

Molecular Basis for Interaction of *let-7* MicroRNAs with Lin28

Yunsun Nam,^{1,2,3} Casandra Chen,^{1,2,3} Richard I. Gregory,^{1,3} James J. Chou,¹ and Piotr Sliz^{1,2,3,*}

¹Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

²Laboratory of Molecular Medicine

³Stem Cell Program

Children's Hospital, Boston, MA 02115, USA

*Correspondence: piotr_sliz@hms.harvard.edu

DOI 10.1016/j.cell.2011.10.020

SUMMARY

MicroRNAs (miRNAs) are small noncoding RNA molecules that regulate gene expression. Among these, members of the *let-7* miRNA family control many cell-fate determination genes to influence pluripotency, differentiation, and transformation. Lin28 is a specific, posttranscriptional inhibitor of *let-7* biogenesis. We report crystal structures of mouse Lin28 in complex with sequences from *let-7d*, *let-7-f1*, and *let-7g* precursors. The two folded domains of Lin28 recognize two distinct regions of the RNA and are sufficient for inhibition of *let-7* in vivo. We also show by NMR spectroscopy that the linker connecting the two folded domains is flexible, accommodating Lin28 binding to diverse *let-7* family members. Protein-RNA complex formation imposes specific conformations on both components that could affect downstream recognition by other processing factors. Our data provide a molecular explanation for Lin28 specificity and a model for how it regulates *let-7*.

INTRODUCTION

Since the discovery of the first human microRNAs (miRNAs) about a decade ago, examples of miRNA regulation have been found for virtually every cellular process (Kim et al., 2009; Krol et al., 2010). Precursors of miRNAs undergo a series of processing steps after transcription to generate an active product. In this canonical pathway, a newly transcribed primary miRNA (pri-miRNA) with at least one hairpin structure is cleaved within the nucleus by an RNaseIII enzyme, Drosha, that acts in complex with DGCR8. The resulting pre-miRNA is exported to the cytoplasm, where another RNaseIII, Dicer, removes the “terminal loop region,” or pre-element (preE), to yield the mature miRNA (Figure 1A). Mechanisms of transcriptional control have been analyzed for many miRNAs, but the recent identification of post-transcriptional regulators of miRNA biogenesis now provides a way to investigate the molecular details of miRNA maturation and regulation (Davis and Hata, 2010; Siomi and Siomi, 2010).

The *let-7* family of miRNAs regulates many factors that control cell-fate decisions, including oncogenes (c-Myc, Ras, HMGA-2) and cell-cycle factors (CyclinD1, D2) (Büssing et al., 2008; Viswanathan and Daley, 2010). Deregulation of *let-7* influences tumorigenicity of breast cancer stem cells (Yu et al., 2007a). Moreover, IL-6 is a target of *let-7*, thereby bridging the inflammation and cell-transformation signaling pathways (Iliopoulos et al., 2009). There are several *let-7* family members in mammals, with similar mature regions but divergent sequences in the preE removed by Dicer (Figure 1A). Despite low sequence identity, most preEs in *let-7* are predicted to contain conserved structural elements (stem, bulge, and loop) that may be important for regulation of pre-miRNAs (Figure 1B).

LIN28, originally discovered as a heterochronic gene regulating developmental timing in worms (Moss et al., 1997), blocks *let-7* biogenesis (Heo et al., 2008; Lehrbach et al., 2009; Newman et al., 2008; Rybak et al., 2008; Viswanathan et al., 2008). Its effects on gene expression are profound enough to make Lin28 one of the four factors sufficient to reprogram human somatic cells into induced pluripotent stem cells (iPSCs) (Yu et al., 2007b). Lin28 is activated in many human tumors (~15%) and appears to be associated with less differentiated cancers (Viswanathan et al., 2009). Studies with patient samples show correlation between overexpression or mutation of Lin28 with ovarian cancer (Peng et al., 2010; Permuth-Wey et al., 2011) and colon cancer (King et al., 2011). Variations in Lin28 have also been linked to developmental traits such as height and timing of puberty onset in humans and mice (Lettre et al., 2008; Lu et al., 2009; Ong et al., 2009; Perry et al., 2009; Sulem et al., 2009; Zhu et al., 2010).

Because it is one of few specific inhibitors of miRNA maturation to be discovered thus far, understanding Lin28 activity provides an avenue for investigating the mechanisms of miRNA biogenesis and regulation. Lin28 contains two well-known nucleic acid interaction domains—a cold shock domain (CSD) and two tandem Cys-Cys-His-Cys (CCHC)-type zinc-binding motifs (CCHCx2). Mammals have two paralogs, Lin28a and Lin28b, with different physiological expression patterns but similar behavior in vitro (Guo et al., 2006; Heo et al., 2008; Viswanathan et al., 2008; Yang and Moss, 2003). Lin28 binds precursor forms of *let-7* miRNAs and can inhibit both pri-*let-7* processing by Drosha (Newman et al., 2008; Viswanathan et al., 2008) and pre-*let-7* processing by Dicer (Heo et al., 2008; Lehrbach et al.,

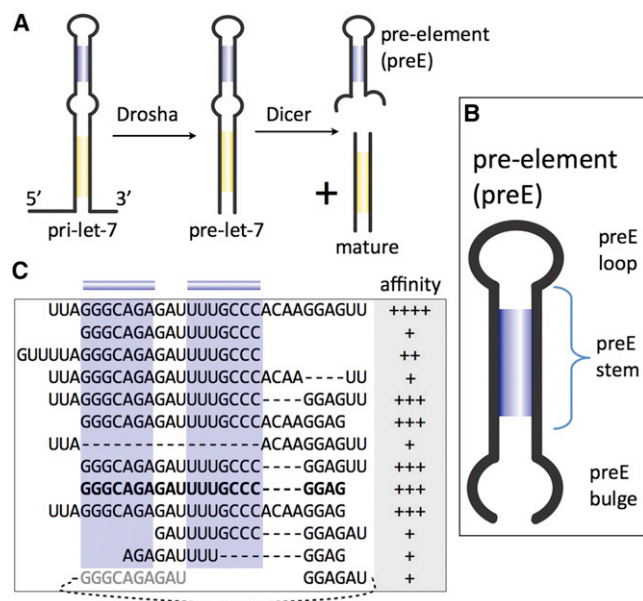


Figure 1. Mapping of Lin28-Binding Sites on Pre-let-7

(A) Processing steps in canonical miRNA biogenesis. (B) Architecture of preEs. (C) Fragments of preE-let-7d tested on EMSA for association with Lin28. Affinity is indicated by apparent dissociation constant ranges: +++++, 0.2–1.5 μ M; +++, 1.5–3 μ M; ++, 3–15 μ M; +, >15 μ M. Predicted stem is highlighted in blue. Minimal fragment (preE_M) identified is bolded. See also Figure S1.

2009; Rybak et al., 2008). Furthermore, Lin28 can recruit a terminal uridylyl transferase (TUTase) that adds uridine to the 3' end of pre-miRNA to increase decay (Hagan et al., 2009; Heo et al., 2009; Lehrbach et al., 2009). Although parts of the preE segment are dispensable for pri-miRNA processing by Drosha (Han et al., 2006), point mutations in the preE can disrupt interactions with Lin28 (Heo et al., 2009; Lehrbach et al., 2009; Newman et al., 2008; Piskounova et al., 2008), thereby derepressing Drosha-mediated processing (Newman et al., 2008). Sequence variability among preEs in *let-7* (Figure S1A available online) has hindered interpretation of these results and extension of the conclusions to other *let-7*s, highlighting the need for an atomic-level view of divergent Lin28:*let-7* complexes.

We present here high-resolution crystal structures of mouse Lin28a in complex with three preE constructs of *let-7d*, *let-7f-1*, and *let-7g*. These structures provide a direct view of a protein interacting with the terminal loop region of a miRNA. We identify sequence-specific interactions between Lin28 and *let-7* precursors that give direct structural evidence for the role of preEs in miRNA regulation. The Lin28 CSD and the CCHC “zinc knuckles” make extensive contacts with the preE elements in two distinct regions. We also describe NMR studies and biochemical assays showing that the linker between the CSD and CCHCx2 regions introduces flexibility to accommodate variable preE sequences and lengths while preserving the joint contribution of the two interaction sites to overall affinity. We show that neither the terminal nor linker regions outside of the folded domains are essential for blocking *let-7* in vivo. Mutagen-

esis of preE fragments and full-length pre-miRNA molecules confirms our conclusions from the structure concerning specificity of Lin28 and allows us to predict how Lin28 recognizes other *let-7*s. Complex formation induces in both Lin28 and preE-*let-7* a specific conformation that can affect recognition by downstream factors such as Drosha, Dicer, and TUTase, and changes in the CCHCx2 region are particularly detrimental to Lin28 activity in vivo.

RESULTS

Two Discrete Binding Sites in Pre-let-7 for Lin28 Binding

As a first step to understanding how pre-*let-7* is recognized by Lin28, we tested a series of deletions in pre-*let-7d* for binding to the protein. Pre-*let-7d* has a relatively high affinity for Lin28 both in vivo and in vitro (Hagan et al., 2009; Heo et al., 2009; Newman et al., 2008), and secondary structure predictions indicate that it has the most stable preE stem among mouse pre-*let-7*s, without interrupting bulges (Markham and Zuker, 2005). We focused our analysis on the preE, as mutagenesis studies had indicated its importance in direct association with Lin28 (Heo et al., 2009; Newman et al., 2008; Piskounova et al., 2008; Rybak et al., 2008). We observed that an isolated preE segment, containing none of the mature-region nucleotides, can bind Lin28 and that two distinct regions are critical for binding to Lin28, thereby defining a minimal preE-*let-7d* (preE_M-*let-7d*) sufficient for high-affinity binding (Figures 1C, S1B, and S1C). The first required region includes the preE stem and the preE loop; truncating the stem reduces binding. The other is the GGAG motif, which occurs at the 3' end of the preE bulge. Although overall preE sequence conservation is low, even within the preE stem and loop, the GGAG tetranucleotide element is well conserved throughout the *let-7* family (Figure S1A). Our mapping results suggest that the GGAG element provides an independent binding site, as deleting the neighboring nucleotides, thereby altering the distance to the CSD-binding site, does not abolish Lin28 binding. The presence of two independent binding sites explains how diverse preE-*let-7*s containing variable linker sequences can all bind Lin28 with high specificity and affinity.

Domains of Lin28 Tethered to Each Other Are Sufficient for Inhibiting *let-7* Processing

Lin28 has two folded regions, the CSD and the CCHCx2, connected by a positively charged linker of ~15 amino acids, with extensions of ~30 residues at both the amino and carboxyl termini. Mutagenesis studies have implicated both folded domains in repression of *let-7* (Heo et al., 2009; Piskounova et al., 2008). The region C-terminal to the CCHCx2 domain also promotes translation of certain mRNA targets (Jin et al., 2011; Peng et al., 2011; Qiu et al., 2009). Using limited proteolysis and electrophoretic mobility shift assay (EMSA), we analyzed a series of truncation constructs of Lin28 to identify the essential region for interaction with preE-*let-7*. Both the N- and C-terminal regions can be removed without affecting affinity for RNA, but removal of either the CSD or the CCHCx2 abolishes high-affinity preE-*let-7* binding (Figure S2A).

We used NMR spectroscopy to study the dynamics of Lin28:preE_M-*let-7d* complexes in more detail (Figure 2A). We

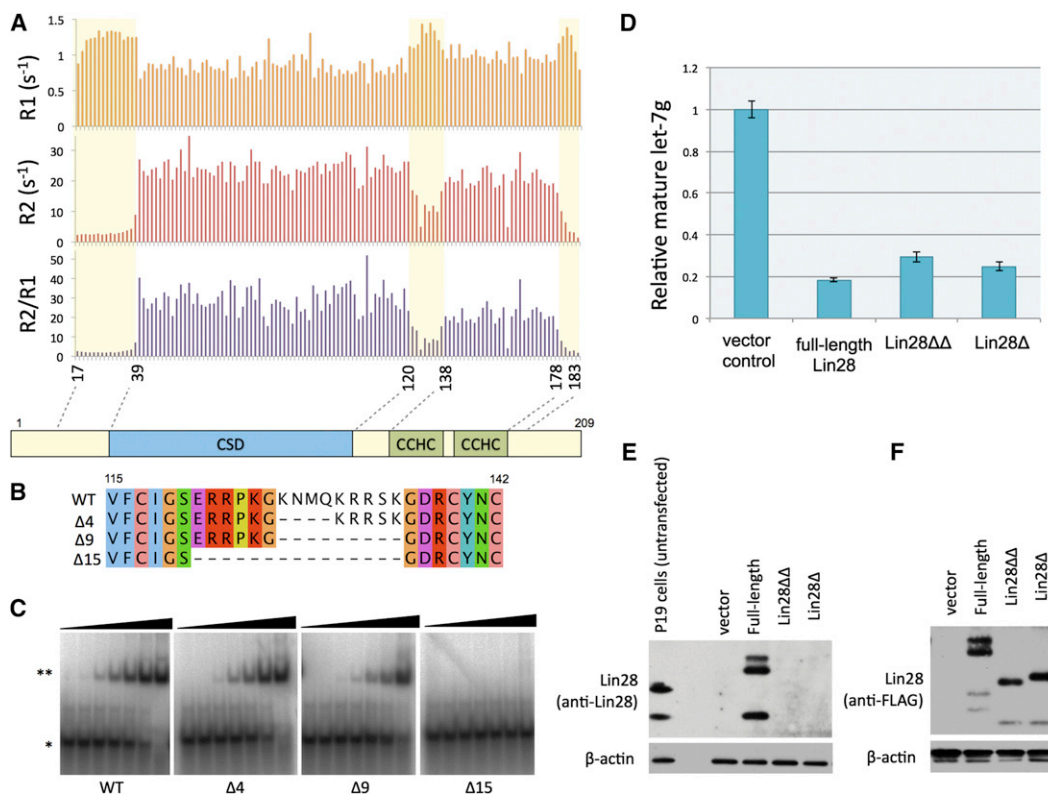


Figure 2. Linker between the CSD and CCHC \times 2 Is Flexible; See also Figure S2

(A) Longitudinal (R_1) and transverse (R_2) relaxation rates and the ratio (R_2/R_1), plotted against the residue number. Relatively more dynamic regions are marked with a light yellow box.

(B) Alignment of internal deletions in the linker, indicated with the number of amino acids deleted on left.

(C) EMSAs with preE-*let-7d* as probe, mixed with increasing concentrations (0.005, 0.02, 0.08, 0.3, 1.2, 5, 20 μ M) of linker deletion constructs of Lin28 (16–184): *, free probe; **, complex.

(D) Quantitative RT-PCR results for in vivo levels of mature *let-7g*. Lin28 Δ is truncated at both N and C termini. Lin28 $\Delta\Delta$ has both of the terminal extensions and the linker removed. The standard deviation is calculated from triplicate experiments; U6 RNA levels were used for normalization.

(E and F) Western blots of Trizol bottom layer for transfections shown in (D). Anti-Lin28 antibodies do not recognize truncation constructs, so anti-FLAG was used in (F) to compare the relative expression levels of different Lin28 constructs.

measured longitudinal (R_1) and transverse (R_2) relaxation rates to probe backbone dynamics. The R_2/R_1 ratio, which is a measure of correlation time, is an indicator of tumbling rate in solution. This ratio is similar for the folded domains but much lower for the terminal segments and the intervening linker, indicating more rapid motion in those regions. We conclude that the linker sequence lacks secondary structure, an inference consistent with absence of inter-residue backbone NOE cross-peaks in ^{15}N -NOESY (Figure S2B). Comparing the α , β , and C' chemical shifts to random coil chemical shifts also indicates that the linker region lacks secondary structure (Figure S2C). Deletion of up to 9 amino acids in the linker region supports binding to preE-*let-7d* or preE-*let-7f-1*, although further deletion prevents complex formation (Figures 2B, 2C, and S2D). We conclude that a Lin28 fragment (31–187) with N- and C-terminal truncations and a 9 residue linker deletion (Lin28 $\Delta\Delta$) is sufficient for binding to preE-*let-7* in vitro.

To test whether Lin28 $\Delta\Delta$ can inhibit *let-7* processing in cells, we compared the intracellular levels of processed mature *let-7g* when pri-*let-7g* is cotransfected with different Lin28 trun-

cation constructs. Relative to vector alone, Lin28 $\Delta\Delta$ significantly reduces the level of mature *let-7g*, although not as much as the full-length Lin28 construct, probably due to slightly lower affinity (Figures 2C and 2D). Processing of pri-miR-122 or pri-miR-16 is not inhibited by either Lin28 construct (Figure S2E). Ectopically expressed Lin28 levels are similar to the endogenous levels observed in P19 mouse embryonic carcinoma cells and also among all Lin28 constructs (Figures 2E and 2F). The Lin28 $\Delta\Delta$ construct we have identified is therefore comparable to the full-length protein in its ability to inhibit processing in vivo as well as to bind *let-7* precursors in vitro.

High-Resolution Crystal Structures of Lin28 with *let-7* MicroRNAs

We determined crystal structures of Lin28 $\Delta\Delta$ in complex with preE-*let-7s* derived from *let-7d*, *let-7f-1*, and *let-7g*, at resolutions 2.9 Å, 2.8 Å, and 2.0 Å, respectively, from three different crystal forms (Figures 3 and S3A). We used single-wavelength anomalous dispersion (SAD), with the bound zinc atoms as the anomalous scatterers, to determine the structure of the

Lin28 $\Delta\Delta$:preE_M-*let-7d* complex; we determined the other structures by molecular replacement. Although the overall architectures of the three complexes are similar (Lin28 C α root-mean-square deviation [rmsd] < 1.3 Å), there are several local differences due to divergent RNA sequences (Figures S3B and S3C, and see CSD and CCHC sections below).

The structures reveal that the CSD and CCHCx2 domains of Lin28 interact with two distinct single-stranded (ss) regions of the RNA fragment (Figure 3A). The preE loop encircles a protrusion of the CSD as a necktie would wrap around a collar, with the extensive contacts around the circle made possible by the presence of the preE stem, which functions as the necktie's knot. The CCHC zinc knuckles interact with the GGAG motif at the 3' end, and several sequence-specific interactions shape the single-stranded segment around the knuckles to introduce a distinctive kink in the RNA backbone. Positively charged surfaces on both domains interact with RNA throughout the complex (Figure 3B).

The shortened linker between the CSD and CCHCx2 is the most variable region among the different complexes, as might have been expected from its flexibility (Figure S3D). In all three crystal forms, we observe a domain swap in which the Lin28 CSD interacts with the loop of one RNA molecule, and the CCHCx2 interacts with the GGAG of a second RNA (Figure 3C). That is, each Lin28 monomer in the crystal interacts with distinct elements of two separate preE_M-*let-7d* molecules. In sedimentation equilibrium ultracentrifugation experiments under more physiological conditions, we observe only monomeric complexes of Lin28:preE-*let-7d*, with or without internal deletions in the Lin28 linker (Figure S3E). An unswapped complex conformation can be modeled with a small rearrangement of the C-terminal extension of the CSD (residues 112–121) and a rotation of the 7 residue linker to span the 18–30 Å distance between the CSD and CCHCx2 on the same RNA (Figure 3A). Moreover, the longer, 16 residue linker in wild-type Lin28 would accommodate even longer RNA substrates, including pre-*let-7d* without internal deletions. The monomeric model is also consistent with our observation that high-affinity RNA binding by Lin28 requires both Lin28-binding sites on the same molecule (Figure 1C). As all biochemical evidence points to a monomeric complex in solution, we restrict our description to a 1:1 complex, with the CSD and CCHCx2 bound in *cis* to a single RNA.

Specific Binding of PreE-*let-7* with the CSD

A detailed analysis of the contacts between the CSD and the preE-*let-7* stem loops suggests that specificity relies on both the sequence and the conformation of the RNA. Most of the direct contacts lie in a ≥ 9 nucleotide segment that includes the preE loop (Figures 4, S4A, and S4B). As the loop wraps around the CSD, the bases project and make a number of π -stacking interactions with aromatic side chains. Complementary to the Velcro-like effects of the hydrophobic interactions, hydrogen bonding and steric exclusion create nucleotide preferences and enhance specificity. From inspection of the binding pocket of each nucleotide, we can imagine an ideal RNA substrate for the CSD of Lin28. To simplify the discussion, we define the middle position of preE-*let-7d* that docks into the pocket lined by Phe73 and Lys102 as the “center,” or position 0. Purines are preferred at positions 0 and –1, near the tip of

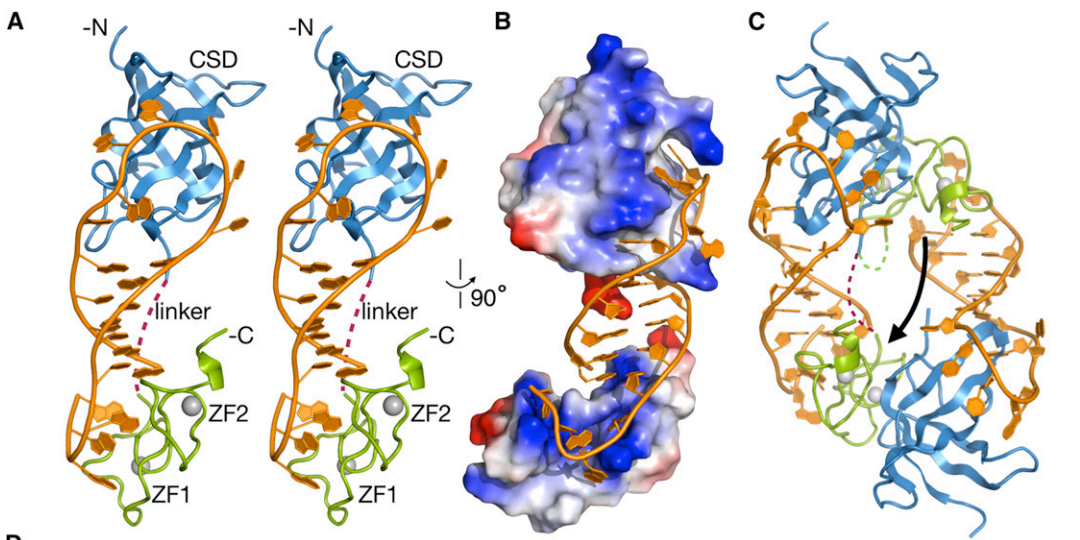
the loop, so that the bulky bases can reach the protein. Position 1, on the other hand, is limited to a pyrimidine, as Lys45 and Asp71 impose steric hindrance. A deeper pocket at position –3 makes a purine more favorable because a larger ring is necessary to stack over Phe84 (in d and f-1) and also to make favorable contacts with the Lin28 backbone (in all three). The hydrogen-bonding networks around –3, –1, and 0 are specific for G, G, and A, respectively.

We evaluated the effect of several point mutations in the co-crystallized preEs at positions where specific interactions are observed in the structures (Figures 4C and S4C). Most of the mutant probes have lower affinity for Lin28 than wild-type. Although Gua is strongly preferred over Ade at positions –3 and –1, substitution of Ade0 with a Gua is not as deleterious. Ade replaces Gua(–3) in the Lin28:*let-7g* complex, and as a result some favorable hydrogen bonds are absent in comparison to other structures. Due to the small size of the pocket, a pyrimidine is strongly preferred at position 1. Some of the previously reported mutations of preE-*let-7g* include a transversion (purine to pyrimidine) at position 0 (Newman et al., 2008) and changes in the preE stem that disrupt base pairing (Piskounova et al., 2008). Although our studies focused on mouse Lin28a, the observed effects of preE_M point mutations on complex stability are equivalent for human Lin28a and Lin28b (Figure S4C).

Comparing the structure of Lin28 bound to the divergent preE-*let-7g* with those of the preE-*let-7d* and -*7f-1* complexes illustrates how the CSD accommodates variability in substrate RNAs. The short preE loops in *let-7d* and -*7f-1* require that base pairs be broken to fit around the CSD. In order to tighten the longer loop in preE-*let-7g* (Figure 4B), Arg50 moves in to mimic a base, pairing with Cyt5 and stacking against Ade5. The other extra bases have π -stacking interactions: two with the side chains of Arg122 and Arg123 at the amino-terminal end of the interdomain linker, and Ade2 and Cyt3 with each other. A closed RNA loop appears to be important to maintain full contact with the CSD, perhaps explaining the more extensive interactions here than in other CSD:RNA complex structures (Frazão et al., 2006; Max et al., 2006, 2007) (Figure S4D).

Interactions of Zinc Knuckles with PreE-*let-7*

The CCHC knuckles maximize favorable interactions with a small number of nucleotides by making many contacts with the bases (Figures 5A–5C). The intimate interaction between GGAG and CCHCx2 produces a distinctive kink in the RNA backbone. Most of the protein atoms participating in the extensive hydrogen-bonding network lie in relatively rigid regions of the protein, such as adjacent to zinc-coordinating residues or in a proline-rich linker, thereby imposing a specific, rigid conformation on the 3' end of the RNA (Figures 5C, S5A, and S5B). Ring stacking and hydrophobic interactions with side chains of the CCHCx2 further stabilize the particular conformation by aligning the bases. One of the key residues is Y140, which establishes the kinked conformation by sandwiching between the last two bases (AG) and interacting with H162, which braces the first (G). Although the adenine base does not have as many polar contacts with Lin28, it packs closely against the first (G) and makes a hydrogen bond that assists in bending the RNA backbone. The resulting conformation of the ssRNA resembles that



D

	Lin28:preE _M -let-7d Native	Lin28:preE _M -let-7d Zn-SAD	Lin28:preE _M -let-7f-1 Native	Lin28:preE _M -let-7g Native
Data collection ^a				
Beamline	NSLS x25	NSLS x25	NSLS x25	APS 24-ID-C
Resolution (Å)	66.6-2.9 (2.910-2.900)	70.0-3.22 (3.230-3.220)	73-2.76 (2.772- 2.763)	25.8-2.01 (2.014-2.007)
Space Group	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	C222 ₁
Wavelength (Å)	1.0000	1.2549	0.9795	0.9795
Unit cell (Å)	a=b=143.8, c=177.7	a=b=144.0, c=179.0	a=b=139.9, c=85.6	a=46.0, b=109.2, c=182.9
R _{sym} ^b	0.071 (0.586)	0.109 (0.622)	0.06 (0.629)	0.046 (0.433)
I / σ(I)	21.4 (3.2)	21.3 (4.4)	28.2 (3.3)	16.9 (2.1)
Completeness (%)	98.5 (99.8)	99.9 (100)*	100 (100)	98.3 (92.5)
Redundancy	5.6 (5.5)	4.6 (4.6)*	8 (8.3)	3.2 (2.4)
Mosaicity (°)	0.145	0.187	0.103	0.164
Phasing				
Mean FOM, initial		0.32		
Mean FOM, after NCS and DM ^c		0.65		
Refinement				
R _{work} /R _{free} (%) ^d	17.93/20.80		19.30/21.45	18.95/21.73
Test Size	2086 (5.06%)		1148 (5.15%)	1550 (5%)
Number of complexes in ASU	6		2	2
Average B /Wilson (Å ²)	70.3 / 83.2		74.8 / 78.1	49.6 / 40.3
Ramachandran (%) ^e	88.1 / 11.9 / 0 / 0		88.2 / 11.8 / 0 / 0	93.2 / 6.8 / 0 / 0
Rmsd Bond Length (Å)	0.01		0.01	0.01
Rmsd Bond Angles (°)	1.42		1.36	1.25
Peptide Omega Torsion (°)	3.35		3.4	3.60
Other Torsion (°)	24.89		22.23	19.58
solvent content	0.681		0.759	0.559

^aData for the outermost shell are given in parentheses, and asterisks reflect anomalous.
^bR_{sym} = 100 ∑ ||h - <lh>| / ∑ ||h|, where <lh> is the average intensity over symmetry equivalents.
^cFigure of merit after non-crystallographic symmetry averaging over six copies and density modification in AutoSol/ RESOLVE.
^dR_{work} or free = 100 ∑ ||F_o - F_c| / ∑ ||F_o|. R_{work} and R_{free} were calculated from the working and test sets, respectively.
^eRamachandran (%) is given for each 1:1 complex, favorable/allowed/generously allowed/disallowed regions, respectively.

Figure 3. Structure of the Lin28:PreE_M-let-7d Complex
Cartoon representations were colored by domain: blue, CSD; green, CCHC2; gray, zinc; orange, RNA.
(A) Stereo representation of the monomeric complex. Interdomain linker is represented by a purple dotted line.

of the so-called “K-turn,” which often participates in specific protein-RNA interactions (Klein et al., 2001).

The CCHCx2 regions from all our structures align well with each other, except for slight differences, due to crystal contacts, in one of the two noncrystallographic copies of preE_M-*let-7g* (Figures S5C and S5D). When compared with the conformation seen in the solution structure of an isolated Lin28 zinc-knuckle fragment (Protein Data Bank [PDB] 2CQF), however, there is a large rearrangement of the inter-knuckle joint in Lin28 (Figure 5D). Therefore, association of the CCHCx2 with GGAG imposes specific conformational constraints on both the RNA and the protein; this reciprocal effect may be functionally important for regulation.

Two NMR structures of CCHC motifs from HIV NCp1 have been determined previously, in which the knuckles bind a tetraloop of sequence GGAG or GGUG in two stem loops (SL2 and SL3) of the ψ site (Amarasinghe et al., 2000; De Guzman et al., 1998). The conformation of the GGAG motif in complex with Lin28 is very different from its conformation in complex with HIV NCp1, indicating that the conformation we observe is specific to Lin28 (Figures S5E and S5F).

Lin28 Interactions with Full-Length Pre-*let-7*

To test our conclusions from the model provided by the crystal structures, and to verify that the truncations and deletions we have made for crystallization do not affect specificity, we generated mutant forms of full-length Lin28 and pre-*let-7g*. Alteration of the key binding sites of the CSD (near position 0) or CCHCx2 (GGAG) in pre-*let-7g* reduces affinity, consistent with the mutagenesis studies with preE fragments (Figures 6A and S6A). In addition, mutation of RNA-contacting residues in the CSD and CCHCx2 also interferes with complex formation, especially when aromatic side chains are replaced with Ala (Figures 6B and S6B). We then conducted binding assays using combinations of protein and RNA mutants (Figures 6C and S6C). The D71 side chain, which is near nucleotide position 1, limits the size of the pocket and restricts it to pyrimidine rings. Presumably due to the additional free space provided by a glycine, a D71G mutant no longer discriminates against a purine at position 1 (Figure 6C, D71G block).

The bipartite character of the Lin28:*let-7* interactions implies that one should observe strong synergy when combining a mutation in one of the two *let-7* interaction sites with a mutation in the Lin28 domain that recognizes the other *let-7* interaction site. Indeed, a CSD mutation (F73A) has much greater effect on binding with RNA bearing a mutation in the GGAG motif (to GGAU or deletion) than it does on binding with RNA bearing a preE loop mutation near the CSD-binding site (Figure 6C, F73A block). Similarly, for binding with a mutated CCHCx2 (Y140A), GGAG mutations are not as detrimental as a CSD-binding-site mutation (Figure 6C, Y140A block). We have also tested binding of individual domains of Lin28 to various pre-

let-7g mutants (Figure 6C, CSD and CCHCx2 blocks). Neither isolated domain binds to *let-7* as specifically or tightly as does full-length Lin28. Nevertheless, RNA mutations at each binding site affect only the affinity of the corresponding domain, consistent with our model. In summary, the results of all these mutational studies are all consistent with the conclusion that Lin28 binds full-length pre-*let-7* in the same way as does the truncated form present in our crystals.

The GGAG motif is conserved among *let-7*s not only in its sequence but also in its proximal position with respect to the Dicer site in the context of the full pre-*let-7* molecule. The last G is 4 bases from the Dicer cleavage site on the 3' strand and only 2 bases from the position at which complementarity to the mature strand begins. Using previously determined structures of Dicer and the proposed location of the cut site (Du et al., 2008; Macrae et al., 2006), we have modeled how Lin28 would interfere with Dicer binding (Figure S6D). Because their binding sites on RNA are close together and because Lin28 bends the RNA backbone, Lin28, especially its CCHCx2, may hinder Dicer directly. To test whether binding of Lin28 with pre-*let-7g* is sufficient to inhibit Dicer processing, we used different mutants in an in vitro Dicer assay (Figure 6D). The mutations that disrupt association between Lin28 and pre-*let-7* lead to increased Dicer cleavage, compared with wild-type control. Our data are thus consistent with a direct effect of Lin28 on Dicer processing of pre-*let-7*.

We also tested the effect of the described mutations on in vivo processing of *let-7* (Figures 6E–6G). Mutations that affect CSD binding derepress processing of pri-*let-7g* only modestly, perhaps because the presence of other cellular factors partially compensates for the affinity change (<10-fold). Altering the CCHCx2:GGAG interaction—by changes in RNA or protein—is more detrimental to Lin28 activity. Levels of mature *let-7* in our in vivo assay depend on both complex formation between Lin28 and *let-7* precursors and downstream effects of Lin28, such as hindering Drosha and Dicer while recruiting TUTase. Our results indicate that although both the CSD and CCHCx2 contribute to affinity and specificity for *let-7* precursors, the CCHCx2:GGAG interaction is more critical for the effector function of Lin28.

DISCUSSION

Deciphering Lin28 Specificity

The structural and biochemical studies presented here reveal how Lin28 recognizes *let-7* precursors and allow us to postulate how Lin28 might bind diverse pre-*let-7*s. We propose a preferred sequence consensus for CSD binding: NGNGA₀YNNN (Y = pyrimidine; N = any base). The sequences and distances between the CSD-binding site and the CCHCx2-binding GGAG motif are variable, but the two sites can be identified in many of the preE-*let-7* sequences (Figure S7A). In cases where no significant preE stem structure is predicted (e.g., in *let-7a-2* or

(B) Same complex in (A) represented with surface colored by electrostatic potential and rotated.

(C) Domain-swapped dimer. Arrow points from the domain-swapped to the biologically relevant CCHCx2 domain. Linker connecting swapped domains is marked in green, dotted line. Linker connecting unswapped domains is marked in purple, dotted line. See also Figure S3.

(D) X-ray data collection and refinement statistics.

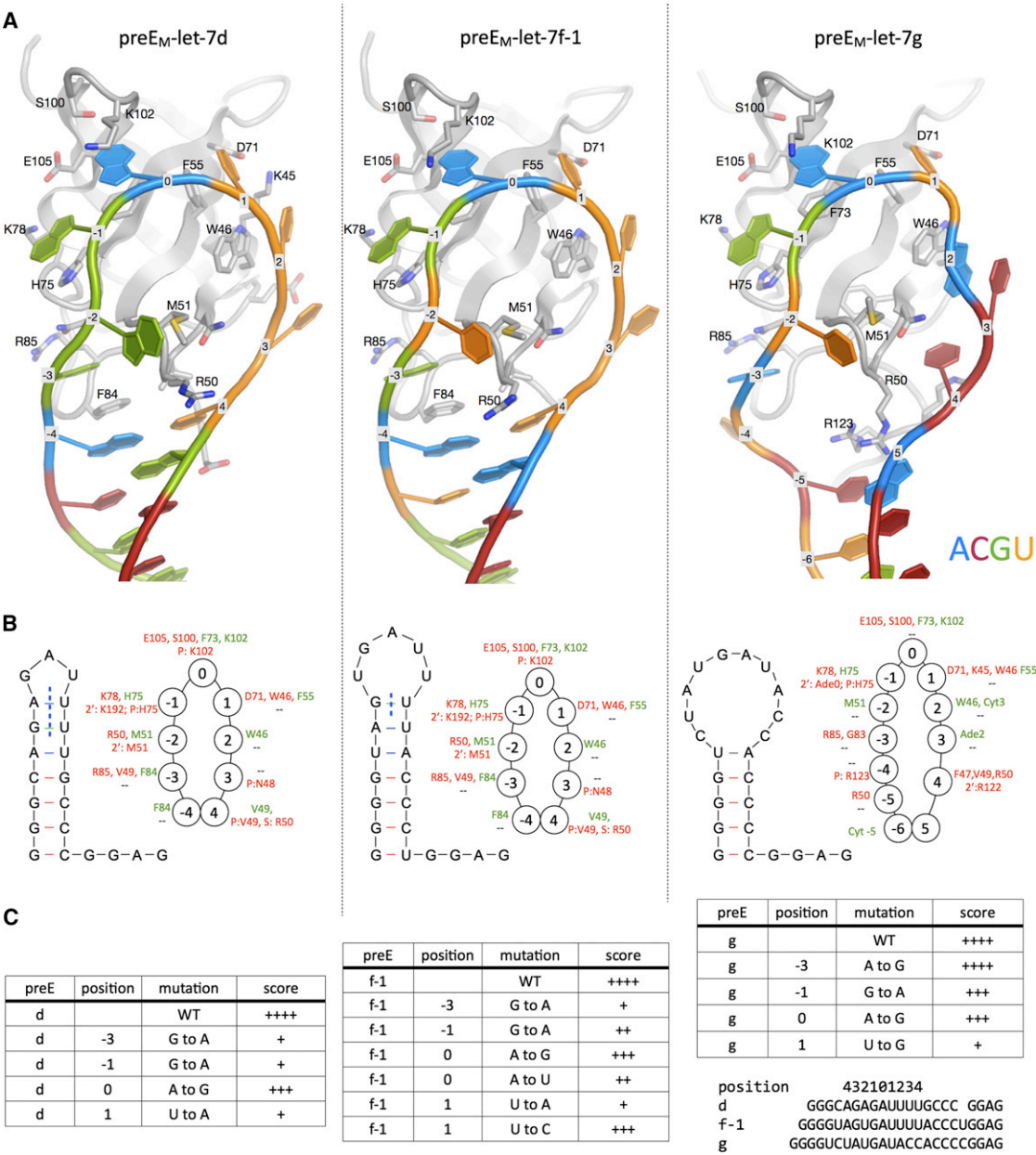


Figure 4. CSD:RNA Interactions; See also Figure S4

(A) Close-up view of the CSD (backbone as gray cartoon; key residues also shown with sticks and labels) interacting with the preE_M loops as labeled, shown in the same orientation. RNA is colored by base identity, (A)zure, (A)de; (C)rimson, (C)yt; (G)green, (G)ua; (U)mber, (U)ra, and marked by position number on backbone as defined in text.

(B) Schematic drawing of predicted structures of preE_M sequences used for crystallization. Some of the predicted base pairs are broken (blue vertical dotted line) in the complex structures. Rings of circles show the protein:RNA interactions at each nucleotide position, marked with interacting residues (green, hydrophobic or π -stacking; red, H-bond). Top line is for base contacts and bottom line is for sugar or phosphate interactions at each position.

(C) Comparison of Lin28 affinity for various preE_M-let-7 mutants, described by the parent let-7, position of the mutation, and the base identity. Accompanying gels are shown in Figure S4.

let-7c-1), the nearby mature region with its stable double-stranded (ds) helix may aid in closing the loop around the CSD. Loss of one or a few favorable interactions in other preE-let-7s might not completely exclude the RNAs from binding to Lin28 but rather result in differences in affinity that could affect the sensitivity of particular let-7s to Lin28 regulation in vivo. Indeed, understanding Lin28 specificity from preE-let-7d and preE-let-7f-1 allowed us to crystallize the preE-let-7g complex, which binds to Lin28 in an energetically less stable conformation (Figures S7B and S7C).

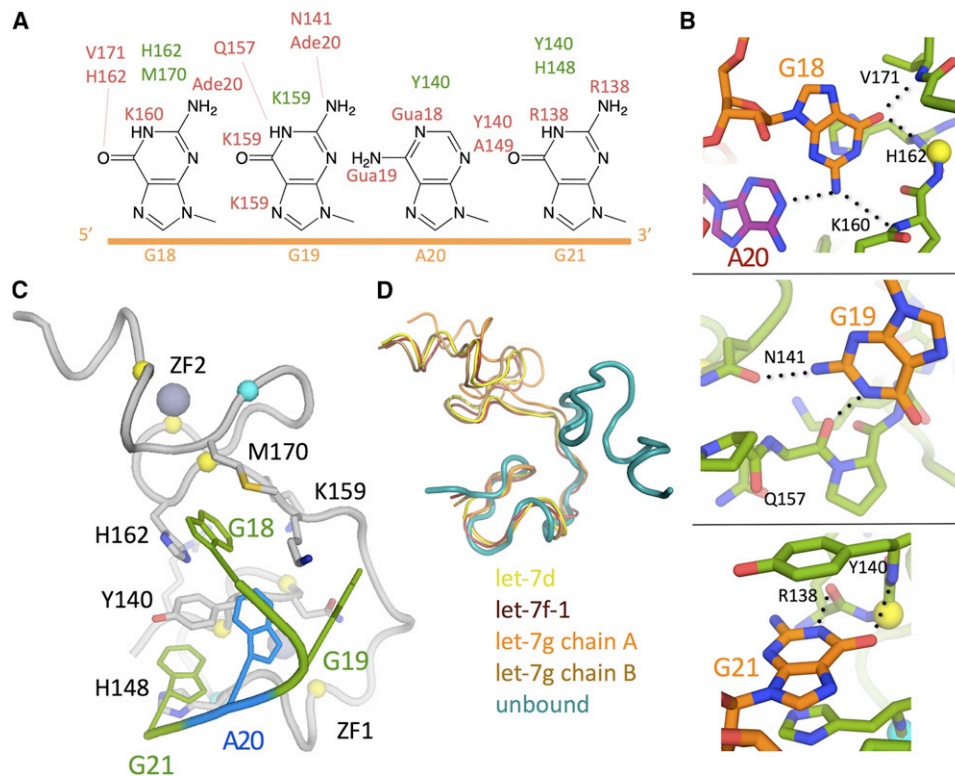


Figure 5. CCHCx2:RNA interactions; See also Figure S5

(A) Schematic drawing of GGAG, and atoms making contact are marked with amino acid/nucleotide numbers (green, hydrophobic or π -stacking; red, H-bond). (B) Close-up view of base interactions. H-bonds are marked with dashed lines. (C) View of GGAG interactions with the CCHCx2. Lin28 is represented with gray cartoon, and GGAG are colored by sequence (G)reen, (G)ua; (A)zure, (A)de. Zinc (large gray spheres) coordinating residues are represented with small spheres at C α positions (yellow, Cys; cyan, His). (D) Comparison of the CCHCx2 region of Lin28 in different states by superimposition of the first CCHC motif (unbound, PDB code = 2CQF).

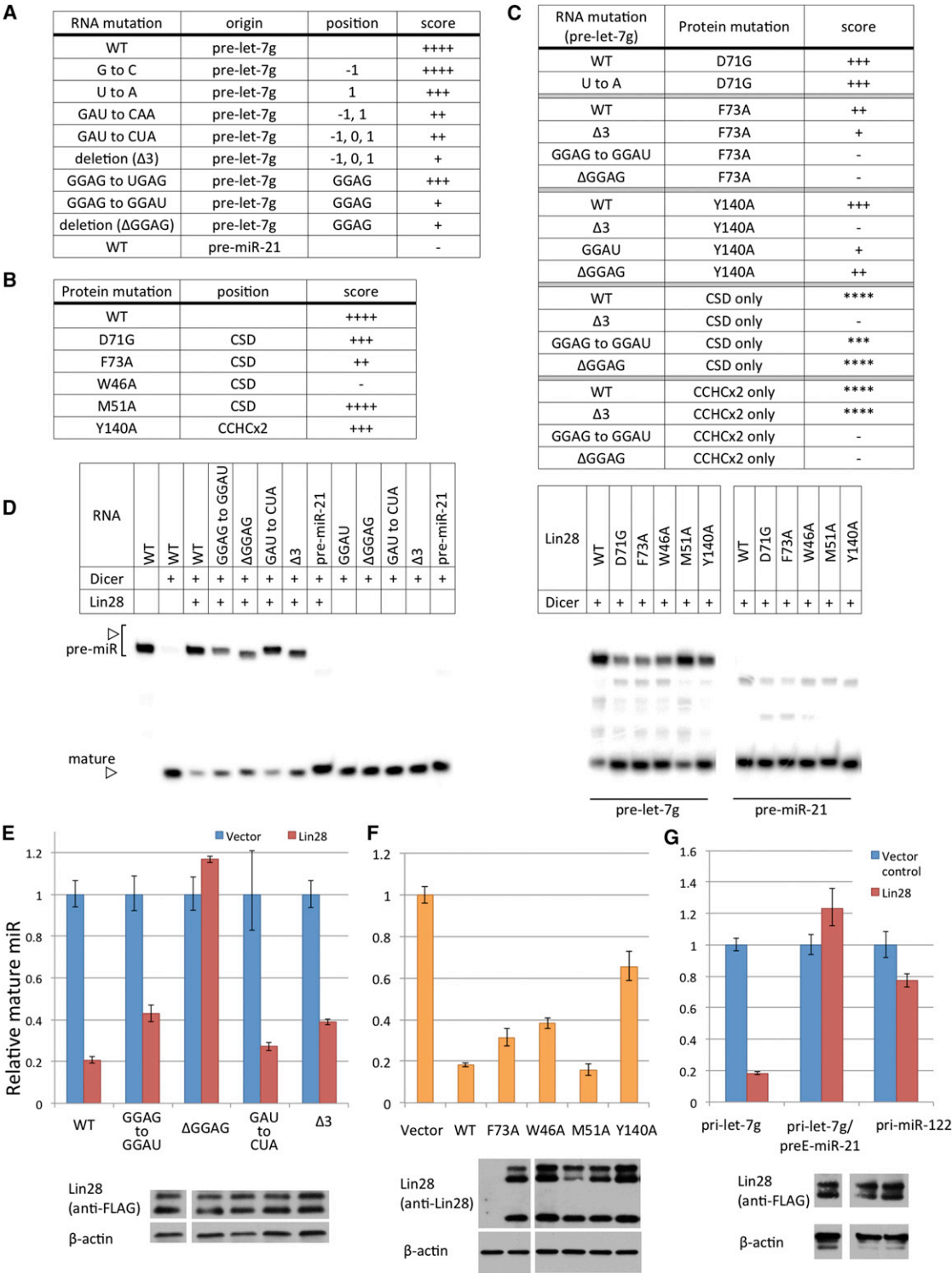
The sequence of the linker between the CSD and CCHCx2 has a strong net positive charge, probably to interact with the negatively charged RNA sugar-phosphate backbone, or to compensate for any unpaired bases, as seen in the case of the preE-*let-7g* complex. Evolutionary conservation of the electrostatic property suggests that the linker does play some role, even though its sequence is not crucial for binding specificity. The length of the linker varies in some organisms, and shorter linkers occur in those with only one copy of *let-7* containing a shorter preE sequence. Longer, more flexible linkers might have evolved in higher eukaryotes to recognize longer and divergent *let-7* precursors. Our preE-*let-7g* complex structure illustrates how the linker can adapt to different RNA substrates; Arg122 and Arg123 at the amino-terminal end of the interdomain linker stack against extra bases near the ds-ss junction (Figure S5B).

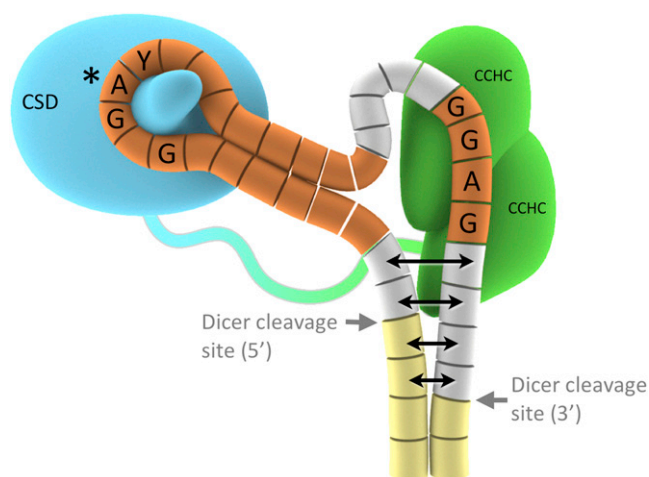
The GGAG tetranucleotide motif is well conserved among the members of the *let-7* family within a particular species. In evolutionarily distant organisms such as worms and fruit flies, however, other sequences (such as GGUG or AUCA) are found in place of GGAG, perhaps due to coevolution of RNA and protein. Although not included in the crystal structure, the two nucleotides following GGAG are A and U in most *let-7* sequences. In the context of full-length molecules, there may be more contacts between the bulges near the GGAG and

CCHCx2. The importance of the GGAG motif has been explored previously, by introducing a GGAG motif into an unrelated RNA sequence, miR-16, to generate a chimeric pre-miRNA that has gained affinity for Lin28 (Heo et al., 2009). From our binding experiments and structural data, the GGAG motif alone cannot confer robust binding with Lin28, and shifting its position by a base or two relative to the CSD-binding site does not affect Lin28 binding significantly. In the case of the chimeric RNA with miR-16, its preE also coincidentally contains a sequence similar to the preferred CSD-binding site (UAAGAUUCU versus NGNGAYNN), at the 5' side of the GGAG motif, explaining why this chimera could bind Lin28. Our structural and biochemical data thus provide a molecular explanation for Lin28 specificity, making it possible to investigate further its role in *let-7* biogenesis as well as its function in binding various mRNA targets (Jin et al., 2011; Peng et al., 2011; Qiu et al., 2009).

Implications for miRNA Regulatory Mechanisms

Although Drosha and Dicer are known to cut at opposite ends of the mature miRNA, there are still major questions regarding how they recognize their target and how the cleavage can be regulated. Our structures of Lin28:preE-*let-7* complexes combined with known structural data for Dicer have allowed us to postulate how the Lin28-binding event itself can inhibit processing of





Electrophoretic Mobility Shift Assay

For preE probes, RNA oligonucleotides were synthesized (IDT), and full pre-miR probes were purified by PAGE after in vitro transcription followed by double ribozyme cleavage, as detailed in Walker et al. (2003). RNAs were radiolabeled with ATP[γ - 32 P] using T4 polynucleotide kinase, incubated with protein in a buffer containing 20 mM Tris 7.5, 100 mM NaCl, 10 mM DTT, 50 μ M ZnCl₂, 15 μ g/ μ l yeast tRNA, and 1U/ μ l RNase inhibitor.

NMR Spectroscopy

All NMR samples were prepared as 0.5 mM Lin28:preE_M-let-7d complexes. Sequence-specific chemical shifts for backbone atoms were determined for 157 residues (out of 166 total, including 13 prolines), using the TROSY versions of HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HNCO, and HN(CA)CO, with a 15 N, 13 C, and 85% 2 H-labeled protein combined with unlabeled RNA. Experiments were conducted at 30°C on Bruker spectrometers equipped with cryogenic probes, operating at 1H frequencies of 600 MHz (sequence assignment and relaxation experiments) or 750 MHz (NOESYs).

Crystallography

Crystals of all three complexes were produced by vapor diffusion, using reservoir solution containing 0.6 M NaH₂PO₄, 1.4 M K₂HPO₄, and 5% glycerol for preE_M-let-7d and preE_M-let-7f-1 complexes, and 0.1 M Tris (pH 8.0), 32% w/v PEG 4000, and 0.2 M sodium acetate for the preE_M-let-7g complex. Experimental phases were obtained for the preE_M-let-7d complex by anomalous scattering from zinc atoms (SAD), and the structures of preE_M-let-7f-1 and preE_M-let-7g complexes were solved by molecular replacement with Lin28:pre-let-7d as a search model.

Dicer In Vitro Processing Assay

Dicer expression construct (Addgene plasmid 19873) and purification are described as in Landthaler et al. (2008), and radiolabeled pre-miR constructs were prepared similarly to EMSA probes. Dicer assays were carried out as described in De and Macrae (2011), using a buffer containing 20 mM Tris 7.5, 5% glycerol, 3.2 mM MgCl₂, 5 mM DTT, 50 mM NaCl, and 100 μ M ZnCl₂.

MicroRNA In Vivo Processing Assay

Ability of Lin28 constructs to block let-7 processing in cells was compared as outlined in Viswanathan et al. (2008). Briefly, pri-let-7g was cotransfected with FLAG-tagged Lin28 constructs (25 ng unless otherwise noted) or vector control into 293T cells (12-well) using lipofectamine. Total RNA was isolated using TriZol reagent and treated with DNase I, and quantitative RT-PCR was used with miRNA-specific stem-loop primers as previously described (Wan et al., 2010). Relative levels of mature miRNAs were analyzed by $\Delta\Delta$ Ct method and normalized by U6 snRNA levels.

ACCESSION NUMBERS

Coordinates and structure factors for the structures of Lin28:preE_M-let-7d, Lin28:preE_M-let-7f-1, and Lin28:preE_M-let-7g complexes have been deposited with the Protein Data Bank under accession codes 3TRZ, 3TS0, and 3TS2, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures and seven figures and can be found with this article online at doi:10.1016/j.cell.2011.10.020.

ACKNOWLEDGMENTS

We thank Stephen C. Harrison for discussion and critical review of the manuscript, members of the Chou Lab for discussion about NMR experiments, Janet Iwasa for help with design of Figure 7, George Q. Daley for the pri-miR constructs for in vivo processing assays, Thomas Tuschl for Dicer expression construct, Katarzyna Szatkowski for technical assistance, and Elizabeth O'Day for participation during the early stages of the project. This work is

based upon research conducted at the Advanced Photon Source (North-eastern Collaborative Access Team beamlines) and Brookhaven National Laboratory (beamline X25). This work was partially supported by Center for Molecular and Cellular Dynamics at Harvard Medical School, by NIH grant 5U54GM094608, and by postdoctoral fellowships to Y.N. from the Damon Runyon Cancer Research Foundation (DRG-#1953-07) and the Charles A. King Trust, N.A., Bank of America, Co-Trustee. Y.N. performed experiments. C.C. provided technical assistance with experiments. R.G. provided the Lin28 clone and participated in discussions. J.C. helped with NMR spectroscopy. Y.N. and P.S. designed and interpreted experiments and prepared the manuscript.

Received: April 27, 2011

Revised: August 5, 2011

Accepted: October 6, 2011

Published online: November 10, 2011

REFERENCES

- Amarasinghe, G.K., De Guzman, R.N., Turner, R.B., Chancellor, K.J., Wu, Z.R., and Summers, M.F. (2000). 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.
- Büssing, I., Slack, F.J., and Grosshans, H. (2008). let-7 microRNAs in development, stem cells and cancer. *Trends Mol. Med.* 14, 400–409.
- Davis, B.N., and Hata, A. (2010). microRNA in Cancer—The involvement of aberrant microRNA biogenesis regulatory pathways. *Genes & Cancer* 1, 1100–1114.
- De Guzman, R.N., Wu, Z.R., Stalling, C.C., Pappalardo, L., Borer, P.N., and Summers, M.F. (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* 279, 384–388.
- De, N., and Macrae, I.J. (2011). Purification and assembly of human Argonaute, Dicer, and TRBP complexes. *Methods Mol. Biol.* 725, 107–119.
- Du, Z., Lee, J.K., Tjhen, R., Stroud, R.M., and James, T.L. (2008). Structural and biochemical insights into the dicing mechanism of mouse Dicer: a conserved lysine is critical for dsRNA cleavage. *Proc. Natl. Acad. Sci. USA* 105, 2391–2396.
- Frazão, C., McVey, C.E., Amblar, M., Barbas, A., Vornrhein, C., Arraiano, C.M., and Carrondo, M.A. (2006). Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. *Nature* 443, 110–114.
- Guo, Y., Chen, Y., Ito, H., Watanabe, A., Ge, X., Kodama, T., and Aburatani, H. (2006). Identification and characterization of lin-28 homolog B (LIN28B) in human hepatocellular carcinoma. *Gene* 384, 51–61.
- Hagan, J.P., Piskounova, E., and Gregory, R.I. (2009). Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* 16, 1021–1025.
- Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., Sohn, S.Y., Cho, Y., Zhang, B.-T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* 125, 887–901.
- Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V.N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol. Cell* 32, 276–284.
- Heo, I., Joo, C., Kim, Y.-K., Ha, M., Yoon, M.-J., Cho, J., Yeom, K.-H., Han, J., and Kim, V.N. (2009). TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* 138, 696–708.
- Iliopoulos, D., Hirsch, H.A., and Struhl, K. (2009). An epigenetic switch involving NF- κ B, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. *Cell* 139, 693–706.
- Jin, J., Jing, W., Lei, X.-X., Feng, C., Peng, S., Boris-Lawrie, K., and Huang, Y. (2011). Evidence that Lin28 stimulates translation by recruiting RNA helicase A to polysomes. *Nucleic Acids Res.* 39, 3724–3734.
- Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* 10, 126–139.

- King, C., Cuatrecasas, M., Castells, A., Sepulveda, A., Lee, J.S., and Rustgi, A.K. (2011). Lin28b promotes colon cancer progression and metastasis. *Cancer Res.* **71**, 4260–4268.
- Klein, D.J., Schmeing, T.M., Moore, P.B., and Steitz, T.A. (2001). The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**, 4214–4221.
- Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* **11**, 597–610.
- Landthaler, M., Gaidatzis, D., Rothballer, A., Chen, P.Y., Soll, S.J., Dinic, L., Ojo, T., Hafner, M., Zavolan, M., and Tuschl, T. (2008). Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA* **14**, 2580–2596.
- Lehrbach, N.J., Armissen, J., Lightfoot, H.L., Murfitt, K.J., Bugaut, A., Balasubramanian, S., and Miska, E.A. (2009). LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in *Caenorhabditis elegans*. *Nat. Struct. Mol. Biol.* **16**, 1016–1020.
- Lette, G., Jackson, A.U., Gieger, C., Schumacher, F.R., Berndt, S.I., Sanna, S., Eyheramendy, S., Voight, B.F., Butler, J.L., Guiducci, C., et al. Diabetes Genetics Initiative; FUSION; KORA; Prostate, Lung Colorectal and Ovarian Cancer Screening Trial; Nurses' Health Study; SardiNIA. (2008). Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat. Genet.* **40**, 584–591.
- Lu, L., Katsaros, D., Shaverdashvili, K., Qian, B., Wu, Y., de la Longrais, I.A.R., Preti, M., Menato, G., and Yu, H. (2009). Pluripotent factor lin-28 and its homologue lin-28b in epithelial ovarian cancer and their associations with disease outcomes and expression of let-7a and IGF-II. *Eur. J. Cancer* **45**, 2212–2218.
- Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195–198.
- Markham, N.R., and Zuker, M. (2005). DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.* **33**(Web Server issue), W577–W581.
- Max, K.E.A., Zeeb, M., Bienert, R., Balbach, J., and Heinemann, U. (2006). T-rich DNA single strands bind to a preformed site on the bacterial cold shock protein Bs-CspB. *J. Mol. Biol.* **360**, 702–714.
- Max, K.E.A., Zeeb, M., Bienert, R., Balbach, J., and Heinemann, U. (2007). Common mode of DNA binding to cold shock domains. Crystal structure of hexathymidine bound to the domain-swapped form of a major cold shock protein from *Bacillus caldolyticus*. *FEBS J.* **274**, 1265–1279.
- Michlewski, G., and Cáceres, J.F. (2010). Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nat. Struct. Mol. Biol.* **17**, 1011–1018.
- Michlewski, G., Guil, S., Semple, C.A., and Cáceres, J.F. (2008). Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol. Cell* **32**, 383–393.
- Moss, E.G., Lee, R.C., and Ambros, V. (1997). The cold shock domain protein LIN-28 controls developmental timing in *C. elegans* and is regulated by the lin-4 RNA. *Cell* **88**, 637–646.
- Newman, M.A., Thomson, J.M., and Hammond, S.M. (2008). Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA* **14**, 1539–1549.
- Ong, K.K., Elks, C.E., Li, S., Zhao, J.H., Luan, J., Andersen, L.B., Bingham, S.A., Brage, S., Smith, G.D., Ekelund, U., et al. (2009). Genetic variation in LIN28B is associated with the timing of puberty. *Nat. Genet.* **41**, 729–733.
- Peng, S., Maihle, N.J., and Huang, Y. (2010). Pluripotency factors Lin28 and Oct4 identify a sub-population of stem cell-like cells in ovarian cancer. *Oncogene* **29**, 2153–2159.
- Peng, S., Chen, L.-L., Lei, X.-X., Yang, L., Lin, H., Carmichael, G.G., and Huang, Y. (2011). 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.
- Permeth-Wey, J., Kim, D., Tsai, Y.-Y., Lin, H.-Y., Chen, Y.A., Barnholtz-Sloan, J., Birrer, M.J., Bloom, G., Chanock, S.J., Chen, Z., et al. (2011). LIN28B polymorphisms influence susceptibility to epithelial ovarian cancer. *Cancer Res.* **71**, 3896–3903.
- Perry, J.R.B., Stolk, L., Franceschini, N., Lunetta, K.L., Zhai, G., McArdle, P.F., Smith, A.V., Aspelund, T., Bandinelli, S., Boerwinkle, E., et al. (2009). Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. *Nat. Genet.* **41**, 648–650.
- Piskounova, E., Viswanathan, S.R., Janas, M., LaPierre, R.J., Daley, G.Q., Sliz, P., and Gregory, R.I. (2008). Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. *J. Biol. Chem.* **283**, 21310–21314.
- Qiu, C., Ma, Y., Wang, J., Peng, S., and Huang, Y. (2009). Lin28-mediated post-transcriptional regulation of Oct4 expression in human embryonic stem cells. *Nucleic Acids Res.* **38**, 1240–1248.
- Rybak, A., Fuchs, H., Smirnova, L., Brandt, C., Pohl, E., Nitsch, R., and Wulczyn, F. (2008). A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat. Cell Biol.* **10**, 987–993.
- Siomi, H., and Siomi, M.C. (2010). Posttranscriptional regulation of microRNA biogenesis in animals. *Mol. Cell* **38**, 323–332.
- Sulem, P., Gudbjartsson, D.F., Rafnar, T., Holm, H., Olafsdottir, E.J., Olafsdottir, G.H., Jonsson, T., Alexandersen, P., Feenstra, B., Boyd, H.A., et al. (2009). Genome-wide association study identifies sequence variants on 6q21 associated with age at menarche. *Nat. Genet.* **41**, 734–738.
- Viswanathan, S.R., and Daley, G.Q. (2010). Lin28: A microRNA regulator with a macro role. *Cell* **140**, 445–449.
- Viswanathan, S.R., Daley, G.Q., and Gregory, R.I. (2008). Selective blockade of microRNA processing by Lin28. *Science* **320**, 97–100.
- Viswanathan, S.R., Powers, J.T., Einhorn, W., Hoshida, Y., Ng, T.L., Toffanin, S., O'Sullivan, M., Lu, J., Phillips, L.A., Lockhart, V.L., et al. (2009). Lin28 promotes transformation and is associated with advanced human malignancies. *Nat. Genet.* **41**, 843–848.
- Walker, S.C., Avis, J.M., and Conn, G.L. (2003). General plasmids for producing RNA in vitro transcripts with homogeneous ends. *Nucleic Acids Res.* **31**, e82.
- Wan, G., Lim, Q.E., and Too, H.P. (2010). High-performance quantification of mature microRNAs by real-time RT-PCR using deoxyuridine-incorporated oligonucleotides and hemi-nested primers. *RNA* **16**, 1436–1445.
- Yang, D.H., and Moss, E.G. (2003). Temporally regulated expression of Lin-28 in diverse tissues of the developing mouse. *Gene Expr. Patterns* **3**, 719–726.
- Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., Huang, Y., Hu, X., Su, F., Lieberman, J., and Song, E. (2007a). let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell* **131**, 1109–1123.
- Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007b). Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920.
- Zeng, Y., and Cullen, B.R. (2003). Sequence requirements for micro RNA processing and function in human cells. *RNA* **9**, 112–123.
- Zeng, Y., and Cullen, B.R. (2005). Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J. Biol. Chem.* **280**, 27595–27603.
- Zeng, Y., Yi, R., and Cullen, B.R. (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J.* **24**, 138–148.
- Zhang, X., and Zeng, Y. (2010). The terminal loop region controls microRNA processing by Drosha and Dicer. *Nucleic Acids Res.* **38**, 7689–7697.
- Zhu, H., Shah, S., Shyh-Chang, N., Shinoda, G., Einhorn, W.S., Viswanathan, S.R., Takeuchi, A., Grasemann, C., Rinn, J.L., Lopez, M.F., et al. (2010). Lin28a transgenic mice manifest size and puberty phenotypes identified in human genetic association studies. *Nat. Genet.* **42**, 626–630.

EXTENDED EXPERIMENTAL PROCEDURES

Constructs

All Lin28 crystallization constructs were derived from mouse Lin28a (NP_665832). Expression constructs are in pETDuet-1 (Novagen), with an N-terminal hexahistidine tag followed by a TEV cleavage site. For NMR studies and electrophoretic mobility shift assays, Lin28 (residues 16–184) contained the wild-type linker sequence, unless otherwise noted. For crystallization a similar construct (35–187) with nine internally deleted residues in the linker ($\Delta 9$) was used (Figure 2B). Isolated CSD construct contains residues 16–126 and CCHCx2 construct contains residues 135–184.

Protein Purification and Complex Preparation

Lin28 constructs were overexpressed in *E. coli* strain BL21(DE3) Rosetta pLysS. After initial affinity chromatography step using Ni-NTA beads (QIAGEN), His-tags were removed by incubating with recombinant TEV protease. Cation exchange chromatography (HiTrap S, GE Healthcare) was performed using a buffer containing 20 mM BisTris (pH 6.0), 5 mM dithiothreitol (DTT), 5% glycerol, and 50 μ M ZnCl₂, over 0.1–1 M NaCl gradient. Further purification was accomplished by size-exclusion chromatography (Superdex 200, GE Healthcare) in the same buffer. Complexes with RNA oligonucleotides were prepared for NMR studies and crystallization trials, by mixing at 1:1.2 (protein:RNA) molar ratio, and free RNA was removed by another size-exclusion chromatography step.

NMR Spectroscopy

All NMR samples were prepared as 0.5 mM Lin28 (16–184) and preE_M-let-7d complex, in a buffer containing 20 mM BisTris (pH 7.0), 100 mM NaCl, 5% glycerol, 5 mM dithiothreitol, 50 μ M ZnCl₂, and 0.2% sodium azide. Backbone distance restraints were obtained using uniform ¹⁵N, ¹³C-labeled protein in complex with unlabeled RNA, from 3D ¹⁵N-edited NOESY (mixing time = 120 ms). To measure ¹⁵N R₁ and R₂, a sample containing ¹⁵N, ¹³C, and 85% ²H-labeled protein combined with unlabeled RNA was used with standard pulse schemes (Kay et al., 1989). Secondary chemical shifts were calculated by the comparing the recorded chemical shift to sequence adjusted random coil chemical shift (Schwarzinger et al., 2001).

All spectra were processed and analyzed with NMRPipe (Delaglio et al., 1995) and CCPNmr Analysis (Vranken et al., 2005).

Crystallography

Crystals of all three complexes were produced by vapor diffusion, using the hanging drop method. Concentrated complexes (10 mg/ml) were mixed 1 μ l:1 μ l with reservoir solution, and crystals grew overnight. Reservoir solution contained 0.6 M NaH₂PO₄, 1.4 M K₂HPO₄ and 5% glycerol for preE_M-let-7d and preE_M-let-7f-1 complexes; for preE_M-let-7g, it contained 0.1M Tris (pH 8.0), 32% w/v PEG 4000, and 0.2 M Sodium Acetate. Crystals were harvested with mother liquor supplemented with 20% glycerol and frozen in liquid nitrogen. Diffraction data were indexed and scaled using XDS (Kabsch, 2010) and SCALA (Evans, 2006) in a workflow provided by autoPROC (Vonnrhein et al., 2011). Experimental phases were obtained for Lin28:pre-let-7d complex by anomalous scattering from zinc atoms (SAD), using HKL2MAP (Schneider and Sheldrick, 2002) and AutoSol (Terwilliger et al., 2009). The structures of Lin28:preE_M-let-7f-1 and Lin28:preE_M-let-7g were solved by molecular replacement with Lin28:preE_M-let-7d as search model using Phaser (McCoy et al., 2007). Density modification and NCS averaging over 6 or 2 copies were performed with PHENIX (Adams et al., 2010) to obtain electron density maps for model building with COOT (Emsley and Cowtan, 2004) and PHENIX was used for further refinement. Final rounds of refinement were carried out using BUSTER with local structure similarity restraints (LSSR) and TLS (Bricogne et al., 2011).

Equilibrium Sedimentation

Complexes containing indicated protein and RNA constructs were purified as described above. Three concentrations (Absorbance₂₈₀ = 0.2, 0.4, 0.6) were measured for each complex. Data were collected on a Beckman Optima XL-A ultracentrifuge at 4 speeds (15, 18, 21, 24K RPM) and analyzed by fitting to a single-species model using Origin. Partial specific volumes for each complex was calculated by using NucProt (Voss and Gerstein, 2005).

SUPPLEMENTAL REFERENCES

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221.
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O.S., Vonnrhein, C., et al. (2011). BUSTER version 2.11.1 (Cambridge, UK: Global Phasing Ltd).
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277–293.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* 62, 72–82.
- Kabsch, W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125–132.

- Kay, L.E., Torchia, D.A., and Bax, A. (1989). Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28, 8972–8979.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Cryst.* 40, 658–674.
- Schneider, T.R., and Sheldrick, G.M. (2002). Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* 58, 1772–1779.
- Schwarzinger, S., Kroon, G.J., Foss, T.R., Chung, J., Wright, P.E., and Dyson, H.J. (2001). Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* 123, 2970–2978.
- Terwilliger, T.C., Adams, P.D., Read, R.J., McCoy, A.J., Moriarty, N.W., Grosse-Kunstleve, R.W., Afonine, P.V., Zwart, P.H., and Hung, L.-W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D Biol. Crystallogr.* 65, 582–601.
- Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011). Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* 67, 293–302.
- Voss, N.R., and Gerstein, M. (2005). Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. *J. Mol. Biol.* 346, 477–492.
- Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59, 687–696.

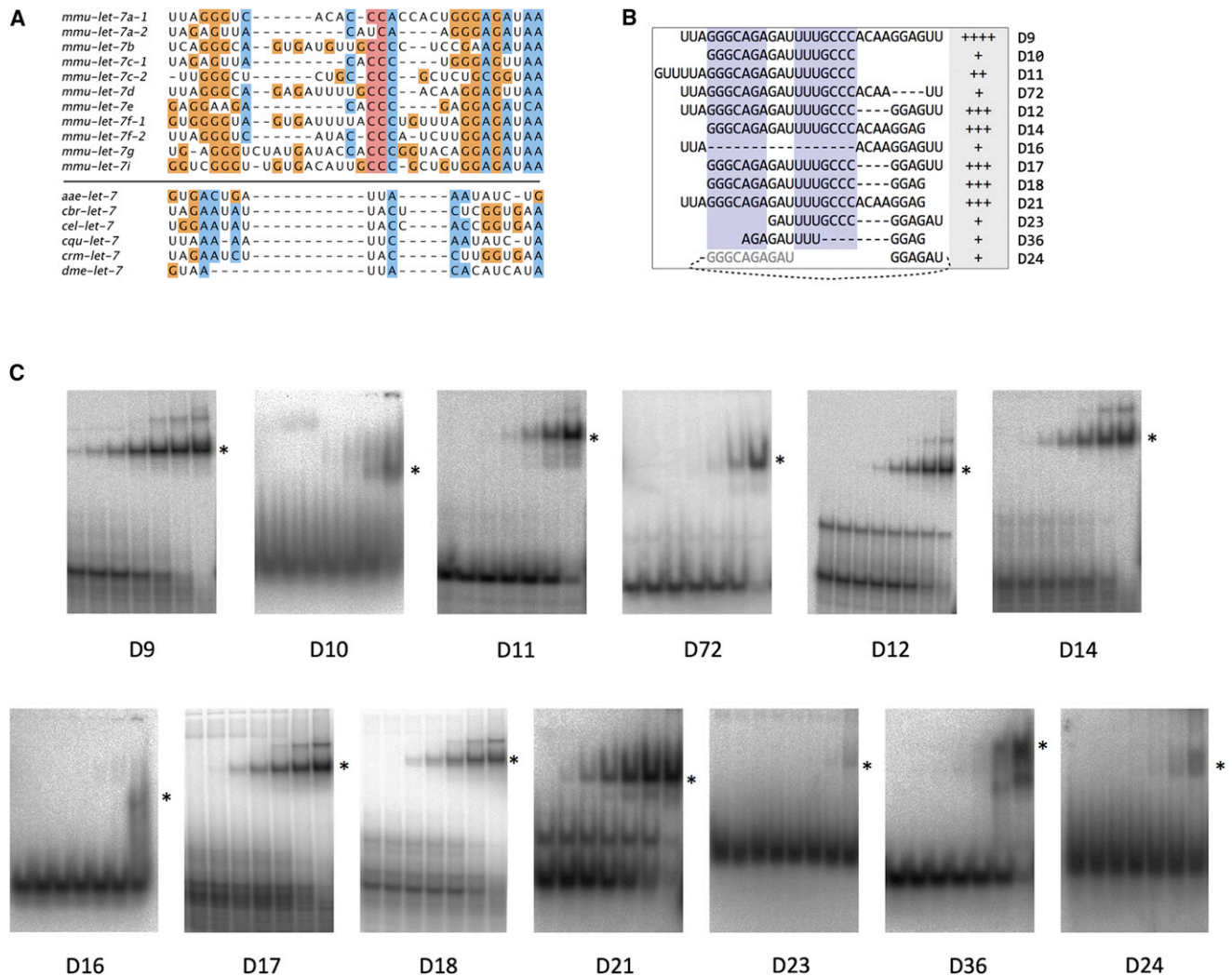


Figure S1. Mapping of *let-7* Precursors for Binding to Lin28, Related to Figure 1

(A) Sequence alignment of pre-elements (preEs) from *let-7* homologs. Upper block contains mouse homologs, and the lower block contains *dme* (*Drosophila melanogaster*), *crm* (*Caenorhabditis remanei*), *cqu* (*Culex quinquefasciatus*), *cel* (*Caenorhabditis elegans*), *cbr* (*Caenorhabditis briggsae*), and *aae* (*Aedes aegypti*). The colors follow Clustalx scheme.

(B) Alignment of the tested oligonucleotide sequences for electrophoretic mobility shift assays. Scores for binding are defined by apparent dissociation constant ranges: +, +, +, +, 0.2–1.5 μ M; +, +, 1.5–3 μ M; +, 3–15 μ M; +, >15 μ M. D24 starts with GGAG with the gray sequence attached to the 3' end of the GGAG motif.

(C) Representative EMSAs using probes indicated, titrated with increasing concentrations of Lin28(16–184) from left to right (8 nM, 33 nM, 130 nM, 520 nM, 2.1 μ M, 8.3 μ M, 33.3 μ M). Binding affinities were scored using the major complex band (marked with an asterisk at the right side). Other minor bands sometimes appear, but they are not reproducible and seem to depend on the protein/RNA batch.

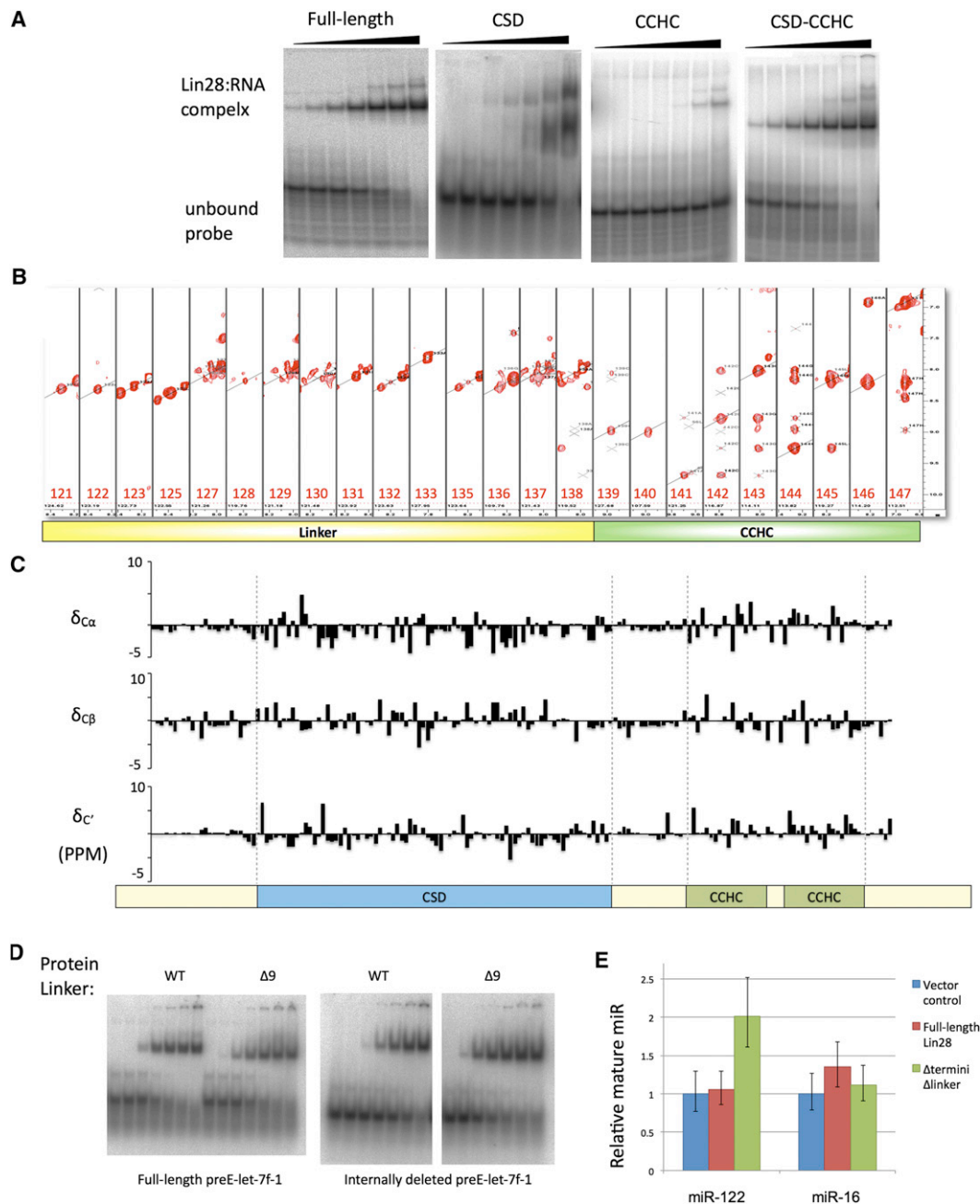


Figure S2. Domain Mapping and Structural Analysis of Lin28 Protein, Related to Figure 2

(A) EMSA of Lin28 truncations. Full-length (1–209) and CSD-CCHC (16–184) constructs have comparable affinity but isolated CSD (16–126) or isolated CCHC \times 2 (134–184) do not. Full-length preE-let-7d was used as probe, titrated with protein constructs indicated (left to right: 8 nM, 33 nM, 130 nM, 520 nM, 2.1 μ M, 8.3 μ M, 33.3 μ M).

(B) 15 N-NOESY of the linker region. Homonuclear $^1\text{H}/^1\text{H}$ NOE spectral strips belonging to the residue numbers marked in red were taken from a 3D ^{15}N -selected NOESY-HSQC, and only the amide region is shown for inter-residue backbone NOEs. Diagonal peaks are marked with a diagonal line and crosspeaks are marked with a cross. The linker region is missing inter-residue NOEs, whereas CCHC has amide-amide interactions evident from crosspeaks.

(C) Plot of secondary shifts for $C\alpha$, $C\beta$, and C' versus residue number. Domain boundaries, consistent as in Figure 2A, are marked with dashed gray lines.

(D) EMSA of protein linker deletion constructs. Increasing amounts of protein (Lin28 35–187, with or without 9 residue internal deletion in the linker region) were added to radioactively labeled fragment of pre-let-7f-1 indicated. The concentrations of protein in each lane are as follows: 5 nM, 20 nM, 78 nM, 313 nM, 1.2 μ M, 5 μ M, 20 μ M.

(E) In vivo processing assay of pri-miR-122 and pri-miR-16, similar to main Figure 2D. 293T cells (12 well) were cotransfected with Lin28 (100 ng) with pri-miR-122 or pri-miR-15-16 (750 ng). Standard deviations from three experiments are shown.

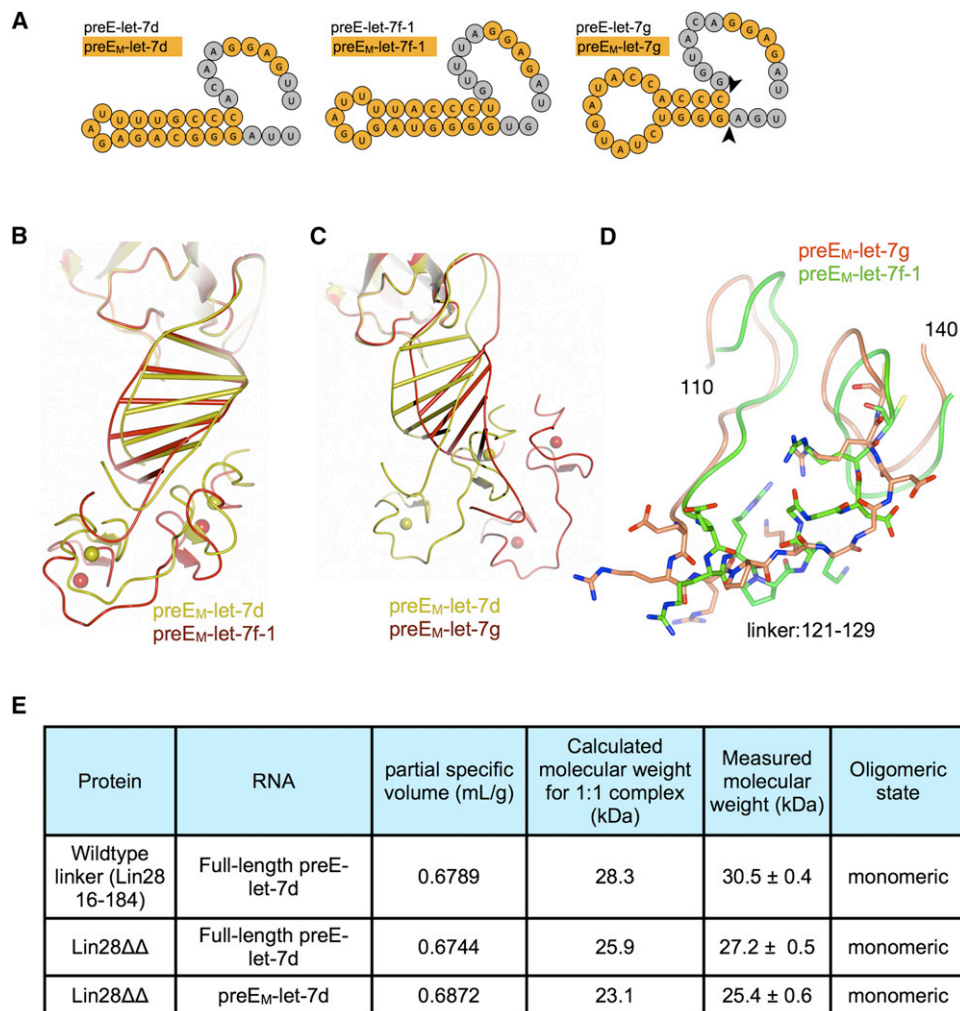


Figure S3. CocrySTALLIZATION of Lin28 with Sequences from Three *let-7* Precursors, Related to Figure 3

(A) Minimal preE (preE_M, orange) used for cocrySTALLIZATION is shown for each *let-7*, in the context of the full-length preE. Grey nucleotides were removed to reduce flexibility for crystallization. For preE-*let-7g*, another GC pair was added at the position marked with arrow heads to stabilize the stem structure.

(B) Comparison of Lin28:preE-*let-7d* structure with preE-*let-7f-1* complex shows that when the CSDs are superimposed, the CCHCx2 shifts according to the longer stem length of preE-*let-7f-1*.

(C) Comparison of Lin28:preE-*let-7d* structure with preE-*let-7g* complex shows that due to differences near the ds-ss junction in the CSD-binding region, the overall axis of the preE-stem is tilted. Again, the CCHCx2 follows the GGAG motif, indicating specific binding.

(D) Interdomain linker of preE-*let-7g* is compared to preE-*let-7f-1* and shows variability. Lack of clear density prevented modeling of the preE-*let-7d* linker.

(E) Equilibrium sedimentation of wild-type and truncated linker constructs shows that the complexes are monomeric in solution, in contrast to what is seen in cocrySTALLS.

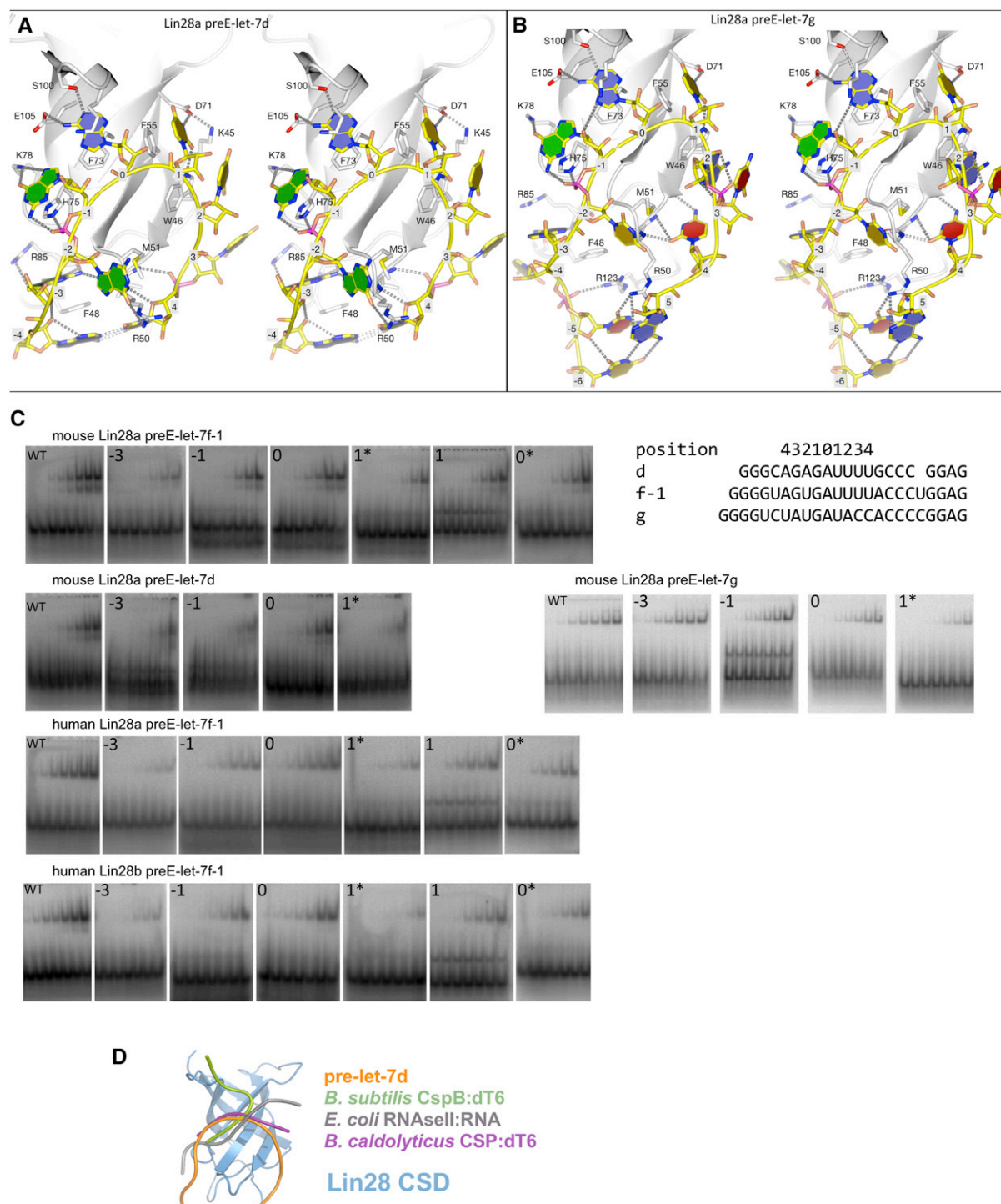


Figure S4. CSD:RNA Interactions, Related to Figure 4

(A and B) Stereo representation of a detailed view of the CSD:preE loop of the complex indicated.

(C) EMSA using the protein and preE_M RNA combination indicated above each row. Each panel represents a titration using the particular mutation at position marked at the left top corner. An asterisk indicates a transversion mutation, and plain numbers indicate transition mutations. Actual sequences for the probes and mutations are shown in main Figure 4.

(D) Superimposition of previously determined CSD:RNA complexes with Lin28:pre-let-7. Only the CSD from Lin28 is shown because all four protein models overlap well (blue cartoon), and RNA backbone is shown in indicated colors.

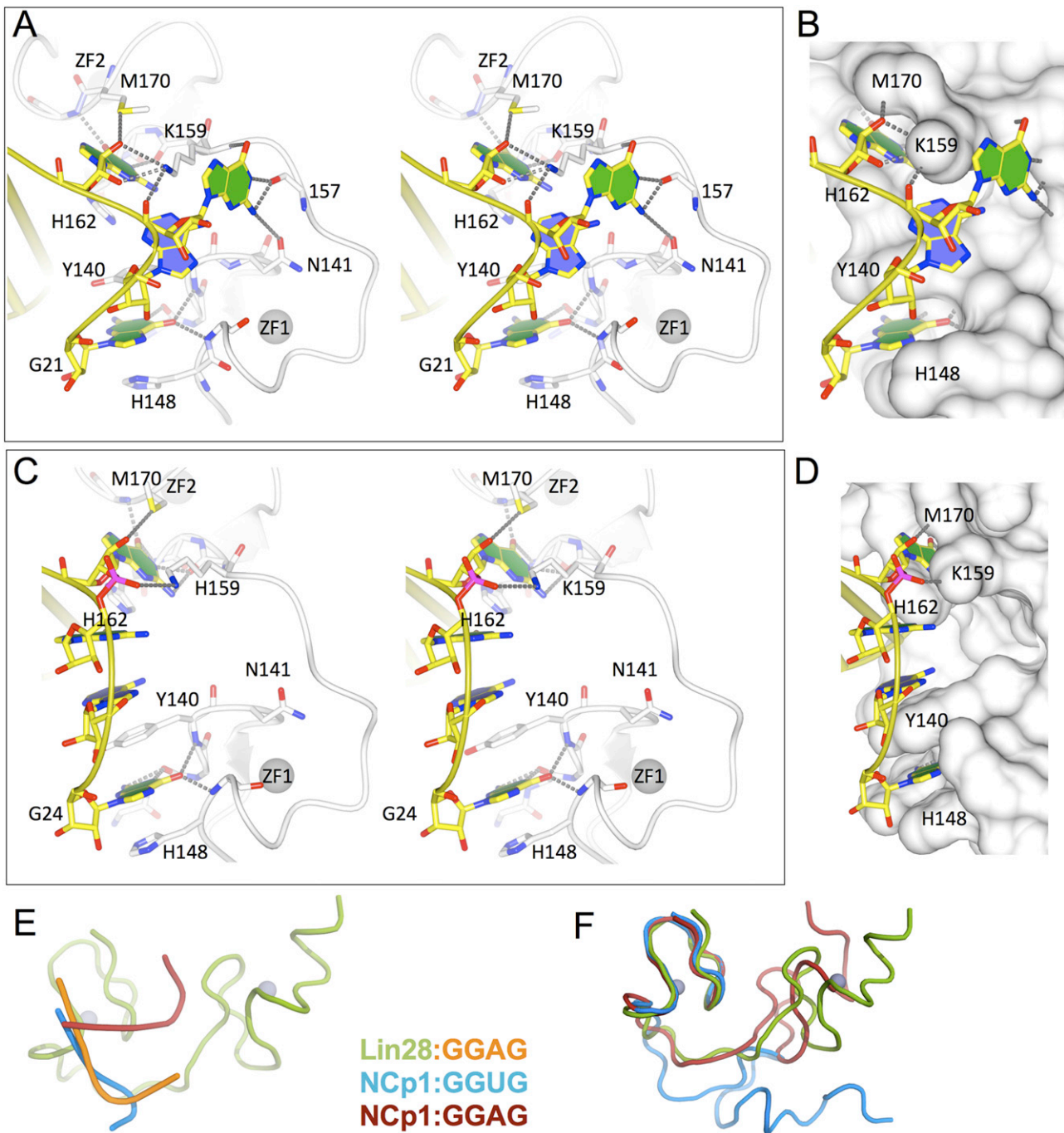


Figure S5. CCHCx2:RNA Interactions, Related to Figure 5

(A) Stereo representation of a detailed view of CCHCx2:GGAG interactions, using the Lin28:preE_M-let-7d complex structure. This conformation is observed in all copies of preE-let-7d, both copies of preE-let-7f, and chain B of preE-let-7g.
 (B) Identical orientation as in (A), but protein is shown with surface representation.
 (C) Same as (A) but for the Lin28:preE_M-let-7g structure, chain A.
 (D) Identical orientation as (C), but protein is shown with surface representation.
 (E and F) CCHCx2 conformations are variable. The CCHCx2 from the Lin28 (green):preE_M-let-7d (orange) complex is compared with NMR structures of HIV NCp1 CCHCx2:RNA complexes by superimposition of the first CCHC. Only RNA backbone is shown in (E) against Lin28, and only protein backbone is shown in (F) for clarity.

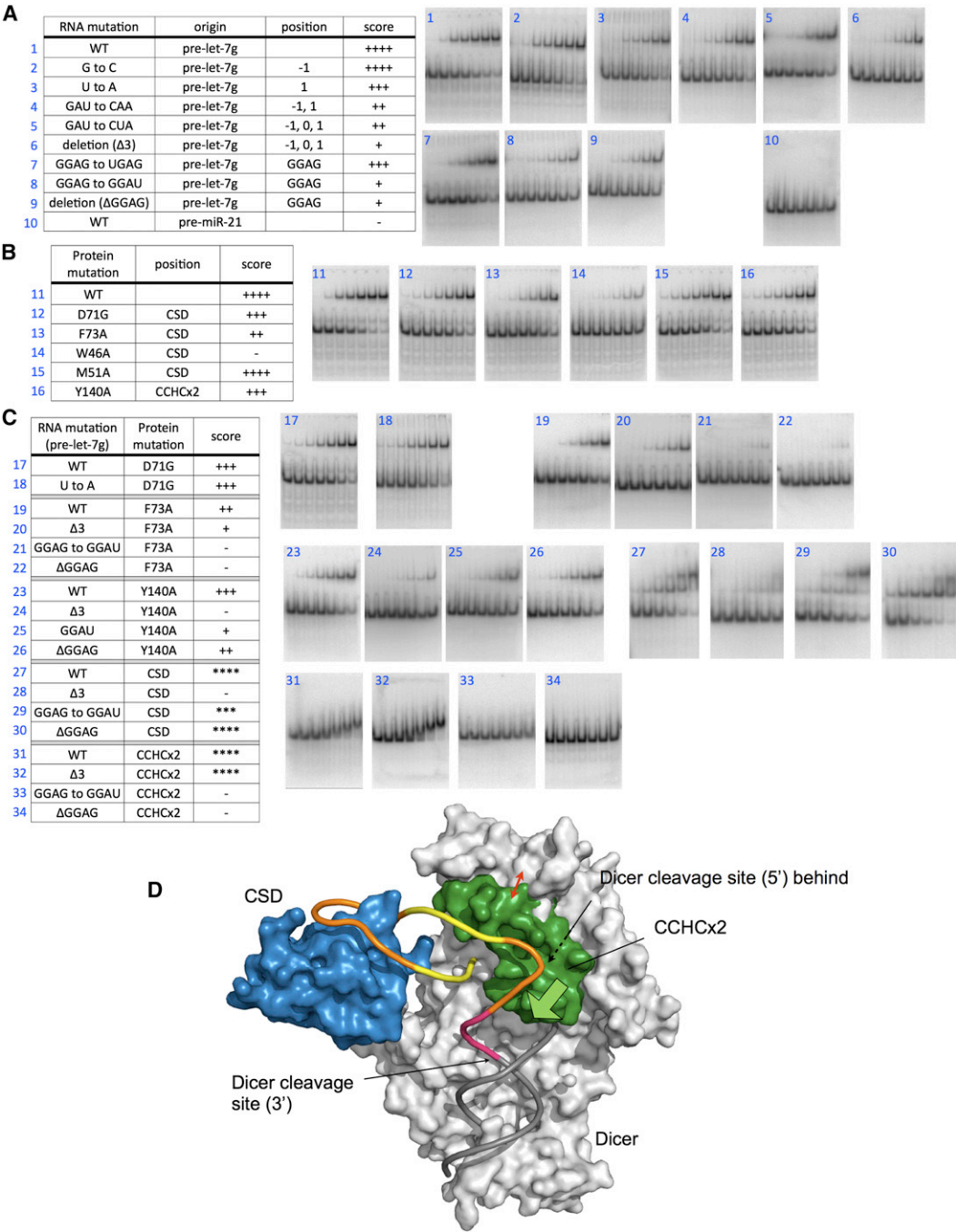


Figure S6. Interactions of Full-Length Lin28 with Full-Length Pre-let-7, Related to Figure 6
(A–C) EMSA using full-length molecules. Gels to accompany the tables shown in main Figure 6 are shown here, ordered left to right and top to bottom, according to the order in each table.
(D) Model of the Lin28:pre-let-7 complex binding to Dicer. Dicer and dsRNA substrate complex was modeled as referenced in text. The composite model with Lin28:preE-let-7 was generated by connecting the preE structure to an ideal dsRNA helix, connected by a flexible pink linker. Due to the limited number of bases on the 5' region of preE (hidden behind the CCHCx2), Lin28:preE-let-7 crystal structure portion cannot be peeled away from Dicer more without melting much of the mature region (gray loops). The direction of the CCHCx2 to wedge into the dsRNA is shown with a green arrow. Both CCHCx2 protein volume and unmodeled linker between the CSD and CCHCx2 would clash with Dicer, as shown with a red arrow. (Blue, CSD; green, CCHCx2; orange and yellow, preE-let-7f-1 included in crystal structure, orange is for direct contacts with Lin28; pink, preE-let-7f-1 not included in the structure; gray loop, portion of mature region of pre-let-7; gray surface, dsRBD and RNaseIIIb dimer from mouse Dicer.)

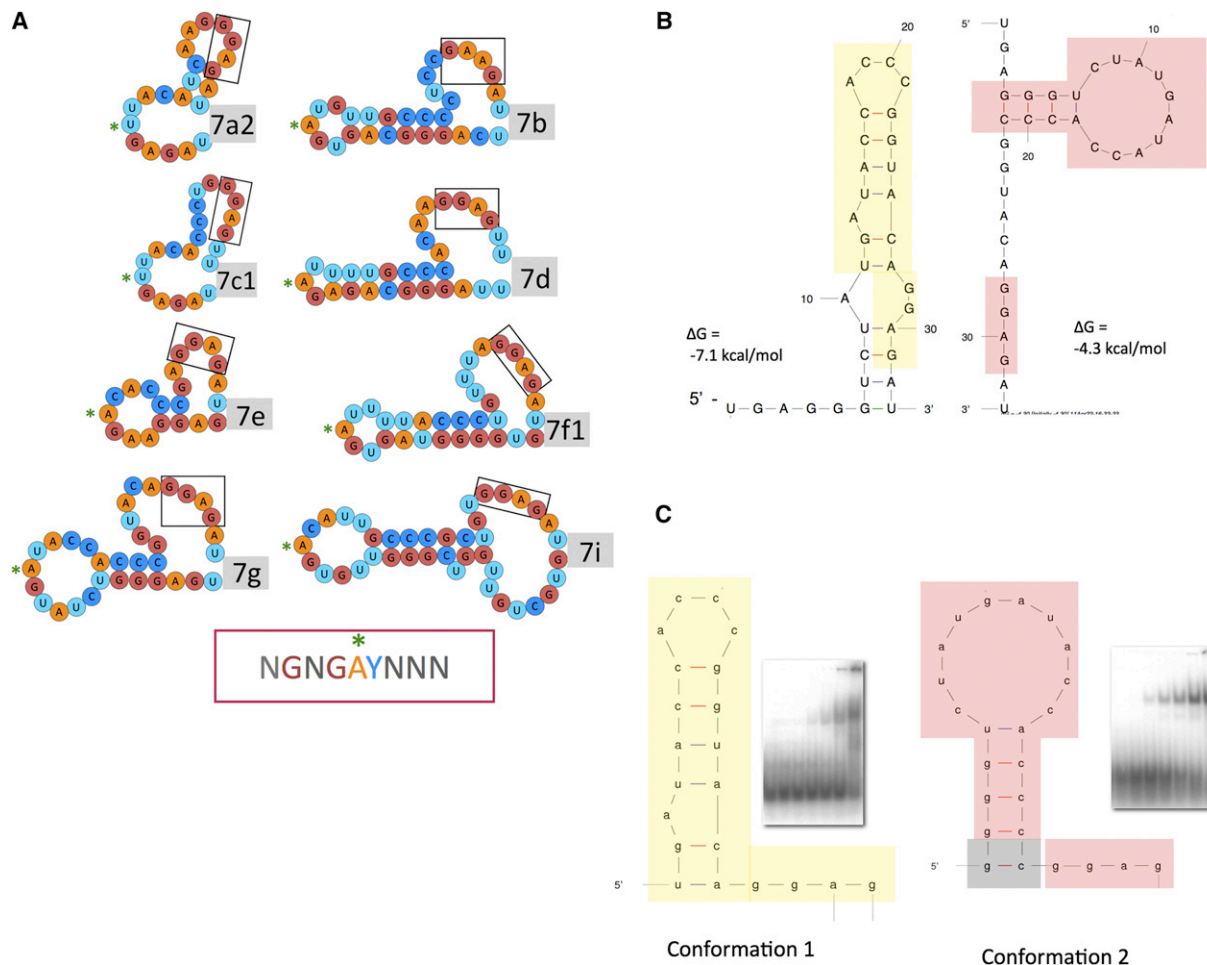


Figure S7. Structural Model Suggests How Lin28 Would Bind to Various *let-7* Family Members, Related to Figure 7

(A) Proposed Lin28-binding sites on a selection of mouse preE-*let-7*s. The entire preE-*let-7* sequences are shown, as seen in Figure S1A. The optimal conformation for binding was chosen among the top 3 structures calculated by mfold. The minimal RNA for CSD binding is shown in red box, with sequence preference (Y = pyrimidine, N = any). The central position in the loop (zero as defined in the text) is marked with an asterisk and the GGAG motif is indicated with boxes outlined in black. Although some CSD-binding sites do not exactly match the preferred sequence, our mutagenesis results show that these RNA mutations only slightly reduce affinity to Lin28.

(B) Predicted structures of preE-*let-7g* by mfold as referenced in text. Colored blocks indicate portions used to generate minimal Lin28-binding constructs shown in (C).

(C) Predicted structures of fragments of preE-*let-7g* designed for complex formation with Lin28. Two constructs of stem-loop with a 3' tail were designed using the two predicted structures in (A). An extra G-C base pair was added distal to the preE-loop to aid with crystallization. EMSA results using the diagrammed probes in (C) show that conformation 2 construct binds Lin28 with much higher affinity.