1-250	Y130A, H152G
human TUT4 constructs			
pENTR4	TUT4	1-1644	WT
pENTR4	TUT4	1-1264	ΔC, ΔCCHC II, III
pENTR4	TUT4	1-1381	ΔC
pENTR4	TUT4	348-1644	ΔN
pENTR4	TUT4	348-1381	ΔN, ΔC
pENTR4	TUT4	348-1264	ΔN, ΔC, ΔCCHC II, III
pENTR4	TUT4	895-1264	ΔNTD*, ΔC, ΔCCHC II, III
pENTR4	TUT4	348-711	NTD*
pENTR4	TUT4	934-1264	NTD
pENTR4	TUT4	934-1381	NTD+CCHC II, III
pENTR4	TUT4	934-1381	NTD+CCHC I, II, III
pDEST15	TUT4	1-1644	WT
pDEST15	TUT4	1-1264	ΔC, ΔCCHC II, III
pDEST15	TUT4	1-1381	ΔC
pDEST15	TUT4	348-1644	ΔN
pDEST15	TUT4	348-1381	ΔN, ΔC
pDEST15	TUT4	348-1264	ΔN, ΔC, ΔCCHC II, III
pDEST15	TUT4	895-1264	ΔNTD*, ΔC, ΔCCHC II, III
pDEST15	TUT4	348-711	NTD*
pDEST15	TUT4	934-1264	NTD
pDEST15	TUT4	934-1381	NTD+CCHC II, III
pDEST15	TUT4	934-1381	NTD+CCHC I, II, III
pDEST-*Firefly*	TUT4	1-1644	WT

Table 24 – Continued on next page

Table 24 – Continued from previous page

VECTOR	PROTEIN	LENGTH	COMMENT
pDEST-*Firefly*	TUT4	1-1264	ΔC, ΔCCHC II, III
pDEST-*Firefly*	TUT4	1-1381	ΔC
pDEST-*Firefly*	TUT4	348-1644	ΔN
pDEST-*Firefly*	TUT4	348-1381	ΔN, ΔC
pDEST-*Firefly*	TUT4	348-1264	ΔN, ΔC, ΔCCHC II, III
pcDNA 3.1/nFlag	TUT4	1-1644	WT
pcDNA 3.1/nFlag	TUT4	1-1264	ΔC, ΔCCHC II, III
pcDNA 3.1/nFlag	TUT4	1-1381	ΔC
pcDNA 3.1/nFlag	TUT4	348-1644	ΔN
pcDNA 3.1/nFlag	TUT4	348-1381	ΔN, ΔC
pcDNA 3.1/nFlag	TUT4	348-1264	ΔN, ΔC, ΔCCHC II, III
pcDNA 3.1/nFlag	TUT4	895-1264	ΔNTD*, ΔC, ΔCCHC II, III
mouse Lin28A constructs			
pQLinkH	mLin28A	2-209	WT
pQLinkH	mLin28A	33-179	CSD+ZKD
pQLinkH	mLin28A	33-185	CSD+ZKD
pQLinkH	mLin28A	39-179	CSD+ZKD
pQLinkH	mLin28A	39-185	CSD+ZKD
mouse Lin28B constructs			
pQLinkH	mLin28B	20-189	CSD+ZKD
pQLinkH	mLin28B	20-199	CSD+ZKD
pQLinkH	mLin28B	27-189	CSD+ZKD
pQLinkH	mLin28B	27-199	CSD+ZKD
Xenopus tropicalis Lin28A constructs			
pQLinkH	*Xtr*Lin28A	2-195	WT
pQLinkH	*Xtr*Lin28A	26-171	CSD+ZKD
pQLinkH	*Xtr*Lin28A	26-179	CSD+ZKD
pQLinkH	*Xtr*Lin28A	33-171	CSD+ZKD
pQLinkH	*Xtr*Lin28A	33-179	CSD+ZKD
pQLinkH	*Xtr*Lin28A	26-107	CSD
pQLinkH	*Xtr*Lin28A	26-116	CSD
pQLinkH	*Xtr*Lin28A	129-171	ZKD
pQLinkH	*Xtr*Lin28A	124-171	ZKD

Table 24 – Continued on next page

Table 24 – Continued from previous page

VECTOR	PROTEIN	LENGTH	COMMENT
Xenopus tropicalis Lin28B constructs			
pQLinkH	*Xtr*Lin28B	2-253	WT
pQLinkH	*Xtr*Lin28B	27-170	CSD+ZKD
pQLinkH	*Xtr*Lin28B	27-175	CSD+ZKD
pQLinkH	*Xtr*Lin28B	32-170	CSD+ZKD
pQLinkH	*Xtr*Lin28B	32-175	CSD+ZKD
pQLinkH	*Xtr*Lin28B	27-186	CSD+ZKD
pQLinkH	*Xtr*Lin28B	27-197	CSD+ZKD
pQLinkH	*Xtr*Lin28B	27-220	CSD+ZKD
pQLinkH	*Xtr*Lin28B	27-243	CSD+ZKD
pQLinkH	*Xtr*Lin28B	27-152	CSD+CCHC I
pQLinkH	*Xtr*Lin28B	27-108	CSD
pQLinkH	*Xtr*Lin28B	27-114	CSD
pQLinkH	*Xtr*Lin28B	114-170	ZKD
pQLinkH	*Xtr*Lin28B	125-170	ZKD
pQLinkH	*Xtr*Lin28B	125-175	ZKD
pQLinkH	*Xtr*Lin28B	128-175	ZKD
pQLinkH	*Xtr*Lin28B	27-170	Δ115-118, replaced by AAAA
pQLinkH	*Xtr*Lin28B	27-170	Δ115-129, replaced by AAAA
pQLinkH	*Xtr*Lin28B	2-253	W39A
pQLinkH	*Xtr*Lin28B	2-253	F48A
pQLinkH	*Xtr*Lin28B	2-253	F66A
pQLinkH	*Xtr*Lin28B	2-253	H68A
pQLinkH	*Xtr*Lin28B	2-253	F77A
pQLinkH	*Xtr*Lin28B	2-253	F77A, R78G
pQLinkH	*Xtr*Lin28B	2-253	Y133A
pQLinkH	*Xtr*Lin28B	2-253	H155A
pQLinkH	*Xtr*Lin28B	27-114	CSD W39A
pQLinkH	*Xtr*Lin28B	27-114	CSD F48A
pQLinkH	*Xtr*Lin28B	27-114	CSD F66A
pQLinkH	*Xtr*Lin28B	27-114	CSD H68A
pQLinkH	*Xtr*Lin28B	27-114	CSD F77A
pQLinkH	*Xtr*Lin28B	27-114	CSD F77A, R78G
pQLinkH	*Xtr*Lin28B	125-175	ZKD Y133A
pQLinkH	*Xtr*Lin28B	125-175	ZKD H155A

A.6.2 *Plasmids encoding for miRNAs*

For simplification, in the Results and Discussion part of this thesis, the species of the used miRNAs were not named. The used miRNAs always matched their binding partners, that means EMSAs with *Xtr*Lin28 were performed with *Xtr*-pre-*let-7*, *in vitro* uridylation assays with hLin28 were conducted with *hsa-let-7* and so on.

Table 25: List of all plasmids encoding for miRNAs that were prepared in the present work.

VECTOR	MIRNA	COMMENT
pENTR4	*Cel*-pre-*let-7*	WT
pENTR4	*hsa*-pre-*let-7f*	WT
pENTR4	*hsa*-pre-*let-7g*	WT
pENTR4	*hsa*-pre-*let-7g* GGAG-mut	GGAG mutated to AAAA
pENTR4	*Xtr*-pre-*let-7f*	WT
pENTR4	MS2-HH *Xtr*-pre-*let-7f*	with MS2-tag and Hammerhead ribozyme
pENTR4	*Xtr*-pre-*let-7f* GGAG-mut	GGAG mutated to AAAA
pENTR4	*Xtr*-pre-*let-7f* preE-mut I	terminal loop mutated
pENTR4	*Xtr*-pre-*let-7f* preE-mut II	internal loop mutated
pENTR4	*Xtr*-pre-*let-7f* preE-mut III	pre-mut I and pre-mut II
pENTR4	*Xtr*-pre-*let-7g*	WT

APPENDIX B

B.1 MACROMOLECULAR X-RAY CRYSTALLOGRAPHY

In this section, a brief overview about the theoretical principles of macromolecular X-ray crystallography is provided. For detailed information, the reader is referred to basic textbooks on X-ray crystallography [242, 243, 244].

B.1.1 *Growth of macromolecular crystals*

A prerequisite to perform X-ray diffraction experiments is the availability of suitable crystals containing the macromolecule of interest, since only a well-ordered arrangement of macromolecules yields a detectable diffraction signal. In principle, crystallization is performed using a concentrated solution of the macromolecule of interest and slowly reducing its solubility by increasing the concentration of precipitant. In practice, vapor diffusion methods like hanging-drop or sitting drop vapor diffusion are used to achieve crystallization (Figure 56 A). In these methods, the macromolecule-containing solution is mixed with an equal amount of precipitant and equilibrated in a closed system against a larger reservoir filled with precipitant. Since the concentration of precipitant is higher in the reservoir, the most volatile component evaporates from the drop and equilibrates with the reservoir. Thus, concentrations of macromolecule and precipitant within the drop gradually increase until the macromolecular solubility limit is exceeded. Under optimal conditions, crystallization nuclei form and crystals start to grow leading to depletion of macromolecules in the solution. Some of the initial crystals continue to grow in the growth region (metastable phase) until they reach equilibrium with the saturated macromolecule-containing solution (Figure 56 B).

Success of crystallization depends on several factors such as temperature, precipitant concentration, pH, additives and the kind of precipitant. Since these factors are not predictable, conditions that yield well-ordered and diffracting crystals, must be empirically found.

B.1.2 *Collection of diffraction data*

The smallest entity of a crystal is the asymmetric unit. This building block is sufficient to reproduce the crystallographic unit cell by symmetry operations such as rotations, inversions or reflections. A further translation of the unit cell in three dimensions yields the macroscopic crystal. The dimensions of the unit cell are defined by the length of the three unique edges a, b, c and three unique angles α, β, γ. Knowledge of the unit cell edges and angles together with symmetry operations allows one to define the space group. In principle, 230 space groups are possible, however, due to the chirality of proteins or protein•nucleic

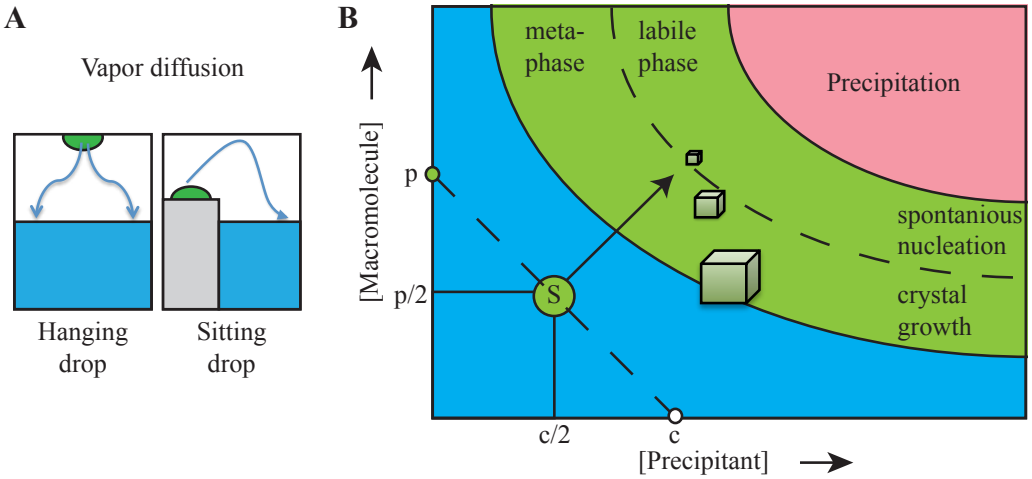


Figure 56: Path of an idealized vapor-diffusion experiment through the crystallization phase diagram. (**A**) Schematic sketch of hanging-drop (left) and sitting-drop (right) vapor diffusion crystallization experimental setups. (**B**) Phase diagram of protein crystallization, illustrating an idealized vapor-diffusion experiment. Equal amounts of macromolecule and precipitant are mixed and equilibrated with pure precipitant in a closed system. Due to concentration differences the volatile component (mostly water) will diffuse to the reservoir leading to an increasing concentration of macromolecule and precipitant in the drop. Once the spontaneous nucleation phase is reached, nucleation takes place, and a crystal begins to grow. The conenctration of free macromolecules drops until until the crystals are in equilibrium with the saturated protein solution.

acid complexes, inversions or mirror planes cannot occur, thereby limiting the number of possible space groups to 65.

In an X-ray diffraction experiment, an incident X-ray beam is elastically scattered by electrons within the crystal. Due to the regular arrangement of a crystalline lattice, constructive or destructive interference of diffracted waves can occur. To simplify matters, a set of equivalent parallel lattice planes can be defined that are thought as sources of diffraction. W. L. Bragg showed that sets of parallel planes with indices hkl and interplanar spacing d_{hkl} generate peaks of scattered intensity when the following condition is fulfilled (Equation 8).

$$2d_{hkl} \cdot \sin\theta = n \cdot \lambda \tag{8}$$

d_{hkl}	interplanar spacing
θ	diffraction angle
λ	wavelength
n	integer

Consequently, at a given wavelength λ, constructive interference occurs when the angle of incidence equals the diffraction angle, and the pathlength difference d_{hkl} is equal to an integer number of the wavelength (Figure 57 A). Under these conditions, scattered X-rays from one set of lattice planes will lead to one

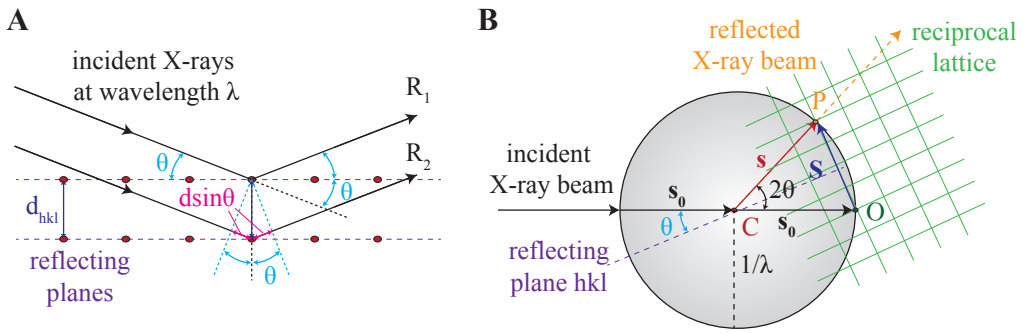


Figure 57: Reflection condition and Ewald construction. (**A**) Graphical interpretation of the Bragg equation. Ray 1 (R_1) and ray 2 (R_2) interfere constructively, when the additional distance traveled by R_2 is an integral multiple of the wavelength λ. (**B**) The Ewald sphere construction can be used to reconstruct the direction of the scattered beam. The sphere has a radius $1/\lambda$ and passes through the origin of the reciprocal lattice (O). The incident X-ray beam is scattered by the crystal in the center of the sphere (C). The Bragg reflection condition is fulfilled for the set of lattice planes hkl when its reciprocal lattice point lies on the Ewald sphere (P). The diffraction vector $\vec{S}$ (indicated in bold) is the difference between the incoming vector $\vec{s_0}$ and the scattering vector $\vec{s}$.

reflection. The reflections are indexed by the Miller indices hkl from the lattice plane. Intensities of each reflection are determined by the local electron density distribution in the unit cell and contains valuable information about the spatial arrangement of atoms within the crystal. In addition, the direction of each reflection contains information about unit cell parameters and crystallographic symmetry. Helpful concepts to determine the direction of diffracted beams are the reciprocal lattice and the Ewald sphere construction (Figure 57 B). Accordingly, a sphere that passes the origin of the reciprocal lattice (O) is drawn around the crystal with a radius of $1/\lambda$. The incident X-ray beam propagates along the x-axis and is scattered at the crystal, thereby cutting the sphere at the reciprocal lattice point P. Consequently, the diffraction pattern is a projection of the reciprocal lattice points on the Ewald's sphere and thus can be mapped back to the corresponding scattering vectors in reciprocal space. Crystal lattice planes, which do not lie on the Ewald sphere can be brought to reflection by rotating the crystal and with it the reciprocal lattice.

B.1.3 *Data processing*

Based on the orientation and space group of the crystal, a data collection strategy is applied that allows, in the ideal case, measurement of all reflections. After data collection, reflections of the diffraction pattern are assigned with Miller indices hkl corresponding to their position in the reciprocal lattice and integrated over their intensity. In the process of data reduction, symmetry-related reflections are summarized and used to put all observations on a common scale. The quality of data can be estimated from R_{meas} (Equation 9), an indicator that deter-

mines deviations of redundant reflections independently from the multiplicity of the dataset [245].

$$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{n_{hkl}}{(n_{hkl}-1)}} \sum_{i}^{n_{hkl}} |\hat{I}_{hkl} - I_{hkl,i}|}{\sum_{hkl} \sum_{i}^{n_{hkl}} I_{hkl,i}} \tag{9}$$

n_{hkl} multiplicity of reflection hkl
$\hat{I}_{hkl}$ mean intensity of reflection hkl
$I_{hkl,i}$ intensity of an individual reflection hkl

The mathematical description between an object and its diffraction pattern can be described by a Fourier sum. Consequently, we can define a Fourier series describing the total scattering from a unit cell with n atoms at positions $\vec{r}_j$ (j = 1,2,3..., n) as:

$$\vec{F}(S) = \sum_{j=1}^{n} f_j \cdot e^{(2\pi i \vec{r}_j \cdot \vec{S})} \tag{10}$$

where $\vec{F}(S)$ is the structure factor, f_j the atomic scattering factor and $\vec{S}$ the diffraction vector ($\vec{S} = \vec{s} - \vec{s}_0$, see also Figure 57 B). The structure factor is a function of electron density distribution in the unit cell. By integrating over all electrons instead of separate atoms we obtain:

$$\vec{F}(S) = \int_{cell} \rho(\vec{r}) \cdot e^{2\pi i \vec{r}_j \cdot \vec{S}} dv \tag{11}$$

$\rho(r)$ is the electron density at position $\vec{r}$ in the unit cell. If V is volume and x, y and z are fractional coordinates in the unit cell ($0 \leqslant x < 1$, also for y and z) we obtain:

$$dv = V \cdot dx\, dy\, dz,$$

and

$$\vec{r} \cdot \vec{S} = (\vec{a} \cdot x + \vec{b} \cdot y + \vec{c} \cdot z) \cdot S = hx + ky + lz$$

Now $\vec{F}(S)$ of equation 11 can be rewritten as $\vec{F}(hkl)$:

$$\vec{F}(hkl) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(xyz) \cdot e^{2\pi i(hx+ky+lz)} dx\, dy\, dz \tag{12}$$

To calculate the electron density ρ at a position x, y and z in the unit cell we Fourier transform $\vec{F}(hkl)$, since $\vec{F}(hkl)$ is the Fourier transform of $\rho(xyz)$

and vice versa. In addition, the Laue condition tell us that the diffraction vector must be oriented in discrete directions with respect to the incident beam. Thus the integral is replaced by a sum. Using $\overrightarrow{F} = |F(hkl)| \cdot e^{i\alpha}$ we finally yield:

$$\rho(xyz) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F(hkl)| \cdot e^{-2\pi i(hx+ky+lz)+i\alpha(hkl)} \tag{13}$$

Since the intensity of a reflection hkl is proportional to the square of the structure factor amplitude ($I_{hkl} \propto |F_{hkl}|^2$), merged intensities from the non-redundant data set can be used to calculate structure factor amplitudes. However, the phase contribution $\alpha(hkl)$ is lost in the diffraction experiment and must be determined by other means.

B.1.4 *Experimental phasing and molecular replacement*

To solve the crystallographic phase problem, the following methods can be employed: i) direct methods for phasing, ii) isomorphous replacement, iii) anomalous scattering and iv) molecular replacement.

At high resolutions (below 1Å) phases can be determined directly via measured intensities using statistical relationships between sets of structure factors. From these relationships, a number of initial phases can be obtained and used to estimate phases of the complete data set. Since this statistical relationship becomes weaker with increasing number of atoms, direct methods are only applied at very high resolutions.

Isomorphous replacement relies on incorporation of heavy atoms into the crystallized protein or nucleic acid. Typically, crystals of the macromolecule of interest are soaked in solutions containing Hg, Pt or Au, whereby the ions bind to specific sites in macromolecules, e.g. Hg^{2+} to Cys. If the derivative crystal is isomorphous to the native crystal and diffracts to a reasonably high resolution, changes in the reflection intensities can be used to determine heavy-atom structure amplitudes. This information allows one to compute a Patterson map, thereby locating the heavy atom in the crystal and subsequently allows one to calculate the heavy-atom phase angles. This latter information is then used to estimate the macromolecule phase angles.

Another method to solve the phase problem is anomalous scattering. A prerequisite of this method is the incorporation of atoms with sufficient anomalous scattering power, meaning an inequality of symmetry-related reflections. At energies of the incident X-ray beam close to the absorption edge of an element, Friedel's law is no longer true and reflection pairs hkl and -h-k-l differ in their intensity. These differences can then be used to locate anomalous scatterers via a Patterson map, and to estimate initial phase angles. In proteins, selenomethionine (SeMet) is often used as an anomalous scatterer, since SeMet-derivatized proteins can be easily produced in cells [246, 247]. In nucleic acids, brominated nucleosides such as 5-bromodeoxyuracil or 5-bromouracil can be incorporated during chemical synthesis and used for phasing [248].

If the structure of a homologous protein is known, it can be used as a search model in molecular replacement. To conduct molecular replacement, the search

model must be placed in the correct orientation and position in the unit cell of the measured crystal. In practice, the search for the best orientation is separated from the search of the best position.

Using the Patterson function it is possible to search the correct orientation independently from the location. The Patterson map can be calculated without phases from squared structure factor amplitudes and contains information about interatomic distance vectors (Equation 14).

$$P(uvw) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)|^2 \cdot e^{-2\pi i(hu+kv+lw)} \tag{14}$$

u, v, w: fractional coordinates of the Patterson map

If the search model is sufficiently similar to the target molecule, they should yield similar Patterson maps. The best orientation of the target molecule can be identified by rotating the Patterson map of a search molecule in small increments over all angles and checking at each time the agreement with the Patterson map calculated from the measured amplitudes (Patterson cross-rotation search). The resulting orientations are fixed and subsequently used for a translational search, in which the observed structure factor amplitudes $|F_o|$ are compared with the structure factor amplitudes of the model $|F_c|$ in its current location by means of the crystallographic R-Factor (see Equation 17). Once the correct orientation and location of the search model is determined, phases $\alpha_c(hkl)$ derived from this model can be used to calculate the electron density of the target molecule according to Equation 13.

B.1.5 *Model building and refinement*

The initial atomic model obtained by molecular replacement must be modified and adjusted to the experimentally determined electron density. Therefore, weighted electron density maps can be calculated and visualized in COOT [219]. The $2F_o - F_c$ map is typically contoured at $\sim 1\sigma$ and allows one to rebuild the atomic model (Equation 15). The difference electron density map $F_o - F_c$ is contoured at $\sim 3\sigma$ (Equation 16). Positive contours show fragments of a structure that are missing in a current model, whereas negative contours appear in regions with incorrectly placed atoms.

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l ||2F_o| - |F_c|| \cdot e^{-2\pi i(hx+ky+lz)+i\alpha_c(hkl)} \tag{15}$$

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l ||F_o| - |F_c|| \cdot e^{-2\pi i(hx+ky+lz)+i\alpha_c(hkl)} \tag{16}$$

Use of the appropriate maps makes it possible to refine the model in real space by adding, removing, or reorienting groups until the model fits best to the observed electron density. The atomic coordinates of the improved model

are then refined in reciprocal space. In this process, atomic coordinates and temperature or B-factors are adjusted to minimize the difference between measured structure factor amplitudes $|F_o|$ and structure factor amplitudes calculated from the current model $|F_c|$. Additional stereochemical restraints tying bond lengths, bond angles, torsion angles and planar groups are incorporated in the refinement process to prevent violation of the general stereochemistry. Furthermore, density modifications such as solvent flattening and non-crystallographic symmetry can help to improve the phases. The refined model is used to calculate a new electron density map with improved phases that can be subsequently used for another round of manual model building and refinement. Hence, an alternation of map interpretation in real space and refinement of atomic coordinates and B-factors in reciprocal space should converge $|F_o|$ and $|F_c|$. This convergence is measured by the residual index, or crystallographic R-factor (Equation 17).

$$R_{work/free} = \frac{\sum ||F_c| - |F_o||}{\sum |F_o|} \tag{17}$$

Whereas R_{work} includes all measured diffraction intensities, R_{free} is calculated with 5% of the data excluded from the refinement. Thus, R_{free} measures how well a current model predicts a subset of measured intensities that were not included in the refinement and therefore cannot be manipulated by the refinement process. Divergence of more than 5% between R_{free} and R_{work} indicates overfitting of the data.

C

APPENDIX C

C.1 LIST OF ABBREVIATIONS

Ago	Argonaute
Ap	Ampicillin
ATP	Adenosine-5′-triphosphate
bp	Base pair
C-terminus	Carboxy-terminus
CD	Circular dichroism
Cm	Chloramphicenol
CSD	Cold-shock domain
Csp	Cold-shock protein
CV	Column volumes
DNase	Deoxyribonuclease
ds	Double-stranded
DTT	Dithiothreitol
eIF	Eukaryotic initiation factor
EMSA	Electrophoretic mobility shift assay
ESC	Embryonic stem cells
FL	*Firefly* luciferase
FPLC	Fast protein liquid chromatography
GAP	GTPase-activating protein
GSH	Reduced glutathione
GST	Glutathione S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human immunodeficiency virus
Ig	Immunoglobulin
iPSC	Induced pluripotent stem cells
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
Kan	Kanamycin
LB	Lysogeny broth
LUMIER	Luminescence-based mammalian interactome assay
m^7G	7-methylguanylate

Table 26 – Continued on next page

Table 26 – Continued from previous page

miRISC	MiRNA induced silencing complex
miRNA	MicroRNA
mRNA	Messenger RNA
N-terminus	Amino-terminus
NC	Nucleocapsid protein
NLS	Nuclear localizing sequence
NoLS	Nucleolar localizing sequence
nt	Nucleotide
NTA	Nitrilotriacetic acid
NTP	Nucleotide triphosphate
OB	oligosaccharide/oligonucleotide-binding
P-bodies	Processing bodies
PCR	Polymerase chain reaction
PDB	Protein data bank
pre-miRNA	Precursor miRNA
preE	Pre-element
pri-miRNA	Primary miRNA
PAP	Poly(A) polymerase
PUP	Poly(U) polymerase
RALS	Right-angle light scattering
RBD	RNA-binding domains
RL	*Renilla* luciferase
RLC	RISC loading complex
RNase	Ribonuclease
RNP	ribonucleoprotein motif
SDS-PAGE	Sodiumdodecylsulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
ss	Single-stranded
TB	Terrific broth
TEV	Tobacco etch virus
TLS	Translation, libration, screw rotation displacement
TUT	Terminal uridyl transferase
UTP	Uridine-5'-triphosphate
UTR	Untranslated region
WT	Wild-type
ZKD	Zn-knuckle domain

C.2 AMINO ACID ABBREVIATIONS

For amino acids in the text, the three-letter code was used. In figures, the one letter code was used: A, Ala: alanine; C, Cys: cystine; D, Asp: aspartate; E, Glu: glutamate; F, Phe; phenylalanine; G, Gly: glycine; H, His: histidine; I, Ile: isoleucine; K; Lys: lysine; L, Leu: leucine; M, Met: methionine; N, Asn: asparagine; P, Pro: proline; Q, Gln: glutamine; R, Arg: arginine; S, Ser: serine; T, Thr: threonine; V, Val: valine; W, Trp: tryptophan; Y, Tyr: tyrosine; X: any amino acid.

C.3 NUCLEOTIDE ABBREVIATIONS

For the description of ribonucleosides and deoxyribonucleosides, the one letter code was used: A: adenosine/deoxyadenosine, G: guanosine/deoxyguanosine, C: cytidine/deoxycytidine, T: thymidine, U: uridine, R: purine nucleoside/purine-deoxynucleoside, Y: pyrimidine nucleoside/deoxynucleoside, H: any nucleoside/deoxynucleoside but guanosine, N: any nucleoside/deoxynucleoside.

BIBLIOGRAPHY

[1] Hou, J., Wang, P., Lin, L., Liu, X., Ma, F., An, H., Wang, Z., Cao, X. MicroRNA-146a feedback inhibits RIG-I-dependent Type I IFN production in macrophages by targeting TRAF6, IRAK1, and IRAK2. *J Immunol*, **183**, 2150–2158 (2009).

[2] Houbaviy, H. B., Murray, M. F., Sharp, P. A. Embryonic stem cell-specific MicroRNAs. *Dev Cell*, **5**, 351–358 (2003).

[3] Judson, R. L., Babiarz, J. E., Venere, M., Blelloch, R. Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat Biotechnol*, **27**, 459–461 (2009).

[4] Lee, R. C., Feinbaum, R. L., Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, **75**, 843–854 (1993).

[5] Jovanovic, M., Hengartner, M. O. miRNAs and apoptosis: RNAs to die for. *Oncogene*, **25**, 6176–6187 (2006).

[6] Taganov, K. D., Boldin, M. P., Baltimore, D. MicroRNAs and immunity: tiny players in a big field. *Immunity*, **26**, 133–137 (2007).

[7] O'Connell, R. M., Taganov, K. D., Boldin, M. P., Cheng, G., Baltimore, D. MicroRNA-155 is induced during the macrophage inflammatory response. *Proc Natl Acad Sci U S A*, **104**, 1604–1609 (2007).

[8] Ambros, V. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell*, **113**, 673–676 (2003).

[9] Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297 (2004).

[10] Wightman, B., Ha, I., Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell*, **75**, 855–862 (1993).

[11] Moss, E. G., Lee, R. C., Ambros, V. The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. *Cell*, **88**, 637–46 (1997).

[12] Reinhart, B., Slack, F., Basson, M., Pasquinelli, A., Bettinger, J., Rougvie, A., Horvitz, H., Ruvkun, G. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, **403**, 901–6 (2000).

[13] Lin, S.-Y., Johnson, S. M., Abraham, M., Vella, M. C., Pasquinelli, A., Gamberi, C., Gottlieb, E., Slack, F. J. The C elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. *Dev Cell*, **4**, 639–650 (2003).

[14] Abrahante, J. E., Daul, A. L., Li, M., Volk, M. L., Tennessen, J. M., Miller, E. A., Rougvie, A. E. The Caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs. *Dev Cell*, **4**, 625–637 (2003).

[15] Slack, F. J., Basson, M., Liu, Z., Ambros, V., Horvitz, H. R., Ruvkun, G. The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Mol Cell*, **5**, 659–669 (2000).

[16] Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degnan, B., Müller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., Ruvkun, G. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, **408**, 86–9 (2000).

[17] Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853–858 (2001).

[18] Lau, N. C., Lim, L. P., Weinstein, E. G., Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science*, **294**, 858–862 (2001).

[19] Lee, R. C., Ambros, V. An extensive class of small RNAs in Caenorhabditis elegans. *Science*, **294**, 862–864 (2001).

[20] Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J.-B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F., Rothberg, J. M. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380 (2005).

[21] Nam, J.-W., Shin, K.-R., Han, J., Lee, Y., Kim, V. N., Zhang, B.-T. Human microRNA prediction through a probabilistic co-learning model of sequence and structure. *Nucleic Acids Res*, **33**, 3570–3581 (2005).

[22] Huang, T.-H., Fan, B., Rothschild, M. F., Hu, Z.-L., Li, K., Zhao, S.-H. MiRFinder: an improved approach and software implementation for genome-wide fast microRNA precursor scans. *BMC Bioinformatics*, **8**, 341 (2007).

[23] Ambros, V., Lee, R. C., Lavanway, A., Williams, P. T., Jewell, D. MicroRNAs and other tiny endogenous RNAs in C. elegans. *Curr Biol*, **13**, 807–818 (2003).

[24] Dostie, J., Mourelatos, Z., Yang, M., Sharma, A., Dreyfuss, G. Numerous microRNPs in neuronal cells containing novel microRNAs. *RNA*, **9**, 180–186 (2003).

[25] Aravin, A. A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., Tuschl, T. The small RNA profile during Drosophila melanogaster development. *Dev Cell*, **5**, 337–350 (2003).

[26] Carthew, R. W., Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell*, **136**, 642–655 (2009).

[27] Kim, V., Han, J., Siomi, M. Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol*, **10**, 126–39 (2009).

[28] Cai, X., Hagedorn, C. H., Cullen, B. R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, **10**, 1957–1966 (2004).

[29] Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., Tuschl, T. New microRNAs from mouse and human. *RNA*, **9**, 175–179 (2003).

[30] Doench, J. G., Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev*, **18**, 504–511 (2004).

[31] Buessing, I., Slack, F. J., Grosshans, H. let-7 microRNAs in development, stem cells and cancer. *Trends Mol Med*, **14**, 400–409 (2008).

[32] Vella, M. C., Choi, E.-Y., Lin, S.-Y., Reinert, K., Slack, F. J. The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR. *Genes Dev*, **18**, 132–137 (2004).

[33] Shin, C., Nam, J.-W., Farh, K. K.-H., Chiang, H. R., Shkumatava, A., Bartel, D. P. Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell*, **38**, 789–802 (2010).

[34] Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233 (2009).

[35] Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S. H., Kim, V. N. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*, **23**, 4051–4060 (2004).

[36] Borchert, G. M., Lanier, W., Davidson, B. L. RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol*, **13**, 1097–1101 (2006).

[37] Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., Kim, V. N. The nuclear RNase III Drosha initiates microRNA processing. *Nature*, **425**, 415–9 (2003).

[38] Denli, A. M., Tops, B. B. J., Plasterk, R. H. A., Ketting, R. F., Hannon, G. J. Processing of primary microRNAs by the Microprocessor complex. *Nature*, **432**, 231–5 (2004).

[39] Gregory, R., Yan, K.-P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., Shiekhattar, R. The Microprocessor complex mediates the genesis of microRNAs. *Nature*, **432**, 235–40 (2004).

[40] Landthaler, M., Yalcin, A., Tuschl, T. The human DiGeorge syndrome critical region gene 8 and its D. melanogaster homolog are required for miRNA biogenesis. *Curr Biol*, **14**, 2162–7 (2004).

[41] Zeng, Y., Cullen, B. R. Sequence requirements for micro RNA processing and function in human cells. *RNA*, **9**, 112–123 (2003).

[42] Zeng, Y., Yi, R., Cullen, B. R. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc Natl Acad Sci USA*, **100**, 9779–9784 (2003).

[43] Zeng, Y., Yi, R., Cullen, B. R. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J*, **24**, 138–148 (2005).

[44] Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., Sohn, S. Y., Cho, Y., Zhang, B.-T., Kim, V. N. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, **125**, 887–901 (2006).

[45] Kim, Y.-K., Kim, V. N. Processing of intronic microRNAs. *EMBO J*, **26**, 775–783 (2007).

[46] Morlando, M., Ballarino, M., Gromak, N., Pagano, F., Bozzoni, I., Proudfoot, N. J. Primary microRNA transcripts are processed cotranscriptionally. *Nat Struct Mol Biol*, **15**, 902–909 (2008).

[47] Pawlicki, J. M., Steitz, J. A. Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production. *J Cell Biol*, **182**, 61–76 (2008).

[48] Berezikov, E., Chung, W.-J., Willis, J., Cuppen, E., Lai, E. C. Mammalian mirtron genes. *Mol Cell*, **28**, 328–336 (2007).

[49] Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M., Lai, E. C. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell*, **130**, 89–100 (2007).

[50] Ruby, J. G., Jan, C. H., Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature*, **448**, 83–86 (2007).

[51] Yi, R., Qin, Y., Macara, I. G., Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev*, **17**, 3011–6 (2003).

[52] Lee, S. J., Jiko, C., Yamashita, E., Tsukihara, T. Selective nuclear export mechanism of small RNAs. *Curr Opin Struct Biol*, **21**, 101–108 (2011).

[53] Leisegang, M. S., Martin, R., Ramírez, A. S., Bohnsack, M. T. Exportin T and Exportin 5: tRNA and miRNA biogenesis - and beyond. *Biol Chem*, **393**, 599–604 (2012).

[54] Okada, C., Yamashita, E., Lee, S. J., Shibata, S., Katahira, J., Nakagawa, A., Yoneda, Y., Tsukihara, T. A high-resolution structure of the pre-microRNA nuclear export machinery. *Science*, **326**, 1275–1279 (2009).

[55] Lund, E., Guettinger, S., Calado, A., Dahlberg, J., Kutay, U. Nuclear export of microRNA precursors. *Science*, **303**, 95–8 (2004).

[56] Bohnsack, M. T., Czaplinski, K., Gorlich, D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA*, **10**, 185–91 (2004).

[57] Gwizdek, C., Ossareh-Nazari, B., Brownawell, A. M., Doglio, A., Bertrand, E., Macara, I. G., Dargemont, C. Exportin-5 mediates nuclear export of minihelix-containing RNAs. *J Biol Chem*, **278**, 5505–5508 (2003).

[58] Bernstein, E., Caudy, A. A., Hammond, S. M., Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, **409**, 363–6 (2001).

[59] Hutvágner, G., McLachlan, J., Pasquinelli, A. E., Bálint, E., Tuschl, T., Zamore, P. D. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science*, **293**, 834–838 (2001).

[60] Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G., Mello, C. C. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell*, **106**, 23–34 (2001).

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

[62] Lau, P.-W., Guiley, K. Z., De, N., Potter, C. S., Carragher, B., MacRae, I. J. The molecular architecture of human Dicer. *Nat Struct Mol Biol*, **19**, 436–440 (2012).

[63] MacRae, I. J., Doudna, J. A. Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Curr Opin Struct Biol*, **17**, 138–145 (2007).

[64] Schwarz, D. S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., Zamore, P. D. Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, **115**, 199–208 (2003).

[65] Khvorova, A., Reynolds, A., Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**, 209–216 (2003).

[66] Okamura, K., Liu, N., Lai, E. C. Distinct mechanisms for microRNA strand selection by Drosophila Argonautes. *Mol Cell*, **36**, 431–444 (2009).

[67] Czech, B., Zhou, R., Erlich, Y., Brennecke, J., Binari, R., Villalta, C., Gordon, A., Perrimon, N., Hannon, G. J. Hierarchical rules for Argonaute loading in Drosophila. *Mol Cell*, **36**, 445–456 (2009).

[68] Haase, A. D., Jaskiewicz, L., Zhang, H., Lainé, S., Sack, R., Gatignol, A., Filipowicz, W. TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep*, **6**, 961–967 (2005).

[69] Lee, Y., Hur, I., Park, S.-Y., Kim, Y.-K., Suh, M. R., Kim, V. N. The role of PACT in the RNA silencing pathway. *EMBO J*, **25**, 522–532 (2006).

[70] Chendrimada, T. P., Gregory, R. I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., Shiekhattar, R. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, **436**, 740–744 (2005).

[71] Maniataki, E., Mourelatos, Z. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev*, **19**, 2979–2990 (2005).

[72] Noland, C. L., Ma, E., Doudna, J. A. siRNA repositioning for guide strand selection by human Dicer complexes. *Mol Cell*, **43**, 110–121 (2011).

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

[74] Meister, G., Landthaler, M., Peters, L., Chen, P. Y., Urlaub, H., Lührmann, R., Tuschl, T. Identification of novel argonaute-associated proteins. *Curr Biol*, **15**, 2149–2155 (2005).

[75] Robb, G. B., Rana, T. M. RNA helicase A interacts with RISC in human cells and functions in RISC loading. *Mol Cell*, **26**, 523–537 (2007).

[76] Salzman, D. W., Shubert-Coleman, J., Furneaux, H. P68 RNA helicase unwinds the human let-7 microRNA precursor duplex and is required for let-7-directed silencing of gene expression. *J Biol Chem*, **282**, 32773–32779 (2007).

[77] Matranga, C., Tomari, Y., Shin, C., Bartel, D. P., Zamore, P. D. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell*, **123**, 607–620 (2005).

[78] Rand, T. A., Petersen, S., Du, F., Wang, X. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell*, **123**, 621–629 (2005).

[79] Diederichs, S., Haber, D. A. Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell*, **131**, 1097–1108 (2007).

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

[81] Boland, A., Tritschler, F., Heimstädt, S., Izaurralde, E., Weichenrieder, O. Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO Rep*, **11**, 522–527 (2010).

[82] Djuranovic, S., Nahvi, A., Green, R. A parsimonious model for gene regulation by miRNAs. *Science*, **331**, 550–553 (2011).

[83] Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J.-J., Hammond, S. M., Joshua-Tor, L., Hannon, G. J. Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, **305**, 1437–1441 (2004).

[84] Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., Tuschl, T. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell*, **15**, 185–197 (2004).

[85] Chi, S. W., Zang, J. B., Mele, A., Darnell, R. B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486 (2009).

[86] Gu, S., Jin, L., Zhang, F., Sarnow, P., Kay, M. A. Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nat Struct Mol Biol*, **16**, 144–150 (2009).

[87] Fang, Z., Rajewsky, N. The impact of miRNA target sites in coding sequences and in 3'UTRs. *PLoS One*, **6**, e18067 (2011).

[88] Llave, C., Kasschau, K. D., Rector, M. A., Carrington, J. C. Endogenous and silencing-associated small RNAs in plants. *Plant Cell*, **14**, 1605–1619 (2002).

[89] Yekta, S., Shih, I.-H., Bartel, D. P. MicroRNA-directed cleavage of HOXB8 mRNA. *Science*, **304**, 594–596 (2004).

[90] Lages, E., Ipas, H., Guttin, A., Nesr, H., Berger, F., Issartel, J.-P. MicroRNAs: molecular features and role in cancer. *Front Biosci*, **17**, 2508–2540 (2012).

[91] Humphreys, D. T., Westman, B. J., Martin, D. I. K., Preiss, T. MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc Natl Acad Sci USA*, **102**, 16961–16966 (2005).

[92] Pillai, R. S., Bhattacharyya, S. N., Artus, C. G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., Filipowicz, W. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science*, **309**, 1573–1576 (2005).

[93] Kiriakidou, M., Tan, G. S., Lamprinaki, S., De Planell-Saguer, M., Nelson, P. T., Mourelatos, Z. An mRNA m7G cap binding-like motif within human Ago2 represses translation. *Cell*, **129**, 1141–1151 (2007).

[94] Chendrimada, T. P., Finn, K. J., Ji, X., Baillat, D., Gregory, R. I., Liebhaber, S. A., Pasquinelli, A. E., Shiekhattar, R. MicroRNA silencing through RISC recruitment of eIF6. *Nature*, **447**, 823–828 (2007).

[95] Petersen, C. P., Bordeleau, M.-E., Pelletier, J., Sharp, P. A. Short RNAs repress translation after initiation in mammalian cells. *Mol Cell*, **21**, 533–542 (2006).

[96] Maroney, P. A., Yu, Y., Fisher, J., Nilsen, T. W. Evidence that microRNAs are associated with translating messenger RNAs in human cells. *Nat Struct Mol Biol*, **13**, 1102–1107 (2006).

[97] Nottrott, S., Simard, M. J., Richter, J. D. Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat Struct Mol Biol*, **13**, 1108–1114 (2006).

[98] Olsen, P. H., Ambros, V. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol*, **216**, 671–680 (1999).

[99] Seggerson, K., Tang, L., Moss, E. G. Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation. *Dev Biol*, **243**, 215–25 (2002).

[100] Wu, L., Belasco, J. G. Micro-RNA regulation of the mammalian lin-28 gene during neuronal differentiation of embryonal carcinoma cells. *Mol Cell Biol*, **25**, 9198–9208 (2005).

[101] Guo, H., Ingolia, N. T., Weissman, J. S., Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, **466**, 835–840 (2010).

[102] Pasquinelli, A. E. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat Rev Genet*, **13**, 271–282 (2012).

[103] Parker, R., Sheth, U. P bodies and the control of mRNA translation and degradation. *Mol Cell*, **25**, 635–646 (2007).

[104] Liu, J., Rivas, F. V., Wohlschlegel, J., Yates, J. R., 3rd, Parker, R., Hannon, G. J. A role for the P-body component GW182 in microRNA function. *Nat Cell Biol*, **7**, 1261–1266 (2005).

[105] Liu, J., Valencia-Sanchez, M. A., Hannon, G. J., Parker, R. MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat Cell Biol*, **7**, 719–723 (2005).

[106] Yang, Z., Jakymiw, A., Wood, M. R., Eystathioy, T., Rubin, R. L., Fritzler, M. J., Chan, E. K. L. GW182 is critical for the stability of GW bodies expressed during the cell cycle and cell proliferation. *J Cell Sci*, **117**, 5567–5578 (2004).

[107] Eulalio, A., Behm-Ansmant, I., Schweizer, D., Izaurralde, E. P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol Cell Biol*, **27**, 3970–3981 (2007).

[108] Kulkarni, M., Ozgur, S., Stoecklin, G. On track with P-bodies. *Biochem Soc Trans*, **38**, 242–251 (2010).

[109] Fukuda, T., Yamagata, K., Fujiyama, S., Matsumoto, T., Koshida, I., Yoshimura, K., Mihara, M., Naitou, M., Endoh, H., Nakamura, T., Akimoto, C., Yamamoto, Y., Katagiri, T., Foulds, C., Takezawa, S., Kitagawa, H., Takeyama, K.-i., O'Malley, B. W., Kato, S. DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nat Cell Biol*, **9**, 604–611 (2007).

[110] Davis, B. N., Hilyard, A. C., Lagna, G., Hata, A. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature*, **454**, 56–61 (2008).

[111] Sakamoto, S., Aoki, K., Higuchi, T., Todaka, H., Morisawa, K., Tamaki, N., Hatano, E., Fukushima, A., Taniguchi, T., Agata, Y. The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway. *Mol Cell Biol*, **29**, 3754–3769 (2009).

[112] Suzuki, H. I., Yamagata, K., Sugimoto, K., Iwamoto, T., Kato, S., Miyazono, K. Modulation of microRNA processing by p53. *Nature*, **460**, 529–533 (2009).

[113] Yamagata, K., Fujiyama, S., Ito, S., Ueda, T., Murata, T., Naitou, M., Takeyama, K.-I., Minami, Y., O'Malley, B. W., Kato, S. Maturation of microRNA is hormonally regulated by a nuclear receptor. *Mol Cell*, **36**, 340–347 (2009).

[114] Pawlicki, J. M., Steitz, J. A. Nuclear networking fashions pre-messenger RNA and primary microRNA transcripts for function. *Trends Cell Biol*, **20**, 52–61 (2010).

[115] Gruber, J. J., Zatechka, D. S., Sabin, L. R., Yong, J., Lum, J. J., Kong, M., Zong, W.-X., Zhang, Z., Lau, C.-K., Rawlings, J., Cherry, S., Ihle, J. N., Dreyfuss, G., Thompson, C. B. Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. *Cell*, **138**, 328–339 (2009).

[116] Paroo, Z., Ye, X., Chen, S., Liu, Q. Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling. *Cell*, **139**, 112–122 (2009).

[117] Luciano, D. J., Mirsky, H., Vendetti, N. J., Maas, S. RNA editing of a miRNA precursor. *RNA*, **10**, 1174–1177 (2004).

[118] Yang, W., Chendrimada, T. P., Wang, Q., Higuchi, M., Seeburg, P. H., Shiekhattar, R., Nishikura, K. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol*, **13**, 13–21 (2006).

[119] Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A. G., Nishikura, K. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science*, **315**, 1137–1140 (2007).

[120] Kawahara, Y., Zinshteyn, B., Chendrimada, T. P., Shiekhattar, R., Nishikura, K. RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. *EMBO Rep*, **8**, 763–769 (2007).

[121] Scadden, A. D. J. The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nat Struct Mol Biol*, **12**, 489–496 (2005).

[122] Kawahara, Y., Megraw, M., Kreider, E., Iizasa, H., Valente, L., Hatzigeorgiou, A. G., Nishikura, K. Frequency and fate of microRNA editing in human brain. *Nucleic Acids Res*, **36**, 5270–5280 (2008).

[123] Guil, S., Cáceres, J. F. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat Struct Mol Biol*, **14**, 591–596 (2007).

[124] Michlewski, G., Guil, S., Semple, C. A., Cáceres, J. F. Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol Cell*, **32**, 383–393 (2008).

[125] Michlewski, G., Guil, S., Cáceres, J. F. Stimulation of pri-miR-18a processing by hnRNP A1. *Adv Exp Med Biol*, **700**, 28–35 (2010).

[126] Michlewski, G., Cáceres, J. F. Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nat Struct Mol Biol*, **17**, 1011–1018 (2010).

[127] Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A. D., Filipowicz, W., Ramos, A., Gherzi, R., Rosenfeld, M. G. The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature*, **459**, 1010–1014 (2009).

[128] Viswanathan, S. R., Daley, G. Q., Gregory, R. I. Selective blockade of microRNA processing by Lin28. *Science*, **320**, 97–100 (2008).

[129] Newman, M. A., Thomson, J. M., Hammond, S. M. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA*, **14**, 1539–49 (2008).

[130] Rybak, A., Fuchs, H., Smirnova, L., Brandt, C., Pohl, E., Nitsch, R., Wulczyn, F. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat Cell Biol*, **10**, 987–93 (2008).

[131] Heo, I., Joo, C., Cho, J., Ha, M., Han, J., Kim, V. N. Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell*, **32**, 276–84 (2008).

[132] Heo, I., Joo, C., Kim, Y.-K., Ha, M., Yoon, M.-J., Cho, J., Yeom, K.-H., Han, J., Kim, V. N. TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell*, **138**, 696–708 (2009).

[133] Ambros, V., Horvitz, H. Heterochronic mutants of the nematode Caenorhabditis elegans. *Science*, **226**, 409–416 (1984).

[134] Ambros, V. A hierarchy of regulatory genes controls a larva-to-adult developmental switch in C. elegans. *Cell*, **57**, 49–57 (1989).

[135] Roush, S., Slack, F. J. The let-7 family of microRNAs. *Trends Cell Biol*, **18**, 505–516 (2008).

[136] Boyerinas, B., Park, S.-M., Hau, A., Murmann, A. E., Peter, M. E. The role of let-7 in cell differentiation and cancer. *Endocr Relat Cancer*, **17**, F19–F36 (2010).

[137] Caygill, E. E., Johnston, L. A. Temporal regulation of metamorphic processes in Drosophila by the let-7 and miR-125 heterochronic microRNAs. *Curr Biol*, **18**, 943–950 (2008).

[138] Sokol, N. S., Xu, P., Jan, Y.-N., Ambros, V. Drosophila let-7 microRNA is required for remodeling of the neuromusculature during metamorphosis. *Genes Dev*, **22**, 1591–1596 (2008).

[139] Lancman, J. J., Caruccio, N. C., Harfe, B. D., Pasquinelli, A. E., Schageman, J. J., Pertsemlidis, A., Fallon, J. F. Analysis of the regulation of lin-41 during chick and mouse limb development. *Dev Dyn*, **234**, 948–960 (2005).

[140] Yu, J., Vodyanik, M., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J., Tian, S., Nie, J., Jonsdottir, G., Ruotti, V., Stewart, R., Slukvin, I., Thomson, J. Induced pluripotent stem cell lines derived from human somatic cells. *Science*, **318**, 1917–20 (2007).

[141] Wulczyn, F., Smirnova, L., Rybak, A., Brandt, C., Kwidzinski, E., Ninnemann, O., Strehle, M., Seiler, A., Schumacher, S., Nitsch, R. Post-transcriptional regulation of the let-7 microRNA during neural cell specification. *FASEB J*, **21**, 415–26 (2007).

[142] Johnson, S., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K., Brown, D., Slack, F. RAS is regulated by the let-7 microRNA family. *Cell*, **120**, 635–47 (2005).

[143] Shi, G., Perle, M. A., Mittal, K., Chen, H., Zou, X., Narita, M., Hernando, E., Lee, P., Wei, J.-J. Let-7 repression leads to HMGA2 overexpression in uterine leiomyosarcoma. *J Cell Mol Med*, **13**, 3898–3905 (2009).

[144] Esquela-Kerscher, A., Slack, F. J. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer*, **6**, 259–269 (2006).

[145] Mayr, C., Hemann, M. T., Bartel, D. P. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science*, **315**, 1576–9 (2007).

[146] Chang, S. S., Jiang, W. W., Smith, I., Poeta, L. M., Begum, S., Glazer, C., Shan, S., Westra, W., Sidransky, D., Califano, J. A. MicroRNA alterations in head and neck squamous cell carcinoma. *Int J Cancer*, **123**, 2791–2797 (2008).

[147] Darr, H., Benvenisty, N. Genetic analysis of the role of the reprogramming gene LIN-28 in human embryonic stem cells. *Stem Cells*, **27**, 352–362 (2009).

[148] Zhu, H., Shah, S., Shyh-Chang, N., Shinoda, G., Einhorn, W., Viswanathan, S., Takeuchi, A., Grasemann, C., Rinn, J., Lopez, M., Hirschhorn, J., Palmert, M., Daley, G. Lin28a transgenic mice manifest size and puberty phenotypes identified in human genetic association studies. *Nat Genet*, **42**, 626–30 (2010).

[149] Zhu, H., Shyh-Chang, N., Segrè, A. V., Shinoda, G., Shah, S. P., Einhorn, W. S., Takeuchi, A., Engreitz, J. M., Hagan, J. P., Kharas, M. G., Urbach, A., Thornton, J. E., Triboulet, R., Gregory, R. I., Altshuler, D., Daley, G. Q. The Lin28/let-7 axis regulates glucose metabolism. *Cell*, **147**, 81–94 (2011).

[150] Huang, Y. A mirror of two faces: Lin28 as a master regulator of both miRNA and mRNA. *Wiley Interdiscip Rev RNA*, **3**, 483–494 (2012).

[151] Piskounova, E., Polytarchou, C., Thornton, J. E., LaPierre, R. J., Pothoulakis, C., Hagan, J. P., Iliopoulos, D., Gregory, R. I. Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. *Cell*, **147**, 1066–1079 (2011).

[152] Moss, E. G., Tang, L. Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. *Dev Biol*, **258**, 432–42 (2003).

[153] Balzer, E., Moss, E. G. Localization of the developmental timing regulator Lin28 to mRNP complexes, P-bodies and stress granules. *RNA Biol*, **4**, 16–25 (2007).

[154] Yang, D. H., Moss, E. G. Temporally regulated expression of Lin-28 in diverse tissues of the developing mouse. *Gene Expr Patterns*, **3**, 719–26 (2003).

[155] Suh, M.-R., Lee, Y., Kim, J. Y., Kim, S.-K., Moon, S.-H., Lee, J. Y., Cha, K.-Y., Chung, H. M., Yoon, H. S., Moon, S. Y., Kim, V. N., Kim, K.-S. Human embryonic stem cells express a unique set of microRNAs. *Dev Biol*, **270**, 488–498 (2004).

[156] Thomson, J. M., Newman, M., Parker, J. S., Morin-Kensicki, E. M., Wright, T., Hammond, S. M. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev*, **20**, 2202–2207 (2006).

[157] Piskounova, E., Viswanathan, S. R., Janas, M., LaPierre, R. J., Daley, G. Q., Sliz, P., Gregory, R. I. Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. *J Biol Chem*, **283**, 21310–4 (2008).

[158] Lightfoot, H. L., Bugaut, A., Armisen, J., Lehrbach, N. J., Miska, E. A., Balasubramanian, S. A LIN28-dependent structural change in pre-let-7g directly inhibits dicer processing. *Biochemistry*, **50**, 7514–21 (2011).

[159] Shen, B., Goodman, H. M. Uridine addition after microRNA-directed cleavage. *Science*, **306**, 997 (2004).

[160] Mullen, T. E., Marzluff, W. F. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5′ to 3′ and 3′ to 5′. *Genes Dev*, **22**, 50–65 (2008).

[161] Hagan, J., Piskounova, E., Gregory, R. Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat Struct Mol Biol*, **16**, 1021–5 (2009).

[162] Heo, I., Ha, M., Lim, J., Yoon, M.-J., Park, J.-E., Kwon, S. C., Chang, H., Kim, V. N. Mono-Uridylation of Pre-MicroRNA as a Key Step in the Biogenesis of Group II let-7 MicroRNAs. *Cell*, **151**, 521–532 (2012).

[163] Lehrbach, N. J., Armisen, J., Lightfoot, H. L., Murfitt, K. J., Bugaut, A., Balasubramanian, S., Miska, E. A. LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in Caenorhabditis elegans. *Nat Struct Mol Biol*, **16**, 1016–20 (2009).

[164] Van Wynsberghe, P., Kai, Z., Massirer, K., Burton, V., Yeo, G., Pasquinelli, A. LIN-28 co-transcriptionally binds primary let-7 to regulate miRNA maturation in Caenorhabditis elegans. *Nat Struct Mol Biol*, **18**, 302–8 (2011).

[165] Thornton, J. E., Gregory, R. I. How does Lin28 let-7 control development and disease? *Trends Cell Biol*, **22**, 474–482 (2012).

[166] Polesskaya, A., Cuvellier, S., Naguibneva, I., Duquet, A., Moss, E., Harel-Bellan, A. Lin-28 binds IGF-2 mRNA and participates in skeletal myogenesis by increasing translation efficiency. *Genes Dev*, **21**, 1125–38 (2007).

[167] Xu, B., Huang, Y. Histone H2a mRNA interacts with Lin28 and contains a Lin28-dependent posttranscriptional regulatory element. *Nucleic Acids Res*, **37**, 4256–63 (2009).

[168] Xu, B., Zhang, K., Huang, Y. Lin28 modulates cell growth and associates with a subset of cell cycle regulator mRNAs in mouse embryonic stem cells. *RNA*, **15**, 357–61 (2009).

[169] Peng, S., Maihle, N. J., Huang, Y. Pluripotency factors Lin28 and Oct4 identify a sub-population of stem cell-like cells in ovarian cancer. *Oncogene*, **29**, 2153–9 (2010).

[170] Lei, X.-X., Xu, J., Ma, W., Qiao, C., Newman, M. A., Hammond, S. M., Huang, Y. Determinants of mRNA recognition and translation regulation by Lin28. *Nucleic Acids Res*, **40**, 3574–3584 (2012).

[171] Peng, S., Chen, L.-L., Lei, X.-X., Yang, L., Lin, H., Carmichael, G. G., Huang, Y. Genome-wide studies reveal that Lin28 enhances the translation of genes important for growth and survival of human embryonic stem cells. *Stem Cells*, **29**, 496–504 (2011).

[172] Jin, J., Jing, W., Lei, X.-X., Feng, C., Peng, S., Boris-Lawrie, K., Huang, Y. Evidence that Lin28 stimulates translation by recruiting RNA helicase A to polysomes. *Nucleic Acids Res*, **39**, 3724–34 (2011).

[173] Wilbert, M. L., Huelga, S. C., Kapeli, K., Stark, T. J., Liang, T. Y., Chen, S. X., Yan, B. Y., Nathanson, J. L., Hutt, K. R., Lovci, M. T., Kazan, H., Vu, A. Q., Massirer, K. B., Morris, Q., Hoon, S., Yeo, G. W. LIN28 Binds Messenger RNAs at GGAGA Motifs and Regulates Splicing Factor Abundance. *Mol Cell*, **48**, 195–206 (2012).

[174] Cho, J., Chang, H., Kwon, S. C., Kim, B., Kim, Y., Choe, J., Ha, M., Kim, Y. K., Kim, V. N. LIN28A is a Suppressor of ER-Associated Translation in Embryonic Stem Cells. *Cell*, **151**, 765–777 (2012).

[175] Mihailovich, M., Militti, C., Gabaldón, T., Gebauer, F. Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. *Bioessays*, **32**, 109–18 (2010).

[176] Horn, G., Hofweber, R., Kremer, W., Kalbitzer, H. R. Structure and function of bacterial cold shock proteins. *Cell Mol Life Sci*, **64**, 1457–70 (2007).

[177] Ermolenko, D. N., Makhatadze, G. I. Bacterial cold-shock proteins. *Cell Mol Life Sci*, **59**, 1902–13 (2002).

[178] Lopez, M. M., Makhatadze, G. I. Major cold shock proteins, CspA from Escherichia coli and CspB from Bacillus subtilis, interact differently with single-stranded DNA templates. *Biochim Biophys Acta*, **1479**, 196–202 (2000).

[179] Lopez, M. M., Yutani, K., Makhatadze, G. I. Interactions of the cold shock protein CspB from Bacillus subtilis with single-stranded DNA. Importance of the T base content and position within the template. *J Biol Chem*, **276**, 15511–8 (2001).

[180] Max, K. E. A., Zeeb, M., Bienert, R., Balbach, J., Heinemann, U. T-rich DNA single strands bind to a preformed site on the bacterial cold shock protein Bs-CspB. *J Mol Biol*, **360**, 702–14 (2006).

[181] Zeeb, M., Max, K. E. A., Weininger, U., Löw, C., Sticht, H., Balbach, J. Recognition of T-rich single-stranded DNA by the cold shock protein Bs-CspB in solution. *Nucleic Acids Res*, **34**, 4561–71 (2006).

[182] Max, K. E. A., Wunderlich, M., Roske, Y., Schmid, F. X., Heinemann, U. Optimized variants of the cold shock protein from in vitro selection: structural basis of their high thermostability. *J Mol Biol*, **369**, 1087–97 (2007).

[183] Morgan, H. P., Estibeiro, P., Wear, M. A., Max, K. E. A., Heinemann, U., Cubeddu, L., Gallagher, M. P., Sadler, P. J., Walkinshaw, M. D. Sequence specificity of single-stranded DNA-binding proteins: a novel DNA microarray approach. *Nucleic Acids Res*, **35**, e75 (2007).

[184] Phadtare, S., Tyagi, S., Inouye, M., Severinov, K. Three amino acids in Escherichia coli CspE surface-exposed aromatic patch are critical for nucleic acid melting activity leading to transcription antitermination and cold acclimation of cells. *J Biol Chem*, **277**, 46706–11 (2002).

[185] Phadtare, S., Severinov, K. Nucleic acid melting by Escherichia coli CspE. *Nucleic Acids Res*, **33**, 5583–90 (2005).

[186] Bae, W., Xia, B., Inouye, M., Severinov, K. Escherichia coli CspA-family RNA chaperones are transcription antiterminators. *Proc Natl Acad Sci USA*, **97**, 7784–7789 (2000).

[187] Sachs, R., Max, K. E. A., Heinemann, U., Balbach, J. RNA single strands bind to a conserved surface of the major cold shock protein in crystals and solution. *RNA*, **18**, 65–76 (2012).

[188] Guerrerio, A. L., Berg, J. M. Design of single-stranded nucleic acid binding peptides based on nucleocapsid CCHC-box zinc-binding domains. *J Am Chem Soc*, **132**, 9638–43 (2010).

[189] Green, L. M., Berg, J. M. A retroviral Cys-Xaa2-Cys-Xaa4-His-Xaa4-Cys peptide binds metal ions: spectroscopic studies and a proposed three-dimensional structure. *Proc Natl Acad Sci USA*, **86**, 4047–4051 (1989).

[190] South, T. L., Blake, P. R., Sowder, R., Arthur, L. O., Henderson, L. E., Summers, M. F. The nucleocapsid protein isolated from HIV-1 particles binds zinc and forms retroviral-type zinc fingers. *Biochemistry*, **29**, 7786–7789 (1990).

[191] De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., Summers, M. F. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science*, **279**, 384–8 (1998).

[192] Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K. J., Wu, Z. R., Summers, M. F. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J Mol Biol*, **301**, 491–511 (2000).

[193] Amarasinghe, G. K., Zhou, J., Miskimon, M., Chancellor, K. J., McDonald, J. A., Matthews, A. G., Miller, R. R., Rouse, M. D., Summers, M. F. Stem-loop SL4 of the HIV-1 psi RNA packaging signal exhibits weak affinity for the nucleocapsid protein. *J Mol Biol*, **314**, 961–70 (2001).

[194] Desjardins, A., Yang, A., Bouvette, J., Omichinski, J. G., Legault, P. Importance of the NCp7-like domain in the recognition of pre-let-7g by the pluripotency factor Lin28. *Nucleic Acids Res*, **40**, 1767–1777 (2012).

[195] Scheich, C., Kuemmel, D., Soumailakakis, D., Heinemann, U., Buessow, K. Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Res*, **35**, e43 (2007).

[196] Sambrook, J., Russel, D. W. F. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, N.Y. (2001).

[197] Chung, C. T., Niemela, S. L., Miller, R. H. One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci USA*, **86**, 2172–2175 (1989).

[198] Rial, D. V., Ceccarelli, E. A. Removal of DnaK contamination during fusion protein purifications. *Protein Expr Purif*, **25**, 503–507 (2002).

[199] Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, **72**, 248–254 (1976).

[200] Lowry, O. H., Rosebrough, N., Farr, A. L., Randall, R. J. Protein measurement with the Folin phenol reagent. *J Biol Chem*, **193**, 265–275 (1951).

[201] Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M., Appel, R., Bairoch, A. *Protein identification and analyses tools on the ExPASy Server*. Humana Press (2005).

[202] Schmid, F. X. Spectral methods of characterizing protein conformation and conformational changes. *Protein structure. A practical approach, 252-285* (1989).

[203] Compton, L. A., Johnson, W., Jr. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal Biochem*, **155**, 155–167 (1986).

[204] Manavalan, P., Johnson, W., Jr. Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal Biochem*, **167**, 76–85 (1987).

[205] Sreerama, N., Woody, R. W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem*, **287**, 252–260 (2000).

[206] Whitmore, L., Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*, **89**, 392–400 (2008).

[207] Eftink, M. R. Fluorescence methods for studying equilibrium macromolecule-ligand interactions. *Fluorescence Spectroscopy*, **278**, 221–257 (1997).

[208] Lohman, T. M., Bujalowski, W. Thermodynamic Methods for Model-Independent Determination of Equilibrium Binding Isotherms for Protein-DNA Interactions - Spectroscopic Approaches to Monitor Binding. *Methods in enzymology*, **208**, 258–290 (1991).

[209] Towbin, H., Staehelin, T., Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Biotechnology*, **24**, 145–149 (1992).

[210] Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res*, **35**, W43–W46 (2007).

[211] Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, **31**, 3406–3415 (2003).

[212] Mueller, U., Nyarsik, L., Horn, M., Rauth, H., Przewieslik, T., Saenger, W., Lehrach, H., Eickhoff, H. Development of a technology for automation and miniaturization of protein crystallization. *J Biotechnol*, **85**, 7–14 (2001).

[213] Heinemann, U., Buessow, K., Mueller, U., Umbach, P. Facilities and methods for the high-throughput crystal structural analysis of human proteins. *Acc Chem Res*, **36**, 157–63 (2003).

[214] Leslie, A. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography, No. 26.* (1992).

[215] Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr*, **66**, 125–132 (2010).

[216] Otwinowski, Z., Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Part A*, **276**, 307–326 (1997).

[217] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., Read, R. J. Phaser crystallographic software. *J Appl Crystallogr*, **40**, 658–674 (2007).

[218] Perrakis, A., Langer, G., Cohen, S. X., Lamzin, V. S. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature protocols*, **3**, 1171–1179 (2008).

[219] Emsley, P., Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography*, **60**, 2126–2132 (2004).

[220] Winn, M. D., Isupov, M. N., Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr*, **57**, 122–133 (2001).

[221] Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr*, **67**, 355–367 (2011).

[222] Painter, J., Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr*, **62**, 439–450 (2006).

[223] Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., Richardson, D. C. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*, **35**, W375–W383 (2007).

[224] Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **66**, 12–21 (2010).

[225] Ferrin, T. E., Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, **25**, 1605–1612 (2004).

[226] Barrios-Rodiles, M., Brown, K. R., Ozdamar, B., Bose, R., Liu, Z., Donovan, R. S., Shinjo, F., Liu, Y., Dembowy, J., Taylor, I. W., Luga, V., Przulj, N., Robinson, M., Suzuki, H., Hayashizaki, Y., Jurisica, I., Wrana, J. L. High-throughput mapping of a dynamic signaling network in mammalian cells. *Science*, **307**, 1621–1625 (2005).

[227] Nam, Y., Chen, C., Gregory, R. I., Chou, J. J., Sliz, P. Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell*, **147**, 1080–91 (2011).

[228] Loughlin, F. E., Gebert, L. F., Towbin, H., Brunschweiger, A., Hall, J., Allain, F. H. Structural basis of pre-let-7 miRNA recognition by the zinc knuckles of pluripotency factor Lin28. *Nat Struct Mol Biol*, **19**, 84–9 (2012).

[229] Zeeb, M., Balbach, J. Single-stranded DNA binding of the cold-shock protein CspB from Bacillus subtilis: NMR mapping and mutational characterization. *Protein Sci*, **12**, 112–23 (2003).

[230] Matthews, B. W. Solvent content of protein crystals. *J Mol Biol*, **33**, 491–497 (1968).

[231] Langer, G., Cohen, S., Lamzin, V. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature protocols* (2008).

[232] Schindelin, H., Jiang, W., Inouye, M., Heinemann, U. Crystal structure of CspA, the major cold shock protein of Escherichia coli. *Proc Natl Acad Sci USA*, **91**, 5119–23 (1994).

[233] Baker, N. A., Sept, D., Joseph, S., Holst, M. J., McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA*, **98**, 10037–41 (2001).

[234] Bienert, R., Zeeb, M., Dostal, L., Feske, A., Magg, C., Max, K., Welfle, H., Balbach, J., Heinemann, U. Single-stranded DNA bound to bacterial cold-shock proteins: preliminary crystallographic and Raman analysis. *Acta Crystallogr D Biol Crystallogr*, **60**, 755–7 (2004).

[235] Shafer, R. H., Smirnov, I. Biological aspects of DNA/RNA quadruplexes. *Biopolymers*, **56**, 209–227 (2000).

[236] Shaik Syed Ali, P., Ghoshdastider, U., Hoffmann, J., Brutschy, B., Filipek, S. Recognition of the let-7g miRNA precursor by human Lin28B. *FEBS Lett*, **586**, 3986–3990 (2012).

[237] Phadtare, S., Inouye, M., Severinov, K. The mechanism of nucleic acid melting by a CspA family protein. *J Mol Biol*, **337**, 147–55 (2004).

[238] Phadtare, S. Unwinding activity of cold shock proteins and RNA metabolism. *RNA biology*, **8** (2011).

[239] Thornton, J. E., Chang, H.-M., Piskounova, E., Gregory, R. I. Lin28-mediated control of let-7 microRNA expression by alternative TUTases Zcchc11 (TUT4) and Zcchc6 (TUT7). *RNA*, **18**, 1875–1885 (2012).

[240] Han, T. W., Kato, M., Xie, S., Wu, L. C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G., McKnight, S. L. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell*, **149**, 768–779 (2012).

[241] Yates, L. A., Fleurdépine, S., Rissland, O. S., De Colibus, L., Harlos, K., Norbury, C. J., Gilbert, R. J. C. Structural basis for the activity of a cytoplasmic RNA terminal uridylyl transferase. *Nat Struct Mol Biol*, **19**, 782–787 (2012).

[242] Rupp, B. *Biomolecular Crystallography: Principles, Practice, and Applications to Structural Biology*. Taylor & Francis Group, 1st edition (2010).

[243] Drenth, J., Mesters, J. *Principles of Protein X-Ray Crystallography*. Springer Verlag, 3rd edition (2007).

[244] Rhodes, G. *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*. Academic Press, 3rd edition (2006).

[245] Diederichs, K., Karplus, P. A. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol*, **4**, 269–275 (1997).

[246] Hendrickson, W. A., Horton, J. R., LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J*, **9**, 1665–1672 (1990).

[247] Cronin, C. N., Lim, K. B., Rogers, J. Production of selenomethionyl-derivatized proteins in baculovirus-infected insect cells. *Protein Sci*, **16**, 2023–2029 (2007).

[248] Correll, C. C., Freeborn, B., Moore, P. B., Steitz, T. A. Use of chemically modified nucleotides to determine a 62-nucleotide RNA crystal structure: a survey of phosphorothioates, Br, Pt and Hg. *J Biomol Struct Dyn*, **15**, 165–172 (1997).

ACKNOWLEDGMENTS

Foremost, I want to thank my PhD supervisor Prof Dr. Udo Heinemann for the opportunity to perform my PhD thesis in his lab and for his support and guidance during my entire work at the MDC Berlin. I am very grateful for his advice, his patience and his motivation.

Next, I want to thank present and past members of the Heinemann group for sharing their knowledge and being open for discussing scientific issues and problems at any time. Thanks to Kerstin, Harald, Jennifer, Chengcheng, Claudia, David, Annette, Ulf, Anja, Janett, Tracy, Yvette, Jürgen, Uli, Andreas, Silke, Heide, Sofia, Anup, Michaela and Maria. Special thanks to Yvette, Jürgen and Uli, who helped me in collecting and processing my crystallographic data sets. I further want to thank Anja, Janett and Tracy, who started the Lin28 project and continued working on it with me. I further would like to thank Maria, Michaela, Nadine and Karin for the excellent technical assistance and work they have done.

Special thanks go also to the Daumke group. I enjoyed it very much with you guys - all the scientific and non-scientific discussions, group seminars, coffee and beer hours. You made work and life at the MDC so much easier.

Being part of the international PhD program of the MDC Berlin was a very special and unforgettable experience for me. I want to thank all my colleagues and fellows, for an unbelievable time - many of you have become really good friends. Thanks to Kristin, Julia, Claudio, Flo, David, Chris, Kathrin, Vreni, Eva, Tilman, Karina and all others that I have not named in person here.

I further want to thank Wiebke, David and Udo for proofreading my manuscript and Claudio for his help with Latex.

Last but not least, thousand thanks to my parents, my sister, my brother and their families. Dear Theo, Edith, Marion, Christian, Roland, Katha, Philipp, Angi, Tim und Lea - thanks for your support, security and love.

Last but not least, I would like to thank you Wiebke. You enriched my life with warmness, security and love. I thank for every day, we have together.

DECLARATION

I hereby declare that the present thesis *Structural and Functional Analysis of Lin28-mediated Inhibition of let-7 miRNA biogenesis* is my own work and effort has not been submitted to any other university for any award. Where other sources of information have been used, they have been acknowledged.

Berlin, January 2013

Florian Mayr