Published online 19 October 2011 *Nucleic Acids Research, 2012, Vol. 40, No. 4* ***1767–1777***
doi:10.1093/nar/gkr808

Importance of the NCp7-like domain in the recognition of pre-let-7g by the pluripotency factor Lin28

Alexandre Desjardins, Ao Yang, Jonathan Bouvette, James G. Omichinski and Pascale Legault*

Département de Biochimie, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, QC, Canada, H3C 3J7

Received June 9, 2011; Revised September 11, 2011; Accepted September 13, 2011

ABSTRACT

The pluripotency factor Lin28 is a highly conserved protein comprising a unique combination of RNA-binding motifs, an N-terminal cold-shock domain and a C-terminal region containing two retroviral-type CCHC zinc-binding domains. An important function of Lin28 is to inhibit the biogenesis of the let-7 family of microRNAs through a direct interaction with let-7 precursors. Here, we systematically characterize the determinants of the interaction between Lin28 and pre-let-7g by investigating the effect of protein and RNA mutations on *in vitro* binding. We determine that Lin28 binds with high affinity to the extended loop of pre-let-7g and that its C-terminal domain contributes predominantly to the affinity of this interaction. We uncover remarkable similarities between this C-terminal domain and the NCp7 protein of HIV-1, not only in terms of primary structure but also in their modes of RNA binding. This NCp7-like domain of Lin28 recognizes a G-rich bulge within pre-let-7g, which is adjacent to one of the Dicer cleavage sites. We hypothesize that the NCp7-like domain initiates RNA binding and partially unfolds the RNA. This partial unfolding would then enable multiple copies of Lin28 to bind the extended loop of pre-let-7g and protect the RNA from cleavage by the pre-microRNA processing enzyme Dicer.

INTRODUCTION

MicroRNAs (miRNAs) are short single-stranded RNAs of ~22 nt found in virus, plant and animal species that act as post-transcriptional regulators of mRNA expression [for recent reviews, see (1–4)]. They are generated from a longer RNA, the primary transcript (pri-miRNA), by a multi-step process. The pri-miRNA is first cleaved by the microprocessor complex containing the endonuclease Drosha and the double-stranded RNA-binding protein DGCR8 to produce a 60–70 nts RNA hairpin known as the precursor miRNA (pre-miRNA). After being exported to the cytoplasm, the pre-miRNA is further cleaved by the endonuclease Dicer to form a ~22-nt dsRNA. The single-stranded mature miRNA is then loaded into the RNA-induced silencing complex to regulate its target mRNAs.

miRNAs play important roles in cell differentiation (5–7), and, in mammals, several miRNAs have been shown to act as oncogenes and tumor suppressors [reviewed in (8–13)]. Among those playing a role as tumor suppressors, the let-7 family of miRNAs have been extensively characterized, and are known inhibitors of oncogenes such as RAS, MYC, HMGA2, and cyclin D1 (10). The let-7 miRNAs are often present in multiple copies in a single genome, with the mature let-7 being highly conserved across species. In human and mouse, there are 10 mature let-7 family sequences (let-7a, let-7b, etc.) produced from 13 precursors.

Although levels of let-7 pri-miRNAs are controlled by transcription factors, post-transcriptional regulation is critical in determining the levels of mature let-7 miRNAs (14–18). Recent studies in embryonic cells have highlighted the importance of Lin28 in post-transcriptional regulation of the let-7 family of miRNAs, where it acts as a selective inhibitor of let-7 miRNAs maturation (19–21). The various members of the let-7 family are not affected to the same degree by Lin28, with let-7a, let-7d and let-7g being among the most affected. Several mechanisms have been proposed to explain the Lin28 inhibition of let-7 biogenesis. Lin28 was shown to interfere with the Drosha cleavage of pri-let-7 (16,19,21) and with the cleavage of pre-let-7 by Dicer (22,23). In addition, Lin28 was shown to induce the uridylylation of pre-let-7 by the recruitment of TUT4 (Zcchc11), which leads to its

*To whom correspondence should be addressed. Tel: 514 343 7326; Fax: 514 343 2210; Email: pascale.legault@umontreal.ca

degradation (22,24–26). Although the relative importance of these mechanisms *in vivo* has not been clearly established (27), they all involve the formation of a complex between Lin28 and the immature forms of the let-7 miRNA.

Lin28 is a highly conserved protein of 209 amino acids known to be an important pluripotency factor (28), and its role in pluripotency is likely related to its function in let-7 biogenesis (19,29). Lin28 contains a unique set of RNA-binding motifs (30,31); an N-terminal cold shock domain (CSD) and a C-terminal region composed of two CCHC-type zinc-binding domains [ZBDs; (30)]. CSDs are found in several RNA- and DNA-binding proteins (32), whereas the CCHC-type ZBDs are most commonly found in retroviral nucleocapsid proteins, such as the NCp7 protein from HIV-1 (33). Although Lin28 has been shown to regulate the stability and translation of selected mRNAs (34–37), it plays a central role in regulating levels of mature let-7.

Several *in vivo* and *in vitro* studies have sought to characterize the interaction between pre-let-7 and Lin28 (19,20,23,24,38). It was demonstrated that both the CSD and the ZBDs of Lin28 are necessary for pre-let-7g binding *in vitro* and maturation inhibition *in vivo* (20). As determined by *in vitro* binding assays, Lin28 binds the extended terminal loop of pre-let-7g (20,38). Mutation of a conserved cytosine in this loop was shown to reduced its *in vitro* affinity for Lin28 (20). A G-rich sequence at the 5′-end of the pre-let-7g terminal loop was found to be strongly protected from ribonuclease cleavage by Lin28 (38). In addition, mutations of a few conserved nucleotides in the terminal loop make the immature miRNA resistant to Lin28 inhibition in P19 embryonal carcinona extract (19). Lin28 also binds the extended terminal loop of pre-let-7a-2, and the sequence composing the mature miRNA (let-7a) can compete with pre-let-7a-2 binding for Lin28 (23). Moreover, a four-nucleotide 5′-GGAG-3’ sequence important for Lin28 binding and its uridylylation by Zcchc11 was identified at the 3′-end of the terminal loop region of pre-let-7a-1 (24). Although several studies have contributed to establish that the RNA-binding domains of Lin28 are important for recognition of the extended terminal loop of pre-let-7, the key determinants of this interaction have not been systematically defined.

In this work, we used electrophoretic mobility shift assay (EMSA), ribonuclease protection assay, in-line probing and NMR spectroscopy with purified molecules to map the interaction between the pre-let-7g RNA and the Lin28 protein from mouse. We determine that the C-terminal domain of Lin28 contributes predominantly to the high-affinity interaction with pre-let-7g and its sequence is very similar to the NCp7 protein of HIV-1. We also uncover several similarities in terms of RNA binding between NCp7 and the C-terminal domain of Lin28.

MATERIALS AND METHODS

Plasmids

Lin28 expression vectors are derived from pGEX4T (GE Healthcare) and were constructed using the Lin28 cDNA from *Mus musculus* (Open Biosystems BC068304). Vectors for RNA transcription were derived either from the pARiBo1 plasmid or the pRSA-VS plasmid (39). Mutant vectors were prepared using the Strategene QuikChangeII site-directed mutagenesis method or by standard cloning of restriction fragments. All plasmids created for this study were verified by DNA sequencing.

RNA preparation for biochemical characterization and NMR studies

Most RNAs used here were transcribed *in vitro* as ARiBo-tagged precursors and purified by batch affinity purification (39,40). In one case (TL-let-7g), the RNA was synthesized *in vitro* as a precursor with a VS ribozyme substrate at its 3′-end and purified as described previously (41). For biochemical characterization (gel-shift, footprinting and in-line probing), the RNAs were [5′-^{32}P]-labeled and further purified by 20% denaturing gel electrophoresis (42). For NMR studies, the RNAs were concentrated and exchanged with an Amicon Ultra-15 3000 NMWL (Millipore) in NMR buffer (10 mM d_{18}-HEPES at pH 6.4, 50 mM NaCl, 0.05 mM NaN_3 and 10% D_2O).

Protein expression and purification for biochemical characterization

Lin28 and related mutants were expressed in *Escherichia coli* strain BL21 cells (Stratagene). The bacterial cultures were grown in LB medium at 37°C and induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 4 h at 30°C. The cells were harvested by centrifugation and resuspended in binding buffer [25 mM Tris, pH 8.0, 1 M NaCl, 0.1% NP-40 alternative (Calbiochem) and 1 mM DTT] supplemented with Complete EDTA-free protease inhibitor (Roche) and 10 U/ml of DNase I recombinant RNase-free (Roche). The cells were lysed by French press and centrifuged at 100 000*g* for 1 h at 4°C. The supernatant was incubated for 1 h at 4°C with GSH-Sepharose 4B resin (GE Healthcare). After incubation, the resin was washed three times with the binding buffer and three times with the S7 buffer (50 mM Tris, pH 8.0, 50 mM NaCl, 5 mM $CaCl_2$ and 1 mM DTT). The washed resin was resuspended in S7 buffer and incubated overnight at room temperature with 5 U/ml of Nuclease S7 (Roche). The resin was subsequently washed three times with S7 buffer and incubated 1 h at room temperature with 100 U of thrombin (Calbiochem). The eluted protein was dialyzed 4 h at 4°C in 2 l of 20 mM Tris, pH 8.0, 2 M NaCl, 1 mM DTT and overnight at 4°C in 20 mM Tris, pH 8.0, 1 M urea, 200 mM NaCl and 1 mM DTT. The dialyzed protein was loaded on an SP-Sepharose high-performance column (GE Healthcare) equilibrated with FPLC-A (20 mM Tris, pH 8.0 and 1 mM DTT). The protein was eluted from the column using a gradient (from 0% to 100% over 525 mL) of FPLC-B (20 mM Tris, pH 8.0, 2 M NaCl and 1 mM DTT). The fractions containing the protein were combined, concentrated with an Amicon Ultra-15 3000 NMWL (Millipore) and dialyzed in storage buffer (100 mM Tris, pH 7.6, 100 mM NaCl, 20% glycerol and 2 mM DTT). The NCp7 protein was expressed and purified as described

previously (43). All proteins purified for this study were verified by mass spectrometry.

Protein expression and purification for NMR studies

For NMR studies, uniform ^{15}N- and $^{15}N/^{13}C$ labeling was obtained by growing the cells in minimal media containing ^{15}N-labeled NH_4Cl and $^{13}C_6$-glucose as the sole sources of nitrogen and carbon, respectively. Protein purification was conducted as described above, but with the following modifications. The selected fractions from the SP-Sepharose column were dialyzed in 5% acetic acid, concentrated on a rotary evaporator and purified on a Vydac C_4 reverse-phase HPLC column using an acetonitrile gradient (from 15% to 35% over 335 ml) in 0.05% TFA. After HPLC purification, the proteins were refolded in the presence of zinc, as described previously (44).

Electrophoretic mobility shift assay

For EMSA, the ^{32}P-labeled RNA was first heated and snap cooled (heated 2 min at 95°C and snap-cooled on ice for 5 min) to promote hairpin formation. The protein samples were diluted in EMSA buffer (50 mM Tris, pH 7.6, 50 mM NaCl, 10% glycerol, 0.05% NP-40 alternative and 2 mM DTT) and their concentrations were adjusted to span from 0.02× to 50× of the estimated K_d. The binding reactions (20 ml) were initiated by mixing 1 pM of ^{32}P-labeled RNA with the diluted proteins and incubated at 4°C for 30 min. For each K_d determination, ~14 binding reactions were loaded directly on an 8% native polyacrylamide gel (37.5:1 polyacrylamide/bisacrylamide) and run in Tris–Glycine buffer (25 mM Tris–Base and 200 mM glycine) at 200 V for 2 h with active water cooling in the cold room. The gels were fixed in 50% methanol and 10% acetic acid for 1 h, washed 15 min in 30% ethanol, quickly rinsed with H_2O and exposed overnight to a storage phosphor screen (Bio-Rad). The ^{32}P-labeled RNA was visualized with a Bio-Rad Molecular Imager FX densitometer, and band intensities were quantified using the QuantityOne software (version 4.6.6 from Bio-Rad). The fraction of bound RNA was plotted against protein concentration, and the data were fitted to the one-site binding equation or to the Hill equation (only for the two cases in Table 1) by nonlinear regression analysis within the Origin 7 SR4 version 7.0552 software (OriginLab, MA, USA). For each protein–RNA complex, at least three independent K_d determination experiments were performed. The reported K_d's and their errors are, respectively, the average values and the standard deviations from these multiple experiments.

In-line probing assay

In-line probing assays were performed as described previously (45). The ^{32}P-labeled RNA was visualized with a Bio-Rad Molecular Imager FX densitometer, and band intensities were quantified using the Image Lab software (version 3.0 from Bio-Rad).

RNA footprinting assay

For RNase footprinting, the ^{32}P-labeled RNA was first heated and snap cooled to promote hairpin formation. The $Lin28_{119–180}$ protein was diluted at various concentrations in EMSA buffer. The protein was first incubated with 1 nM of ^{32}P-labeled RNA (10 ml total volume) for 30 min at 4°C. Then, 1 U of T_1 ribonuclease from *Aspergillus oryzae* (Sigma) was added and the incubation continued for 15 min at 4°C. The reaction was stopped by the addition of Precipitation/Inactivation buffer (Ambion), incubation for 15 min at −20°C and centrifugation at 16 000*g* for 15 min. The RNA pellet was dissolved in Gel loading buffer II (Ambion), loaded on a 20% polyacrylamide/7 M urea sequencing gel and run at 1900 V for 5 h. The sequencing gel was exposed 2 h to a storage phosphor screen (Bio-Rad). The ^{32}P-labeled RNA bands were visualized and quantified as for the EMSA assay.

NMR Spectroscopy

For NMR studies, the following samples were prepared in NMR buffer: 1.0 mM ^{15}N-labeled $Lin28_{136–180}$; 1.3 mM ^{15}N-labeled $Lin28_{119–180}$; 1.1 mM $^{13}C/^{15}N$-labeled $Lin28_{119–180}$; 0.1 mM TL-let-7g:^{15}N-labeled $Lin28_{119–180}$; 0.1 mM Δbulge TL-let-7g:^{15}N-labeled $Lin28_{119–180}$; 0.1 mM G34AG35A TL-let-7g:^{15}N-labeled $Lin28_{119–180}$; 1.0 mM TL-let-7g:^{15}N-labeled $Lin28_{119–180;}$ and 1.3 mM TL-let-7g:$^{13}C/^{15}N$-labeled $Lin28_{119–180}$. For the TL-let-7g:$Lin28_{119–180}$ complexes, the samples were prepared by titration of $Lin28_{119–180}$ into a TL-let-7g sample. All NMR experiments were collected on Varian ^{Unity}INOVA 500 and 600 MHz spectrometers equipped with a pulse-field gradient unit and an actively shielded *z* gradient probe (either a room-temperature probe or a cryogenic probe). The backbone resonances (^1H, ^{15}N and ^{13}C) of $Lin28_{119–180}$ in the free and TL-let-7g-bound form were assigned using the following NMR experiments collected at 35°C: two-dimensional (2D) ^1H-^{15}N HSQC (46); three-dimensional (3D) HNCACB (47–49); and 3D (HB)CBCA(CO)NNH (48,49). ^1H, ^{13}C and ^{15}N chemical shifts were referenced to an external standard of 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) at 0.00 ppm (50). NMR data were processed with NMRPipe/NMRDraw (51) and analyzed with NMRView (52).

RESULTS

Lin28 recognizes the terminal loop of pre-let-7g with its C-terminal domain providing the most important energetic contribution

To identify the domain(s) of Lin28 important for binding the let-7g precursor miRNA (pre-let-7g), we used EMSAs with purified recombinant proteins and *in vitro* transcribed RNAs (Figure 1). It was previously established that Lin28 recognizes pre-let-7g and its terminal loop with a similar affinity [K_d of 1–2 μM; (20,38)]. Thus, we initiated our study by determining the K_d of full-length Lin28 ($Lin28_{1–209}$) for the terminal loop of pre-let-7g (TL-let-7g; Figure 1D). We attempted to fit the data (Figure 2A)

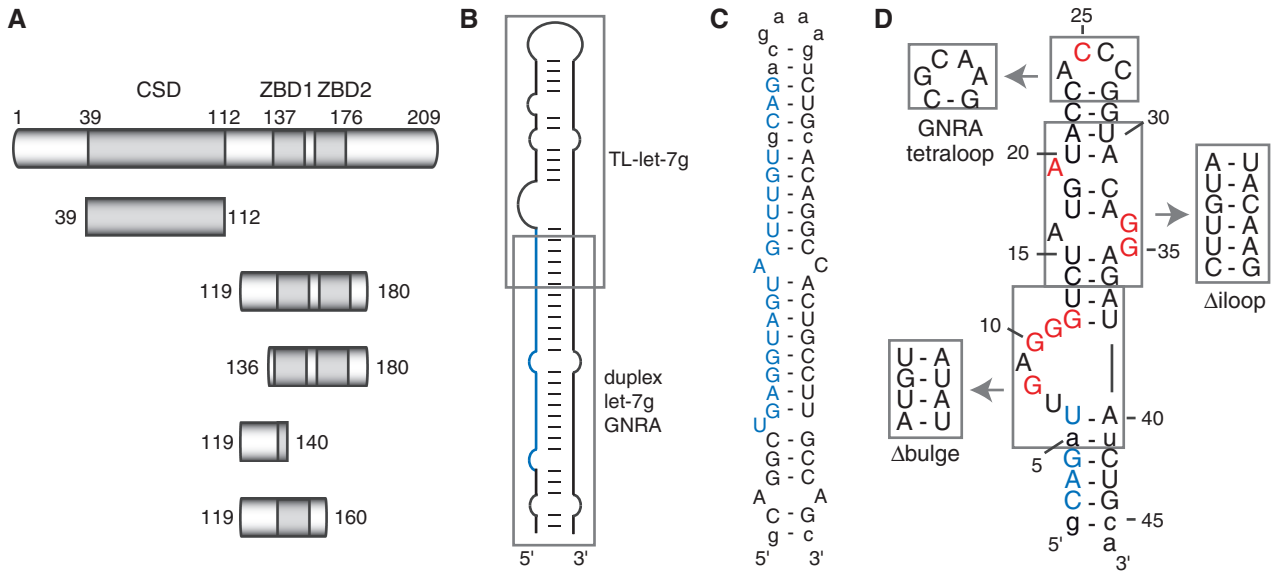

Figure 1. The Lin28 protein, pre-let-7g RNA and related sequences used in this study. (**A**) Schematic representation of the primary structures of Lin28 and deletion fragments. The gray boxes delineate sequences of known RNA-binding motifs: a cold shock domain (CSD) and a pair of retroviral–type CCHC zinc-binding domains (ZBD1 and ZBD2). (**B**) Schematic representation of pre-let-7g, indicating the regions (gray boxes) from which TL-let-7g and duplex let-7g GNRA were derived. (**C** and **D**) Primary and secondary structures of the (C) duplex let-7g GNRA and (D) TL-let-7g. Nucleotides within the mature miRNA sequence are in blue and non-natural nucleotides are shown in lowercase. In (D), site-specific mutations of TL-let-7g are in red and regions that were replaced by alternative structured elements are boxed.

using the classical one-site binding equation, but the fit was rather poor (Figure 2C, red line). Using the Hill equation, a much better fit was obtained (Figure 2C, blue line), from which we derived a K_d of 0.13 ± 0.02 nM with a Hill coefficient of 2.9 ± 0.7 (Table 1). These results clearly demonstrate that Lin28 binds the terminal loop of pre-let-7g with much higher affinity than previously reported (20,38).

Given this unexpected result, we verified the binding of Lin28 to the full-length pre-let-7g. We obtained a K_d of 0.15 ± 0.04 nM with a Hill coefficient of 2.7 ± 0.5 for this interaction (Table 1), confirming that pre-let-7g and its terminal loop have similar affinities for Lin28, as previously established (20,38). As a control, we measured the binding for Lin28 to an RNA that contains only the miRNA stem of pre-let-7g (duplex let-7g GNRA; Figure 1C) and obtained a 60-fold lower K_d (K_d = 9 ± 3 nM), further supporting that the terminal loop is the main determinant of Lin28 binding to pre-let-7g.

Since Lin28 contains two different RNA-binding domains, we determined the affinity of Lin28 fragments containing either the CSD ($Lin28_{39–112}$) or the two ZBDs ($Lin28_{119–180}$) for pre-let-7g, TL-let-7g and duplex let-7g GNRA (Figure 1). For the N-terminal $Lin28_{39–112}$ fragment comprising the CSD (Figure 1A), K_d values of 41 ± 10 nM and 126 ± 41 nM were obtained for pre-let-7g and TL-let-7g, respectively, indicating significantly weaker affinity (<250-fold) compared to the full protein. For the duplex let-7g GNRA, only a minimum K_d value could be obtained (>250 nM), because of aggregation of the $Lin28_{39–112}$ domain detected at concentrations higher than 250 nM. In contrast, the $Lin28_{119–180}$ fragment containing the ZBDs displays only slightly lower affinities than full-length Lin28 toward pre-let-7g and TL-let-7g, with K_d values of 0.6 ± 0.1 nM and 1.3 ± 0.3 nM, respectively (Table 1, Figure 2B and C). These binding data for $Lin28_{119–180}$ can be fitted to a classical one-site binding equation (Figure 2C and Table 1). For the duplex-let-7g GNRA, no specific binding could be observed with $Lin28_{119–180}$. Thus, compared to the full-length protein, the C-terminal $Lin28_{119–180}$ fragment containing the two ZBDs displays similar specificity toward RNAs derived from pre-let-7g. Furthermore, the affinity between $Lin28_{119–180}$ and TL-let-7g is only 10-fold weaker than between the full-length Lin28 protein and pre-let-7g, indicating that $Lin28_{119–180}$ and TL-let-7g encompass the main determinants of the Lin28/pre-let-7g interaction.

Similarities between the C-terminal domain of Lin28 and the HIV-1 NCp7 protein

It has been previously noted that the C-terminal domain of Lin28 contains two ZBDs similar to those found in viral nucleocapsid proteins (30). Given the importance of this RNA-binding domain for pre-let-7g binding, we searched for proteins containing a similar domain in the Swiss-Prot database using BLAST (53). The nucleocapsid proteins from simian and human immunodeficiency viruses give the highest scores after Lin28 proteins from different species. The sequence alignment between the well-characterized HIV-1 nucleocapsid NCp7 and metazoan Lin28 sequences indicates significant similarities in the ZBDs, and also, surprisingly, in the N-terminal KR-rich domain (Figure 3). The percentage of sequence similarity to HIV-1 NCp7 is relatively high for both murine (45%) and human $Lin28_{123–180}$ (45%). All three proteins contain identical zinc-chelating amino acids (CCHC) and spacing between these residues in the two ZBDs. In addition, the

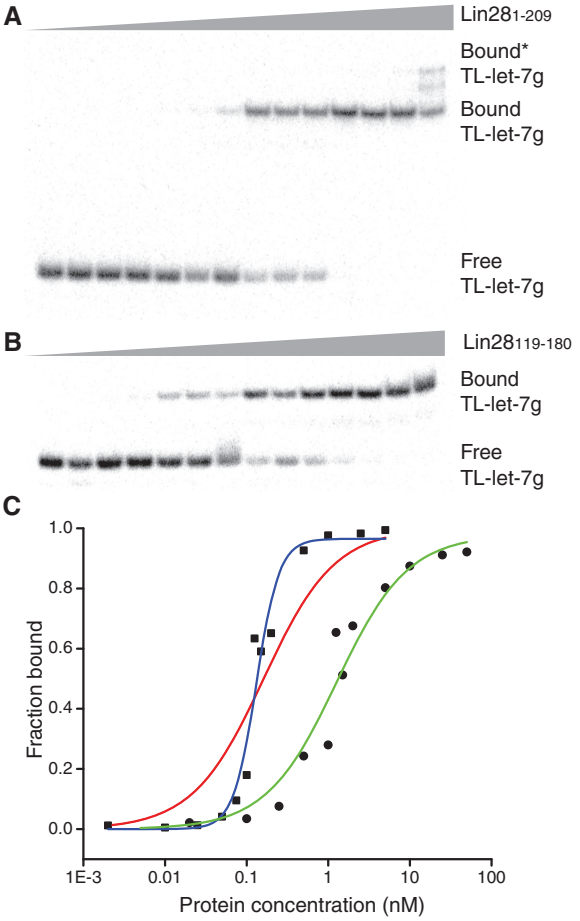

Figure 2. EMSA of TL-let-7g with $Lin28_{1–209}$ and $Lin28_{119–180}$. (**A**) Typical EMSA performed with 1 pM of 5′-[^{32}P]-labeled TL-let-7g and increasing concentrations of $Lin28_{1–209}$ (0.0, 0.002, 0.010, 0.025, 0.050, 0.075, 0.10, 0.15, 0.20, 0.50, 1.0, 2.5 and 5.0 nM). (**B**) Typical EMSA for TL-let-7g and increasing concentrations of $Lin28_{119–180}$ (0.0, 0.02, 0.10, 0.25, 0.50, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 5.0,10, 25 and 50 nM). (**C**) The bound fraction of RNA is plotted against the total concentration of protein. The data for binding of TL-let-7g to $Lin28_{1–209}$ (squares) are fitted to both the one site binding equation (red line; K_d = 0.2 nM) and the Hill equation (blue line; K_d = 0.13 nM). The data for binding of TL-let-7g to $Lin28_{119–180}$ (dots) are fitted to the one site binding equation (green line; K_d = 1.3 nM).

Table 1. Dissociation constants (K_d in nM) of different domains of Lin28 for various pre-let-7g constructs

RNA	$Lin28_{1–209}$	$Lin28_{39–112}$	$Lin28_{119–180}$
Pre-let-7g	0.15 ± 0.04 n = 2.7 ± 0.5[a]	41 ± 10	0.6 ± 0.1
Duplex let-7g GNRA	9 ± 3	>200	n.b.[b]
TL-let-7g	0.13 ± 0.02 n = 2.9 ± 0.7[a]	126 ± 41	1.3 ± 0.3

[a]In these cases, the Hill equation was used to derived the K_d values.
[b]No specific binding observed. The gel mobility shift assays display smearing and multiple shifts, indicating non-specific binding.

spacing between the two ZBDs is similar with seven residues in NCp7 of HIV-1 and eight residues in the Lin28 sequences (Figure 3). Furthermore, this sequence similarity also involves several of the NCp7 residues from the KR-rich domain and the two ZBDs that contribute to the RNA-binding interface as observed in the NMR structures of RNA/NCp7 complexes (33,54).

To further investigate the binding of $Lin28_{119–180}$ to TL-let-7g, we performed NMR chemical shift perturbation experiments. In both the free and bound forms, $Lin28_{119–180}$ displays a well-dispersed ^{1}H-^{15}N HSQC spectrum (Figure 4A), indicating that both adopt a stable homogenous conformation in solution. Analysis of the ^{1}H and ^{15}N chemical shifts reveals that 21 of the 56 amino acid residues analyzed display significant chemical shift differences between the free and RNA-bound form (Figure 4B; $\Delta\delta > 0.4$ ppm). When mapped onto the primary structure of $Lin28_{119–180}$, the residues showing significant chemical shift differences are found in the KR-rich domain, both ZBDs, as well as in the linker between ZBD1 and ZBD2. These results indicate that all these domains participate in RNA binding either by direct contact or through conformational rearrangement of the protein.

To determine if the $Lin28_{119–180}$ fragment could be shortened while maintaining its affinity for the terminal loop of pre-let-7g, we generated several N-terminal and C-terminal deletions (Figure 1A). Of the three deletion fragments that were expressed and purified ($Lin28_{136–180}$, $Lin28_{119–140}$ and $Lin28_{119–160}$), none binds TL-let-7g with high affinity ($K_d > 5\,\mu M$; Table 2), and thus $Lin28_{119–180}$ constitutes the minimal domain required for TL-let-7 g binding. It is particularly striking that removal of the first 17 amino acids encompassing the KR-rich domain is as detrimental to binding as removal of one or two ZBDs. To insure that the absence of binding with $Lin28_{136–180}$ is not due to protein misfolding, we compared the ^{1}H-^{15}N HSQC spectrum of this fragment with that of $Lin28_{119–180}$ (Supplementary Figure S1). The chemical shift similarity between these two spectra indicates that the ZBDs adopt a similar fold in $Lin28_{136–180}$ and $Lin28_{119–180}$. In addition, the importance of the KR-rich domain was further investigated using a mutant of $Lin28_{119–180}$ in which all lysines and arginines of the KR-rich domain are mutated (KR^{-} with mutations R122A, R123A, K125A, K127A, K131A, R132A, R133A and K135G). As expected, $Lin28_{119–180}$ (KR^{-}) does not bind TL-let-7g with high affinity ($K_d > 5\,\mu M$; Table 2). Interestingly, we found that the NCp7 protein of HIV-1 binds with the same affinity to TL-let-7g (1.1 ± 0.3 nM) as $Lin28_{119–180}$ (Table 2). Thus, in addition to sharing sequence similarity with NCp7, $Lin28_{119–180}$ also uses both its KR-rich and ZBDs for RNA recognition and binds with the same affinity as NCp7 to TL-let-7g. To emphasize these similarities with NCp7, we defined $Lin28_{119–180}$ as the NCp7-like domain of Lin28.

Global mapping of the interaction site using ribonuclease protection assay

A ribonuclease protection assay was used to identify the region(s) of TL-let-7g interacting with the NCp7-like domain. As a first step, in-line probing (Supplementary

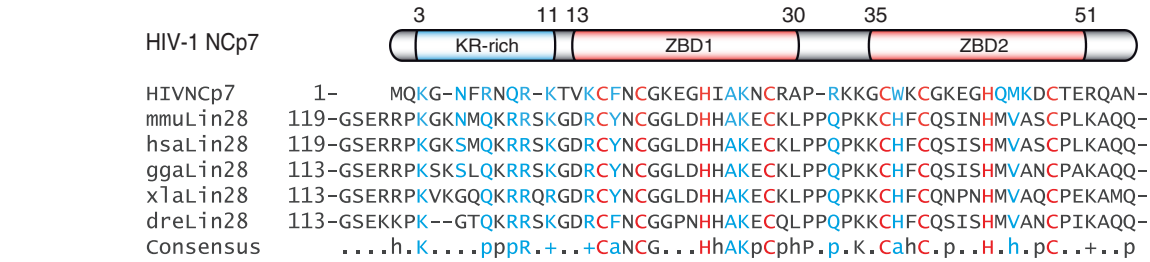

Figure 3. Sequence similarity between the HIV-1 NCp7 and the C-terminal domain of Lin28. The sequences of HIV-1 NCp7 and Lin28 from *Mus musculus* (mmu), *Homo sapiens* (hsa), *Gallus gallus* (gga), *Xenopus laevis* (xla) and *Danio rerio* (dre) were aligned using ClustalW2 (70). A consensus sequence is given with the standard one-letter code in capital letters for amino acids, as well as the following notation: a, aromatic; h, hydrophobic; p, polar; +, positively charged. The schematic representation of NCp7 highlights the domains that contribute to RNA binding: an N-terminal KR-rich domain and two zinc-binding domains (ZBD1 and ZBD2). The residues of NCp7 in red and blue make direct contact with zinc and RNA, respectively (33,54). Those residues that could play an equivalent role in Lin28 are similarly colored.

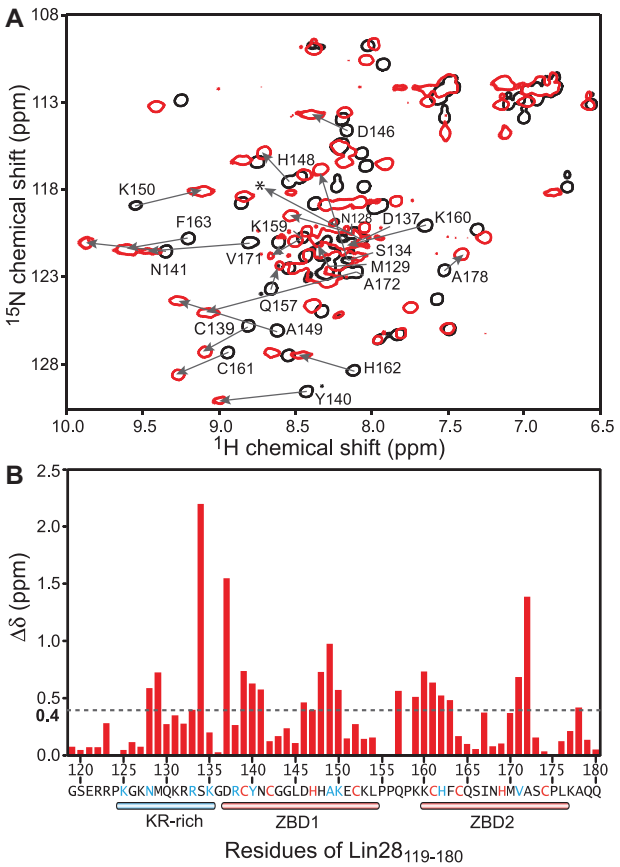

Figure 4. (**A**) Superposition of 2D ^{1}H-^{15}N HSQC spectra of 1 mM ^{13}C/^{15}N-labeled Lin28$_{119–180}$ in the free form (black) and bound to 1 mM TL-let-7g (red). The signals from the free form that display a significant chemical shift change ($\Delta\delta > 0.4$ ppm) as a result of RNA binding are annotated and the change is illustrated with an arrow. A very weak signal for D137 in the complex is indicated by a star. (**B**) Histogram displaying the differences in chemical shifts ($\Delta\delta$ in ppm ± 0.03 ppm) observed after the addition of a molar equivalent of TL-let-7g to 1 mM ^{13}C/^{15}N-labeled Lin28$_{119–180}$. The chemical shift differences ($\Delta\delta$) were calculated according to the formula $\Delta\delta = [(\Delta H^{N})^2 + (0.17\Delta N^{H})^2]^{1/2}$.

Figure S2) was performed to establish the secondary structure of TL-let-7g. These results confirm that the free TL-let-7g adopts a hairpin conformation, with dynamic residues in the G-rich bulge (residues 7–11), the adjoining internal loop (residues 16–19 and 32–35) and the hairpin loop (residues 24–27; Figure 5A). For the ribonuclease protection assay, RNase T_1 was selected because it shares the specificity of ZBDs for single-stranded guanines (55). The results of nuclease mapping in the presence of increasing concentration of Lin28$_{119–180}$ (Figure 5) clearly demonstrate that only the G residues within the G-rich bulge (G8, G10 and G11) and adjacent stems (G4 and G12) are protected by Lin28$_{119–180}$. Only one G residue (G18) becomes more accessible in the presence of Lin28$_{119–180}$, most likely resulting from destabilization of the predicted G18–C12 base pair (see Discussion section). Thus, the ribonuclease protection assay indicates that the G-rich bulge is the main binding site for the NCp7-like domain and that the adjoining internal loop is also affected by binding.

Table 2. Dissociation constants (K_d in nM) of mutants of the NCp7-like domain of Lin28 (Lin28$_{119–180}$) for the TL-let-7g RNA

Protein	K_d (nM)
Lin28$_{119–180}$	1.3 ± 0.3
Lin28$_{136–180}$	>5000
Lin28$_{119–140}$	>2500
Lin28$_{119–160}$	>5000
Lin28$_{119–180}$ (KR^{-})	>5000
HIV–1 NCp7	1.1 ± 0.3

Detailed mapping of the pre-let-7g determinants for Lin28$_{119–180}$ binding

Next, we performed an exhaustive EMSA analysis of TL-let-7g mutants to identify the RNA determinants of the high-affinity interaction between TL-let-7g and Lin28$_{119–180}$ (K_d = 1.3 nM; Table 3). Since NCp7 and several other ZBDs specifically recognize single-stranded nucleic acids (33,54), several mutants of the loop regions of TL-let-7g (Figure 1D) were investigated. Replacement of the ACCC hairpin loop by a stable GNRA tetraloop (GCAA) increases the K_d by a factor of 3, and the punctual C25A mutation increases the K_d by a factor of 1.6. Deletion of unpaired nucleotides in the internal loop to create a stable stem (Figure 1D; Δiloop) causes a 5.3-fold increase in the K_d compared with the wild-type

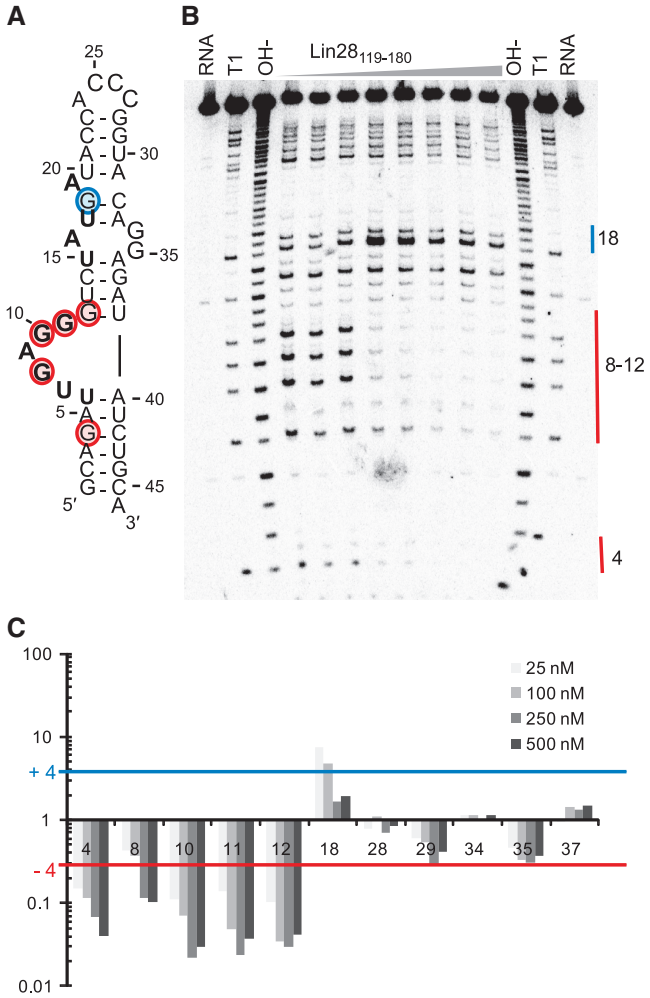

Figure 5. Footprint analysis of TL-let-7g with RNase T_1. (**A**) Secondary structure of TL-let-7g with the mapping of in-line probing and T_1 footprinting data. Residues that are the most susceptible to spontaneous cleavage through in-line attack are in bold (Supplementary Figure S2), and residues that experience a significant reduction ($I_p/I_0 = -4$) or enhancement ($I_p/I_0 = +4$) of T_1 cleavage in the presence of $Lin28_{119–180}$ are shaded in red and blue, respectively. (**B**) Typical RNA footprinting gel of TL-let-7g in the absence and presence of $Lin28_{119–180}$ (at concentrations of 0, 1, 5, 25, 100, 250, 500 and 1000 nM). Lanes with input TL-let-7g (RNA), an alkaline hydrolysis ladder (OH−) and a T_1 hydrolysis ladder (T1) are also included. (**C**) Histogram of normalized band sensitivity (I_p/I_0, where I_p and I_0 are, respectively, the intensity in the presence and absence of protein) for T_1 cleaveage of each guanine obtained at 25–500 nM $Lin28_{119–180}$.

RNA. Similar changes in affinity are also observed with simultaneous mutations of three unpaired nucleotides from the internal loop (A19C/G34A/G35A: 6.2-fold K_d increase) or from the combined effect of two related mutations (1.2- and 3.8-fold K_d increase for the A19C and G34A/G35A mutants, respectively). Thus, the internal loop of TL-let-7g makes a minor contribution to $Lin28_{119–180}$ binding, but the hairpin loop appears to be replaceable.

Several EMSA results indicate that the G-rich bulge contributes significantly to $Lin28_{119–180}$ binding. First, deletion of the G-rich bulge (Figure 1D; Δbulge) has a substantial effect since no specific binding could be observed with this mutant at a protein concentration as high as 5 μM. Similarly, no specific binding could be observed for a mutant in which all guanines at the bulge were mutated (G8C/G10C/G11A/G12A). To identify the guanine residues of the G-rich bulge that are important for $Lin28_{119–180}$ binding, each guanine was individually mutated and the K_d was determined by EMSA (Table 3). Although the G8C and G11A mutations have a negligible effect on binding, the G10C or G12A mutations cause 7- and 19-fold increases in K_d, respectively, which represent the largest changes observed in this study for single nucleotide mutations. Surprisingly, the double G10C/G12A mutation does not completely abolish specific binding, but instead has a similar effect on binding as the single G12A mutation, suggesting that the remaining guanine residues (G8 and G11) may contribute to binding in the double mutant (G10C/G12A). Thus, it appears that G10 and G12 are key residues for the recognition, but that other G residues may also contribute to the affinity of the G-rich bulge for $Lin28_{119–180}$. The G-rich bulge clearly represents the main RNA determinant of the high-affinity interaction with $Lin28_{119–180}$, although the internal loop of TL-let-7g makes a minor contribution to binding.

To provide additional evidence for the importance of the G-rich bulge, we performed NMR studies of $Lin28_{119–180}$ in complex with TL-let-7g mutants (Supplementary Figure S3). As expected, the ^1H-^{15}N HSQC spectrum of the G34AG35A/$Lin28_{119–180}$ complex is almost identical to that of the TL-let-7g/$Lin28_{119–180}$ complex, particularly for residues from the two ZBDs. In contrast, the ^1H-^{15}N HSQC spectrum of the Δbulge/$Lin28_{119–180}$ complex indicates that, at the high concentration (0.1 mM) used for these NMR studies, $Lin28_{119–180}$ interacts with the Δbulge mutant, but in a different manner than observed for the wild-type TL-let-7 g. Thus, these NMR results confirm that the G-rich bulge is the main determinant for high-affinity binding of $Lin28_{119–180}$ to TL-let-7g.

Table 3. Dissociation constants (K_d in nM) of the NCp7-like domain of Lin28 ($Lin28_{119–180}$) for various mutants of the TL-let-7g RNA

TL-let-7g RNA	K_d (nM)	$K_d/[K_d$ (wt)][a]
Wild type (wt)	1.3 ± 0.3	1
GNRA tetraloop	4 ± 1	3.1
C25A	2.1 ± 0.5	1.6
Δiloop	6.9 ± 0.7	5.3
A19C	1.6 ± 0.6	1.2
G34A G35A	5 ± 1	3.8
A19C G34A G35A	8 ± 2	6.2
Δbulge	n.b.[b]	n.b.[b]
G8C G10C G11A G12A	n.b.[b]	n.b.[b]
G8C	1.1 ± 0 1	0.9
G10C	9 ± 2	6.9
G11A	3 ± 1	2.3
G12A	25 ± 7	19
G10C G12A	18 ± 5	14

[a]The $K_d/[K_d$ (wt)] is the ratio of the K_d obtained for the mutant TL-let-7g over the K_d obtained for the wild-type TL-let-7g RNA.
[b]No specific binding observed. The gel mobility shift assays display smearing and multiple shifts, indicating non-specific binding.

Given the importance of the G-rich bulge for $Lin28_{119–180}$ binding, we examined the sequences of all known pre-let-7g RNAs (Supplementary Figure S4). We find that the sequences of the G-rich bulge and adjacent residues are highly conserved in mammals, birds and amphibians and fit the consensus sequence U*AGGGU (* is A, C or G), with the UGAGGGU sequence found in mouse and human being the most common.

DISCUSSION

In this work, we systematically identified the key determinants of the interaction between pre-let-7g and Lin28 using highly-purified proteins and RNAs. A surprising result is the high affinity (K_d of 0.15 nM) measured for Lin28 binding to pre-let-7g, given that a K_d of 1–2 μM has previously been reported for this interaction (20,38). One important factor that may explain the higher affinity measured in this study is the absence of RNA competitor in the binding buffer. We also took great care of removing RNA contaminants during the purification of the Lin28 proteins and found it necessary to use the S7 nuclease. Nevertheless, our results are in general agreement with previous studies, which established that the extended terminal loop of pre-let-7g is the binding site for Lin28 (20,38).

The EMSA data for the binding of full-length Lin28 to pre-let-7g and TL-let-7g could be fitted well by using the Hill equation, but not the classical one-site binding equation (Figure 2C). The Hill coefficients of ~2.8 indicate that these interactions involve a minimum of three binding sites for Lin28 on the target RNA, and most likely reflect positive cooperativity for these binding events (56). The concept that Lin28 can bind multiple sites on the RNA is further supported by supershifts observed at higher protein concentration (>5 nM; Figure 2A). Since the Hill coefficients obtained for TL-let-7g and pre-let-7g are essentially identical, we propose that full-length Lin28 can cooperatively bind a minimum of three sites in the extended terminal loop of pre-let-7g. This ability of Lin28 to cooperatively bind multiple sites is not observed with the isolated NCp7-like domain and could not be identified from our binding data with the isolated CSD due to severe aggregation problems, as previously reported (31). The CSD of Lin28 is similar to bacterial cold shock proteins (30), and other members of this family of proteins are reported to display cooperativity and weak specificity (57). For example, the RNA chaperone CspA is known to destabilize RNA secondary structure by cooperatively binding to single-stranded regions with low sequence specificity (58). Thus, although the CSD of Lin28 does not contribute significantly to the affinity of Lin28 to pre-let-7g, it may mediate cooperative binding in the context of the full-length protein.

The C-terminal domain of Lin28 displays remarkable similarities with the NCp7 protein of HIV-1. The sequence alignment between the NCp7 protein and metazoan Lin28 proteins shows a high degree of similarity, which involves several residues from the KR-rich domain and the two ZBDs of NCp7. Several of these residues have been shown to contribute to RNA binding in NMR structures of NCp7 bound to RNA hairpins derived from the HIV-1 Ψ site (33,54). Here, both NMR and mutational studies confirmed that the KR-rich domain and both ZBDs of Lin28 participate in TL-let-7g binding. In particular, truncation of the KR-rich domain or mutations of K/R residues within the KR-rich domain of $Lin28_{119–180}$ abrogate high-affinity binding to TL-let-7g. Similarly, truncation of the KR-rich domain from HIV-1 NCp7 was previously shown to prevent the specific binding of NCp7 to its Ψ-site RNA target (59).

The similarity between $Lin28_{119–180}$ and the NCp7 protein also extends to their binding affinity and specificity. Indeed, both $Lin28_{119–180}$ and NCp7 display low nanomolar affinities toward their specific RNA targets (60,61) and specifically recognize a G-rich single-stranded region. Furthermore, we find that $Lin28_{119–180}$ and NCp7 bind TL-let-7g with similar affinities under the same conditions. $Lin28_{119–180}$ preferentially binds the G10-X-G12 unit of the G-rich bulge, but may also bind G8-X-X-G11 when G10 and G12 are simultaneously mutated. Our results are compatible with a recent study in which Lin28 was found to strongly protect residues G8, G10, G11 and G12 of the G-rich bulge of pre-let-7g from ribonuclease cleavage (38). The sequence at the G-rich bulge of pre-let-7g is highly conserved in mammals (Supplementary Figure S4), and sequence similarity was also found in this region for most human and mouse let-7 family members (19,20,24). Thus, it is likely that this region is important for Lin28 binding to other pre-let-7 miRNAs. NCp7 from HIV-1 was also reported to bind exposed guanines with its two ZBDs (33,54,60,62,63). Its target sequences in the Ψ-site RNA are generally located in a hairpin loop, where they form a G-X-G motif, but exceptions such as the binding to the 1 × 3 internal loop of SL1 demonstrate flexibility in RNA recognition by NCp7 (60,63). This flexibility in target sequence recognition may be inherent to the adaptive nature of the NCp7 motif and may explain, in part, why Lin28 binds a wide variety of mRNA targets in addition to let-7 precursors (35,36,64,65). $Lin28_{119–180}$ possibly recognizes its RNA target in the same way that HIV-1 NCp7 binds hairpin loops in the HIV Ψ RNA (33,54). In this model, the two ZBDs of Lin28 would each bind an exposed guanine in the G-rich bulge and the KR-rich domain would bind an adjacent stem or enlarged major groove. Interaction of $Lin28_{119–180}$ with a domain adjacent to the G-rich bulge is consistent with our mutational studies, which indicate that the internal loop of TL-let-7g makes a small contribution to binding. Given all these similarities between NCp7 and $Lin28_{119–180}$, it is tempting to speculate on a role for HIV-1 NCp7 in let-7 biogenesis and a common origin for NCp7 and the C-terminal domain of Lin28. The latter is inconsistent with a classical evolutionary model that does not include viruses, but agrees with an alternative model in which viruses play an important role in cellular evolution (66,67).

The high-affinity of Lin28 toward TL-let-7g is mostly due to its C-terminal NCp7-like domain, which has been shown to be required for several functional aspects of

Lin28. It was shown to be required for Lin28 processing inhibition of pre-let-7g *in vivo* (20), its localization to P-bodies (31), its specific binding to let-7 precursors (20,24) and for the Lin28-mediated uridylylation of pre-let-7-a-1 by TUT4 (22). However, the NCp7-like domain is not sufficient for Lin28 function, as previously demonstrated with two Lin28 homologs, Lin28B and Lin28B-S, which are overexpressed in human hepatocellular carcinoma and in several cancer lines (68). The Lin28B-S preserves the NCp7-like domain but contains a truncation of the cold-shock domain. It has been shown that the overexpression of Lin28B-S does not induce cancer cell proliferation in contrast to what is observed with Lin28B (68). In addition, Lin28B-S does not inhibit the processing of pri-let-7g like Lin28 and Lin28B (21). Thus, although the NCp7-like domain of Lin28 likely contributes to its *in vivo* function through high-affinity and specific binding to the terminal loop of pre-let-7g, the cold-shock domain is required for Lin28 to function as an effective oncogene and inhibitor of let-7 biogenesis.

Given its high-affinity for pre-let-7g, the NCp7-like domain may be responsible for the initial targeting of pre-let-7g. After this initial binding event, partial unfolding of the terminal loop of pre-let-7g would make it more accessible for binding multiple copies of Lin28. Both the NCp7 domain and the CSD have been previously described as RNA chaperones (58,69) and could contribute to making the terminal loop more accessible. In agreement with this role of the NCp7-like domain in Lin28, its binding to TL-let-7g makes G18 of the internal loop more accessible to ribonuclease cleavage. The high-affinity and specificity of the NCp7-like domain for the G-rich bulge may also allow Lin28 to bind its RNA target in an orderly fashion to insure that important functional regions of the RNA are protected from binding of miRNA processing enzymes. The G-rich bulge is directly adjacent to one of the Dicer processing site. Thus, it is likely that Lin28 binding at the G-rich bulge protects the pre-let-7g RNA from Dicer cleavage at this site, likely by both steric hindrance and destabilization of the stem region near the G-rich bulge (38).

Our mutagenesis, T1 footprinting and NMR data all indicate that the G-rich bulge of TL-let-7g is the main determinant for high-affinity binding to $Lin28_{119–180}$. In an apparent contradiction with our results, a previous report identified a different G-rich region of pre-let-7 (GGAG residues 34–37 of the internal loop in Figure 5A) to be important for Lin28 binding (24). Both these G-rich regions are highly conserved in mammalian let-7g, and similar G-rich regions are found both at the 5′-end and 3′-end of the terminal loop in most members of the human and mouse let-7 family (19,20,24). Thus, it is likely that both G-rich regions in pre-let-7 miRNAs are important for binding full-length Lin28. For example, initial binding of Lin28 to the G-rich bulge may expose internal loop residues and allow Lin28 binding to the other G-rich region. Alternatively, the levels of Lin28 and other cellular factors may affect binding of Lin28 to the two G-rich regions. Interestingly, both hnRNP A1 and KSRP specifically bind G-rich sequences within the terminal loop of pre-let-7a1 (17,18) and may regulate Lin28 binding at these sites in some pre-let-7 members. Clearly, further studies are needed to more precisely determine how each G-rich region contributes with cellular factors to regulate pre-let-7 biogenesis.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures S1–4, Supplementary References [71–73].

ACKNOWLEDGEMENTS

The authors thank Dominique Chaussé, Geneviève Di Tomasso and Alexis Rompré-Brodeur for sample preparation.

FUNDING

The National Sciences and Engineering Research Council of Canada (NSERC) (to P.L. and J.G.O.). A.D. holds a Ph.D. scholarship from the Canadian Institutes for Health Research. J.B. holds an MSc scholarship from the Université de Montréal. P.L. holds a Canada Research Chair in Structural Biology and Engineering of RNA. Funding for open access charge: National Sciences and Engineering Research Council of Canada.

Conflict of interest statement. None declared.

REFERENCES

1. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
2. Winter,J., Jung,S., Keller,S., Gregory,R.I. and Diederichs,S. (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell. Biol.*, **11**, 228–234.
3. Filipowicz,W., Bhattacharyya,S.N. and Sonenberg,N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.*, **9**, 102–114.
4. Kim,V.N., Han,J. and Siomi,M.C. (2009) Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.*, **10**, 126–139.
5. Zhao,Y., Ransom,J.F., Li,A., Vedantham,V., von Drehle,M., Muth,A.N., Tsuchihashi,T., McManus,M.T., Schwartz,R.J. and Srivastava,D. (2007) Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell*, **129**, 303–317.
6. Peter,M.E. (2009) Let-7 and miR-200 microRNAs: guardians against pluripotency and cancer progression. *Cell Cycle*, **8**, 843–852.
7. Ambros,V. (2011) MicroRNAs and developmental timing. *Curr. Opin. Genet. Dev.*, **21**, 511–517.
8. Esquela-Kerscher,A. and Slack,F.J. (2006) Oncomirs—microRNAs with a role in cancer. *Nat. Rev. Cancer*, **6**, 259–269.
9. Shi,X.B., Tepper,C.G. and deVere White,R.W. (2008) Cancerous miRNAs and their regulation. *Cell Cycle*, **7**, 1529–1538.
10. Roush,S. and Slack,F.J. (2008) The let-7 family of microRNAs. *Trends Cell Biol.*, **18**, 505–516.
11. Bussing,I., Slack,F.J. and Grosshans,H. (2008) Let-7 microRNAs in development, stem cells and cancer. *Trends Mol. Med.*, **14**, 400–409.
12. Visone,R. and Croce,C.M. (2009) MiRNAs and cancer. *Am. J. Pathol.*, **174**, 1131–1138.
13. O'Day,E. and Lal,A. (2010) MicroRNAs and their target gene networks in breast cancer. *Breast Cancer Res.*, **12**, 201.

14. Thomson,J.M., Newman,M., Parker,J.S., Morin-Kensicki,E.M., Wright,T. and Hammond,S.M. (2006) Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev.*, **20**, 2202–2207.
15. Wulczyn,F.G., Smirnova,L., Rybak,A., Brandt,C., Kwidzinski,E., Ninnemann,O., Strehle,M., Seiler,A., Schumacher,S. and Nitsch,R. (2007) Post-transcriptional regulation of the let-7 microRNA during neural cell specification. *FASEB J.*, **21**, 415–426.
16. Van Wynsberghe,P.M., Kai,Z.S., Massirer,K.B., Burton,V.H., Yeo,G.W. and Pasquinelli,A.E. (2011) LIN-28 co-transcriptionally binds primary let-7 to regulate miRNA maturation in Caenorhabditis elegans. *Nat. Struct. Mol. Biol.*, **18**, 302–308.
17. Michlewski,G. and Caceres,J.F. (2010) Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nat. Struct. Mol. Biol.*, **17**, 1011–1018.
18. Trabucchi,M., Briata,P., Garcia-Mayoral,M., Haase,A.D., Filipowicz,W., Ramos,A., Gherzi,R. and Rosenfeld,M.G. (2009) The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature*, **459**, 1010–1014.
19. 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.
20. 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.
21. Viswanathan,S.R., Daley,G.Q. and Gregory,R.I. (2008) Selective blockade of microRNA processing by Lin28. *Science*, **320**, 97–100.
22. 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.
23. Rybak,A., Fuchs,H., Smirnova,L., Brandt,C., Pohl,E.E., Nitsch,R. and Wulczyn,F.G. (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.
24. 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.
25. 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.
26. Lehrbach,N.J., Armisen,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.
27. Newman,M.A. and Hammond,S.M. (2010) Lin-28: an early embryonic sentinel that blocks Let-7 biogenesis. *Int. J. Biochem. Cell Biol.*, **42**, 1330–1333.
28. 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.* (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science*, **318**, 1917–1920.
29. Viswanathan,S.R. and Daley,G.Q. (2010) Lin28: a microRNA regulator with a macro role. *Cell*, **140**, 445–449.
30. 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.
31. Balzer,E. and Moss,E.G. (2007) Localization of the developmental timing regulator Lin28 to mRNP complexes, P-bodies and stress granules. *RNA Biol.*, **4**, 16–25.
32. Landsman,D. (1992) RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain. *Nucleic Acids Res.*, **20**, 2861–2864.
33. 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.
34. Polesskaya,A., Cuvellier,S., Naguibneva,I., Duquet,A., Moss,E.G. and Harel-Bellan,A. (2007) Lin-28 binds IGF-2 mRNA and participates in skeletal myogenesis by increasing translation efficiency. *Genes Dev.*, **21**, 1125–1138.
35. Xu,B. and Huang,Y. (2009) Histone H2a mRNA interacts with Lin28 and contains a Lin28-dependent posttranscriptional regulatory element. *Nucleic Acids Res.*, **37**, 4256–4263.
36. Xu,B., Zhang,K. and Huang,Y. (2009) Lin28 modulates cell growth and associates with a subset of cell cycle regulator mRNAs in mouse embryonic stem cells. *RNA*, **15**, 357–361.
37. Qiu,C., Ma,Y., Wang,J., Peng,S. and Huang,Y. (2010) Lin28-mediated post-transcriptional regulation of Oct4 expression in human embryonic stem cells. *Nucleic Acids Res.*, **38**, 1240–1248.
38. Lightfoot,H.L., Bugaut,A., Armisen,J., Lehrbach,N.J., Miska,E.A. and Balasubramanian,S. (2011) A LIN28-dependent structural change in pre-let-7g directly inhibits dicer processing. *Biochemistry*, **50**, 7514–7521.
39. Di Tomasso,G., Lampron,P., Dagenais,P., Omichinski,J.G. and Legault,P. (2011) The ARiBo tag: a reliable tool for affinity purification of RNAs under native conditions. *Nucleic Acids Res.*, **39**, e18.
40. Di Tomasso,G., Dagenais,P., Desjardins,A., Rompré-Brodeur,A., Delfosse,V. and Legault,P. (2012) Affinity purification of RNA using an ARiBo tag. *Methods Mol. Biol.*, In Press.
41. Delfosse,V., Bouchard,P., Bonneau,E., Dagenais,P., Lemay,J.F., Lafontaine,D.A. and Legault,P. (2010) Riboswitch structure: an internal residue mimicking the purine ligand. *Nucleic Acids Res.*, **38**, 2057–2068.
42. Bouchard,P., Lacroix-Labonté,J., Desjardins,G., Lampron,P., Lisi,V., Lemieux,S., Major,F. and Legault,P. (2008) Role of SLV in SLI substrate recognition by the Neurospora VS ribozyme. *RNA*, **14**, 736–748.
43. Jenkins,L.M., Byrd,J.C., Hara,T., Srivastava,P., Mazur,S.J., Stahl,S.J., Inman,J.K., Appella,E., Omichinski,J.G. and Legault,P. (2005) Studies on the mechanism of inactivation of the HIV-1 nucleocapsid protein NCp7 with 2-mercaptobenzamide thioesters. *J. Med. Chem.*, **48**, 2847–2858.
44. Omichinski,J.G., Clore,G.M., Schaad,O., Felsenfeld,G., Trainor,C., Appella,E., Stahl,S.J. and Gronenborn,A.M. (1993) NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science*, **261**, 438–446.
45. Regulski,E.E. and Breaker,R.R. (2008) In-line probing analysis of riboswitches. *Methods Mol. Biol.*, **419**, 53–67.
46. Kay,L., Keifer,P. and Saarinen,T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.*, **114**, 10663–10665.
47. Wittekind,M. and Mueller,L. (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J. Magn. Reson. Series B*, **101**, 201–205.
48. Grzesiek,S., Dobeli,H., Gentz,R., Garotta,G., Labhardt,A.M. and Bax,A. (1992) ^{1}H, ^{13}C, and ^{15}N NMR backbone assignments and secondary structure of human interferon-gamma. *Biochemistry*, **31**, 8180–8190.
49. Muhandiram,D.R. and Kay,L.E. (1994) Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. Series B*, **103**, 203–216.
50. Wishart,D.S., Bigam,C.G., Yao,J., Dyson,H.J., Oldfield,E., Markley,J.L. and Sykes,B.D. (1995) ^{1}H, ^{13}C, ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR*, **6**, 135–140.
51. 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.
52. Johnson,B.A. and Blevins,R.A. (1994) NMR view: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.
53. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
54. 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.

55. Steyaert,J. (1997) A decade of protein engineering on ribonuclease T1-atomic dissection of the enzyme-substrate interactions. *Eur. J. Biochem.*, **247**, 1–11.
56. Weiss,J.N. (1997) The Hill equation revisited: uses and misuses. *FASEB J.*, **11**, 835–841.
57. Ermolenko,D.N. and Makhatadze,G.I. (2002) Bacterial cold-shock proteins. *Cell Mol. Life Sci.*, **59**, 1902–1913.
58. Jiang,W., Hou,Y. and Inouye,M. (1997) CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. *J. Biol. Chem.*, **272**, 196–202.
59. Dannull,J., Surovoy,A., Jung,G. and Moelling,K. (1994) Specific binding of HIV-1 nucleocapsid protein to PSI RNA in vitro requires N-terminal zinc finger and flanking basic amino acid residues. *EMBO J.*, **13**, 1525–1533.
60. Shubsda,M.F., Paoletti,A.C., Hudson,B.S. and Borer,P.N. (2002) Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein. *Biochemistry*, **41**, 5276–5282.
61. Paoletti,A.C., Shubsda,M.F., Hudson,B.S. and Borer,P.N. (2002) Affinities of the nucleocapsid protein for variants of SL3 RNA in HIV-1. *Biochemistry*, **41**, 15423–15428.
62. Fisher,R.J., Rein,A., Fivash,M., Urbaneja,M.A., Casas-Finet,J.R., Medaglia,M. and Henderson,L.E. (1998) Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. *J. Virol.*, **72**, 1902–1909.
63. Yuan,Y., Kerwood,D.J., Paoletti,A.C., Shubsda,M.F. and Borer,P.N. (2003) Stem of SL1 RNA in HIV-1: structure and nucleocapsid protein binding for a 1×3 internal loop. *Biochemistry*, **42**, 5259–5269.
64. 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.
65. 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.
66. Claverie,J.M. (2006) Viruses take center stage in cellular evolution. *Genome Biol.*, **7**, 110.
67. Bandea,C.I. (2009) The origin and evolution of viruses as molecular organisms. *Nat. Precedings*, npre.2009.3886.2001.
68. 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.
69. Tsuchihashi,Z., Khosla,M. and Herschlag,D. (1993) Protein enhancement of hammerhead ribozyme catalysis. *Science*, **262**, 99–102.
70. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
71. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
72. Parisien,M. and Major,F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.
73. Notredame,C., Higgins,D.G. and Heringa,J. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302**, 205–217.