

Université de Montréal

**Caractérisation de l'interaction entre la protéine Lin28 et
le précurseur du microARN let-7g**

par

Alexandre Desjardins

Département de Biochimie et de Médecine Moléculaire
Faculté de Médecine

Thèse présentée à la Faculté de Médecine
en vue de l'obtention du grade de Philosophiae Doctor (Ph.D)
en Biochimie

Août 2013

© Alexandre Desjardins, 2013

Université de Montréal
Faculté des études supérieures de postdoctorales

Cette thèse est intitulée :

Caractérisation de l'interaction entre la protéine Lin28 et le précurseur du microARN let-7g

Présentée par :
Alexandre Desjardins

a été évaluée par un jury composé des personnes suivantes :

Gerardo Ferbeyre, président-rapporteur
Pascale Legault, directeur de recherche
James G. Omichinski, co-directeur
Lea Harrington, membre du jury
Patrick Provost, examinateur externe
Louis Gaboury, représentant du doyen de la FES

Résumé

La régulation de l'expression des gènes est ce qui permet à nos cellules de s'adapter à leur environnement, de combattre les infections ou, plus généralement, de produire la quantité exacte de protéine nécessaire pour répondre à un besoin spécifique. Parmi les joueurs les plus importants dans cette régulation de l'expression des gènes on retrouve les microARN (miARN). Ces petits ARN de 22 nucléotides sont présents chez la majorité des espèces multicellulaires et sont responsables du contrôle direct de plus de 30% des gènes exprimant des protéines chez les vertébrés. La famille de miARN *lethal-7* (*let-7*) est composée de miARN parmi les plus connus et ayant des fonctions cruciales pour la cellule. La régulation du niveau des miARN *let-7* est essentielle au bon développement cellulaire. La biogenèse de ces miARN, du transcrit primaire jusqu'à leur forme mature, est régulée principalement par Lin28, une protéine pluripotente très conservée. Cette protéine est composée d'un domaine *cold shock* (CSD) et de deux domaines de liaison au zinc. C'est grâce à ces domaines de liaison à l'ARN que Lin28 peut lier et inhiber la maturation des miARN *let-7*.

L'objectif de cette thèse est de caractériser l'interaction entre Lin28 et le microARN précurseur *let-7g* afin de mieux comprendre le rôle de cette protéine dans l'inhibition de la biogenèse du miARN. À l'aide de techniques biochimiques et biophysiques, nous avons d'abord défini les principaux déterminants de l'interaction entre Lin28 et la boucle terminale du miARN précurseur *let-7g* (TL-*let-7g*). Nous avons conclu que le domaine C-terminal de Lin28, composé d'un motif riche en lysines et arginines ainsi que de deux motifs de liaison au zinc, permet à la protéine de lier spécifiquement et avec haute affinité un renflement riche en guanine conservé chez les précurseurs de la famille *let-7*. Aussi, parce que la séquence et la spécificité de liaison à l'ARN de ce domaine C-terminal sont semblables à celles de la protéine NCp7 du VIH, nous avons défini ce dernier comme le domaine *NCp7-like* de Lin28. Par la suite, nous avons caractérisé la multimérisation de trois protéines Lin28 sur la boucle terminale de pre-*let-7g*. Ceci a permis de réconcilier d'apparentes contradictions retrouvées dans la littérature actuelle concernant les sites de liaison de Lin28 lors de sa liaison aux miARN précurseurs. Nous avons identifié trois sites de liaison à haute affinité sur TL-*let-7g* qui sont liés dans un ordre précis par trois protéines Lin28. Lors de la formation du complexe

multimérique, le CSD permet une déstabilisation de l'ARN, ce qui rend accessible plusieurs sites de liaison. Le domaine *NCp7-like* permet plutôt un assemblage ordonné de la protéine et facilite la liaison initiale de cette dernière. Ces nouveaux résultats rendent possible la mise au point d'un nouveau modèle de l'interaction entre Lin28 et le miARN précurseur let-7g. En conclusion, les études réalisées dans cette thèse apportent une meilleure compréhension des mécanismes moléculaires impliqués dans la régulation post-transcriptionnelle d'une importante famille de miARN et permettront de guider les futures études dans le domaine de recherche en pleine effervescence qu'est celui de la biogenèse des miARN.

Mots-clés : Lin28, let-7, microARN, régulation post-transcriptionnelle, interaction protéine-ARN, formation de complexes macromoléculaires, gels natifs.

Abstract

The regulation of gene expression is what allows our cells to adapt to their environment, to fight infections or, more generally, to express the appropriate level of proteins to meet a specific need. The microRNAs (miRNAs) are among the most important players in the regulation of gene expression. These small RNAs of 22 nucleotides are present in most multicellular species and are responsible for the direct control of more than 30% of protein-expressing genes in vertebrates. The miRNA *lethal-7* (*let-7*) family consist of some of the most studied miRNAs and plays crucial roles in the cell. The appropriate regulation of the *let-7* miRNAs level is essential for proper cellular development. The biogenesis of these miRNAs, from the primary transcript to their mature form is mainly regulated by Lin28, a highly-conserved pluripotent protein. This protein is composed of a cold shock domain (CSD) and two zinc-binding domains. These RNA-binding domains allow Lin28 to bind and inhibit the maturation of the *let-7* miRNA.

The objective of this thesis is to characterize the interaction between the Lin28 protein and the *let-7g* miRNA precursor to better understand the role of this protein in the inhibition of miARN biogenesis. Using biochemical and biophysical techniques, we first identified the main determinants of the interaction between Lin28 and the terminal loop of the precursor miRNA *let-7g* (TL-*let-7g*). We concluded that the C-terminal domain of Lin28, composed of a lysine-rich and arginine-rich motif in addition to two zinc-binding motifs, is sufficient to bind with high affinity a conserved guanine-rich bulge located on the TL-*let-7g*. In addition, because the sequence and RNA-binding specificity of this C-terminal domain are similar to those of the HIV protein NCp7, we defined this region as the NCp7-like domain of Lin28. Subsequently, we characterized the multimerization of three Lin28 proteins on the terminal loop of pre-*let-7g*. This study helped to reconcile apparent contradictions found in the current literature regarding the Lin28-binding sites on miRNA precursors. We identified three high-affinity binding sites on TL-*let-7g* that are bound in a stepwise manner by the three Lin28 proteins. As part of the formation of the multimeric complex, both RNA-binding domains of Lin28 play an important role. The CSD destabilizes the RNA and this exposes several binding sites, whereas the NCp7-like domain allows an orderly protein assembly and facilitates the

initial binding of the protein. These results lead us to propose a new model for the interaction between Lin28 and pre-let-7g. In conclusion, these studies provide a better understanding of the molecular mechanisms involved in the post-transcriptional regulation of an important family of miRNAs and will help guide future projects in the expanding research area of miRNA biogenesis.

Keywords: Lin28, let-7, microRNA, post-transcriptional regulation, protein-RNA interaction, macromolecular complex formation, native gels.

Table des matières

Résumé.....	v
Abstract.....	vii
Table des matières.....	ix
Liste des tableaux.....	xii
Liste des figures.....	xiii
Liste des abréviations.....	xvii
Remerciements.....	xx
1.0 Introduction.....	1
1.1 Biogenèse des microARN.....	2
1.1.1 Historique.....	2
1.1.2 Transcription des microARN.....	4
1.1.3 Clivage du pri-miARN par Drosha et exportation au cytoplasme.....	7
1.1.4 Clivage par Dicer du pre-miARN.....	11
1.1.5 Protéines interagissant avec Dicer.....	13
1.1.6 Chargement du miARN sur RISC.....	13
1.1.7 Reconnaissance de l'ARNm cible.....	14
1.1.8 Actions du RISC sur les ARNm.....	17
1.1.9 Voies de biogenèse alternatives.....	20
1.1.10 Comparaison avec la biogenèse des miARN chez les plantes.....	22
1.2 Régulation post-transcriptionnelle des microARN.....	24
1.2.1 Édition des miARN par ADAR.....	24
1.2.2 Mono-uridylation des miRNA du groupe II de la famille let-7.....	26
1.2.3 Régulation par SF2/ASF.....	28
1.2.4 Régulation par KSRP.....	29
1.2.5 Régulation par hnRNP A1.....	30
1.2.6 Introduction au complexe Lin28 et let-7.....	33
1.3 Importance des microARN.....	35
1.3.1 Rôle des microARN dans la régulation temporelle.....	35
1.3.2 Les microARN et le cancer.....	37

1.4	Famille let-7 de microARN	40
1.4.1	Rôle physiologique	41
1.4.2	Suppresseur de tumeur	42
1.5	Protéine Lin28.....	44
1.5.1	Rôles des différents domaines de Lin28	46
1.5.2	Liaison aux ARNm	47
1.6	Inhibition de la biogenèse de let-7 par Lin28	50
1.6.1	Modèles de régulation de pri-let-7 par Lin28	50
1.6.2	Cartographie de l'interaction entre pre-let-7 et Lin28	54
1.6.3	Études structurales de Lin28.....	58
1.7	Objectifs de la thèse	60
1.8	Méthodologie utilisée.....	62
1.8.1	Purification des macromolécules	62
1.8.2	Détermination de l'affinité d'un complexe.....	62
1.8.3	Détermination des sites de liaison.....	63
1.8.4	Visualisation de l'activité chaperonne à ARN.....	64
2.0	Interaction du domaine <i>NCp7-like</i> de Lin28 avec TL-let-7g.....	65
	Mise en contexte	66
	Abstract.....	68
	Introduction.....	69
	Materials and methods	72
	Results.....	76
	Discussion.....	82
	Acknowledgements.....	86
	Tables.....	87
	Figures.....	89
	References.....	95
	Supplementary material	102
	Supplementary references.....	105
3.0	Assemblage séquentiel de plusieurs protéines Lin28 sur TL-let-7g.....	107
	Mise en contexte	108

Abstract.....	110
Introduction.....	111
Materials and methods	114
Results.....	119
Discussion.....	129
Acknowledgements.....	135
Tables.....	136
Figures.....	138
References.....	146
Supplementary material	153
Supplementary references.....	167
4.0 Discussion et conclusion.....	169
4.1 Importance physiologique du modèle proposé	170
4.1.1 Le modèle proposé et la littérature actuelle	171
4.1.2 Effet de l'assemblage séquentiel sur Dicer	173
4.1.3 Possibles rôles physiologiques dans la maturation des miARN	174
4.2 Liaison coopérative de Lin28.....	176
4.2.1 La coopérativité chez les macromolécules	176
4.2.2 Coopérativité locale à l'intérieur d'une protéine	178
4.2.3 La coopérativité et les interactions entre protéines et acides nucléiques.....	179
4.3 Perspectives.....	182
4.3.1 Interactions de Lin28 avec Tut4, Drosha et KSRP	182
4.3.2 Études structurales	182
4.3.3 Séquence consensus	183
4.3.4 Les autres membres de la famille let-7	185
4.3.5 Autres rôles potentiels pour l'assemblage séquentiel de Lin28.....	185
4.4 Conclusion générale.....	187
Bibliographie.....	189

Liste des tableaux

Chapitre 2.0

Table 2-I Dissociation constants (K_d in nM) of different domains of Lin28 for various pre-let-7g constructs.....	87
Table 2-II. Dissociation constants (K_d in nM) of mutants of the NCp7-like domain of Lin28 (Lin28 ₁₁₉₋₁₈₀) for the TL-let-7g RNA.	87
Table 2-III. 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.	88

Chapitre 3.0

Table 3-I. Dissociation constants (K_d) ^a for binding of Lin28 to RNAs derived from TL-let-7g.....	136
Table 3-II. Ability of Lin28 and variants to form multimeric complexes with various RNAs.	137

Liste des figures

Chapitre 1.0

Figure 1-1. Modèle de la régulation successive des gènes hétérochroniques par les miARN <i>lin-4</i> et <i>let-7</i> chez <i>C. elegans</i>	4
Figure 1-2. Organisation des structures possibles des unités de transcription (UT) contenant des loci des miARN.....	6
Figure 1-3. Organisation et rôles des différents domaines de Droscha.	8
Figure 1-4. Voie de biogenèse canonique des microARN.....	10
Figure 1-5. Structures primaire et tertiaire de la protéine Dicer.	12
Figure 1-6. Les différents types de séquences <i>seed</i> des microARN et leur site d'appariement sur l'ARNm.	15
Figure 1-7. Rôle de la protéine GW182 dans l'inhibition de l'expression génique par les microARN.	18
Figure 1-8. Comparaison des voies de biogenèses canonique et alternatives des miARN chez les animaux.	21
Figure 1-9. Représentation des structures adoptées par les pri-miARN des groupes I et II de la famille <i>let-7</i>	27
Figure 1-10. Interaction de KSRP avec la boucle terminale de <i>let-7a-1</i>	29
Figure 1-11. Mécanisme démontrant l'importance de la boucle terminale de <i>let-7a-1</i> dans sa régulation par KSRP et hnRNP A1.....	31
Figure 1-12. Rôles des miARN dans le développement de <i>C.elegans</i>	36
Figure 1-13. Alignement de séquences des pre-miARN <i>let-7</i> retrouvés chez l'humain.	41
Figure 1-14. Conservation des protéines Lin28A et Lin28B chez les vertébrés.....	45
Figure 1-15. Régulation de la biogenèse du miARN <i>let-7</i> par Lin28.....	52
Figure 1-16. Sites de liaison de Lin28 présents sur TL- <i>let-7a-1</i> et TL- <i>let-7g</i>	55

Chapitre 2.0

Figure 2-1. The Lin28 protein, pre-let-7g RNA and related sequences used in this study.....	89
Figure 2-2. Electrophoretic mobility shift assay (EMSA) of TL-let-7g with Lin28 ₁₋₂₀₉ and Lin28 ₁₁₉₋₁₈₀	90
Figure 2-3. Sequence similarity between the HIV-1 NCp7 and the C-terminal domain of Lin28.....	91
Figure 2-4. Chemical shift mapping of Lin28 ₁₁₉₋₁₈₀ binding the TL-let-7g RNA.....	92
Figure 2-5. Footprint analysis of TL-let-7g with RNase T ₁	93
Supplementary Figure 2-1. Superposition of 2D ¹ H- ¹⁵ N HSQC spectra of 1.3 mM ¹⁵ N-labeled Lin28 ₁₁₉₋₁₈₀ (black) and 1 mM ¹⁵ N-labeled Lin28 ₁₃₆₋₁₈₀ (red).	102
Supplementary Figure 2-2. In-line probing assay of TL-let-7g.....	103
Supplementary Figure 2-3. Lin28 ₁₁₉₋₁₈₀ binds to G34AG35A TL-let-7g and to Δbulge TL-let-7g.....	104
Supplementary Figure 2-4. Sequence alignment of all known pre-let-7g RNA.....	105

Chapitre 3.0

Figure 3-1. The Lin28 protein, TL-let-7g RNA and related sequences used in this study.....	138
Figure 3-2. Stoichiometric binding assay by native gel electrophoresis for Lin28 binding to TL-let-7g RNA	139
Figure 3-3. Stoichiometric binding assay by native gel electrophoresis for Lin28 binding to other pre-let-7 terminal loops.	140
Figure 3-4. Effect of Lin28 on the fluorescence intensity of TL-let-7g containing individual 2-AP modifications.....	141
Figure 3-5. Strand displacement assay monitored by FRET to test the RNA melting activity of Lin28.....	142
Figure 3-6. Dicer processing assay of pre-let-7g-U.....	144

Figure 3-7. Schematic representation of the proposed model for the stepwise assembly of Lin28 on the terminal loop of pre-let-7g.	145
Supplementary Figure 3-1. Binding studies of TL-let-7g to Lin28 by EMSA for determination of the dissociation constant.....	153
Supplementary Figure 3-2. Stoichiometric binding assay by native gel electrophoresis for Lin28 binding to different concentrations of TL-let-7g.....	154
Supplementary Figure 3-3. Competitive stoichiometric binding assay by native gel electrophoresis	155
Supplementary Figure 3-4. Stoichiometric binding assay by native gel electrophoresis for Lin28 binding to TL-miR-21a.	156
Supplementary Figure 3-5. Effect of addition of Lin28 on the emission spectra of 2-AP-modified TL-let-7g RNAs.	157
Supplementary Figure 3-6. Effect of addition of Lin28 on the emission spectra of Cy3/Cy5-labeled RNA duplexes.....	159
Supplementary Figure 3-7. Strand displacement assay monitored by native gel electrophoresis with 25 nM of Cy3/Cy5-labeled duplex RNAs.	161
Supplementary Figure 3-8. Strand displacement assay monitored by native gel electrophoresis with 250 nM of Cy3/Cy5-labeled RNA.....	163
Supplementary Figure 3-9. Dicer processing assay of pre-let-7d-U.....	165
Supplementary Figure 3-10. Dicer processing assay of pre-let-7a-1-U.....	166

Chapitre 4.0

Figure 4-1. Mécanismes expliquant la coopérativité chez les macromolécules	177
Figure 4-2. La coopérativité dans l'assemblage de l'unité 30S du ribosome.	180
Figure 4-3. ARN consensus minimal et structuré.....	184

Liste des abréviations

2-AP	2-aminopurine
ac-pre-miARN	pre-microARN clivé par Ago
ADAR	<i>Adenosine deaminases acting on RNA</i> ou Désaminase à adénosine agissant sur l'ARN
Ago	Argonaute
ARNm	ARN messenger
ARNr	ARN ribosomal
ARNt	ARN de transfert
Cdk2	Cyclin-dependent Kinase 2
Cdk4	Cyclin-dependent Kinase 4
Cdk6	Cyclin-dependent Kinase 6
CDC25A	Cell Division Cycle 25 Homolog A
cellules ES	Cellules souches embryonnaires
CSD	Domaine <i>cold shock</i>
Csps	Protéines <i>cold shock</i>
DCL1	Dicer-like 1
DGCR8	DiGeorge Syndrome Critical Region Gene 8
dsRBD	Domaine de liaison à l'ARN double brin
FRET	<i>Fluorescence resonance energy transfer</i> ou Transfert d'énergie par résonance de fluorescence
hsa	<i>Homo sapiens</i>
ITC	<i>Isothermal titration calorimetry</i> ou Calorimétrie par titrage isotherme
kb	Kilobase
KSRP	<i>KH-type splicing regulatory protein</i> ou Protéine régulatrice de l'épissage de type KH
LRE	<i>Lin28-responsive element</i> ou Élément sensible à Lin28
miARN	microARN

NLS	<i>Nuclear localization signal</i> ou Signal de localisation au noyau
nt	Nucléotide
Ntr	Nucléotidyl Transférase
PABP	<i>Poly(A) binding protein</i> ou Protéine liant le Poly(A)
PAP	<i>Polymerase associated domain</i> ou Domaine associé aux polymérases
pb	Paire de bases
pol II	Polymérase II
pol III	Polymérase III
pre-miARN	microARN précurseur
pri-miARN	Transcrit primaire de microARN
Queue poly-U	Queue d'environ 14 uridines
RE	Réticulum endoplasmique
RIIID	Domaine RNase III
RISC	<i>RNA-induced silencing complex</i> ou Complexe inhibant l'expression par l'ARN
RHA	RNA Helicase A
RRM	Motif de reconnaissance de l'ARN
SPR	<i>Surface Plasmon Resonance</i> ou Résonance plasmonique de surface
TIC	Cellules initiatrices du cancer
TL	Boucle terminale
UT	Unité de transcription
TUT	Tutase
UTR	<i>Untranslated region</i> ou Région non traduite
Xtr	<i>Xenopus tropicalis</i>
ZBD	<i>Zinc binding domain</i> ou Domaine de liaison au zinc

*À ma famille pour qui l'éducation
a toujours été une priorité*

Remerciements

Je tiens d'abord à remercier mon directeur de recherche, Dr Pascale Legault, pour l'excellent encadrement dont j'ai bénéficié au cours des six dernières années. Je la remercie de m'avoir accueilli dans son laboratoire, d'abord comme stagiaire, puis en tant qu'étudiant gradué. J'ai appris beaucoup pendant cette période et cela est sans doute dû à la confiance qui m'a été accordée. Je veux également remercier mon co-directeur, Dr James G. Omichinski, qui a toujours été disponible pour m'assister dans mes différents projets.

Je dois aussi remercier tous les membres des laboratoires Legault-Omichinski, actuels et passés, avec lesquels j'ai travaillé depuis 2006. Tout d'abord Philippe Lampron qui m'a enseigné comment travailler en laboratoire, tout en étant un exemple pour tout démonstrateur. Ensuite, Patricia (la matriarche), Julie et Geneviève Desjardins qui ont constitué la première génération dans le laboratoire. Plusieurs autres se sont ensuite ajoutés au groupe. Un merci spécial à Éric, Jonathan, Nicolas, Geneviève DT et Alix. Vous êtes de bons scientifiques, mais surtout de très bons amis qui font en sorte qu'il est toujours plaisant de se rendre au labo, même quand la science ne coopère pas. Un remerciement particulier à Jonathan avec qui j'ai travaillé sur ce projet depuis son arrivé. Il a permis d'accélérer le développement de ce nouveau sujet dans le laboratoire, nous avons formé une très bonne équipe! Bref, merci à tous ces compagnons de voyage, de ski, de camping, d'escalade, de randonnée, de barbecue, de vin et fromage, de DDC, etc...

Pour finir j'aimerais remercier toute ma famille qui m'a toujours encouragé à poursuivre les études et qui n'ont jamais douté de ma réussite. Merci à Françoise pour la correction de toute cette thèse, je sais que cela n'a pas dû être facile. Finalement, un gros merci à Dominique pour m'avoir accompagné, encouragé et cru en moi tout au long de mes études graduées. Je la remercie également d'avoir lu toute cette thèse pour repérer les fautes d'orthographe, tout en ayant un œil critique pour la syntaxe et la reformulation d'idée. De façon plus générale, je remercie tous ceux qui m'ont côtoyé et supporté au cours de ces six dernières et excellentes années.

1.0 Introduction

1.1 Biogenèse des microARN

Les microARN (miARN) sont une nouvelle classe d'ARN d'environ 22 nucléotides (nt). La principale caractéristique de ces miARN est leur capacité d'inhiber l'expression d'un gène cible par la formation de paires de bases avec l'ARN messager (ARNm) de ce dernier. La régulation de l'expression génique par ces miARN est cruciale pour plusieurs voies métaboliques, mais cette thèse porte principalement sur la fonction des miARN dans la régulation temporelle du développement, ainsi que leurs rôles dans certains cas de cancer. Plus spécifiquement, ce travail s'intéresse à la biogenèse des miARN qui s'effectue en une succession d'étapes et demande l'intervention de plusieurs enzymes. C'est cette biogenèse complexe qui permet la régulation post-transcriptionnelle de certains miARN. Finalement, ce doctorat a été entrepris parce qu'il semble de plus en plus évident que ce mode de régulation est crucial à la bonne expression de certains miARN. Dans cette thèse, un nouveau mode de régulation post-transcriptionnelle des miARN a été caractérisé à l'aide de techniques biochimiques et biophysiques.

1.1.1 Historique

La première description d'un miARN a été publiée en 1993 par Rosalind C. Lee, Rhonda L. Feinbaum et Victor Ambros (1). Ils s'intéressaient au gène *lin-4* de *C. elegans* qui est primordial au bon développement larvaire par son inhibition de l'activité de la protéine LIN-14. À ce moment, il était connu que l'extrémité 3'-UTR du transcrit *lin-14* était nécessaire pour son inhibition par *lin-4*, mais le mécanisme par lequel *lin-4* bloquait la traduction de *lin-14* restait à élucider. En clonant *lin-4*, ils ont conclu que ce dernier n'encode aucune protéine, mais produit plutôt deux ARN de 22 et 61 nt. Ils ont également proposé que l'ARN de 61 nt forme une tige-boucle et soit le précurseur de l'ARN le plus court, ce dernier étant produit suite à un procédé post-transcriptionnel. Finalement, les laboratoires Ambros (1) et Ruvkun (2) ont remarqué que la séquence du plus court ARN, produit par le gène *lin-4*, était partiellement complémentaire à sept régions présentes et conservées dans l'extrémité 3'-UTR de *lin-14*. Ils

ont donc proposé un modèle où l'ARN *lin-4* inhibe l'expression de *lin-14* par complémentarité de séquence dans la région 3'-UTR de l'ARNm.

Quelques années plus tard, une deuxième cible de *lin-4* est identifiée (3). Il s'agit de *lin-28*, un gène codant pour une protéine essentielle à la régulation temporelle du développement. En identifiant des régions complémentaires à *lin-4* dans l'extrémité 3'-UTR de *lin-28*, il a été démontré qu'un seul miARN peut avoir plusieurs cibles différentes et donc, jouer un rôle potentiellement important dans la régulation temporelle du développement.

Pendant près de 10 ans, le microARN *lin-4* est resté un cas isolé et restreint à *C. elegans*. En 2000, Reinhart *et al.* (4) identifient un nouveau microARN chez *C. elegans* nommé *let-7*. Celui-ci est un gène temporellement régulé et encode un ARN de 21 nt complémentaire à des éléments retrouvés dans l'extrémité 3'-UTR des gènes hétérochroniques *lin-14*, *lin-28*, *lin-41*, *lin-42* and *daf-12*. En contrôlant l'expression de plusieurs gènes hétérochroniques, des gènes faisant partis d'un réseau hiérarchique contrôlant le développement temporel, les miARN *lin-4* et *let-7* ont donc le pouvoir de réguler le développement de *C. elegans* (voir Figure 1-1). Le modèle proposé pour expliquer l'inhibition par *let-7* est le même que pour *lin-4*; en liant l'extrémité 3'-UTR de sa cible, le court ARN d'environ 22 nt bloque la traduction de l'ARNm par un mécanisme restant à déterminer.

Ce qui a véritablement déclenché l'intérêt pour le champ de recherche sur les microARN, c'est la découverte par Pasquinelli *et al.* que la séquence et l'expression temporelle de *let-7* est conservée au travers plusieurs espèces (5). Par leur démonstration que le miARN *let-7* est retrouvé chez toutes les classes d'animaux bilatéraux et que, tout comme chez *C. elegans*, l'expression de *let-7* y est temporellement régulée, Pasquinelli *et al.* ont identifié un nouveau système de régulation des gènes conservé au cours de l'évolution. Cette nouvelle classe de petits ARN permettant une régulation post-transcriptionnelle a d'abord été nommée petit ARN temporel (*small temporal RNA*, stRNA), puisque les deux membres de cette classe régulaient le développement de *C. elegans*. Rapidement, de nouveaux membres de cette classe d'ARN ont été identifiés, plusieurs ne jouant pas de rôle dans le développement. Le nom de cette classe de petits ARN régulateurs a donc été changé pour *microARN* (miARN) (6-8).

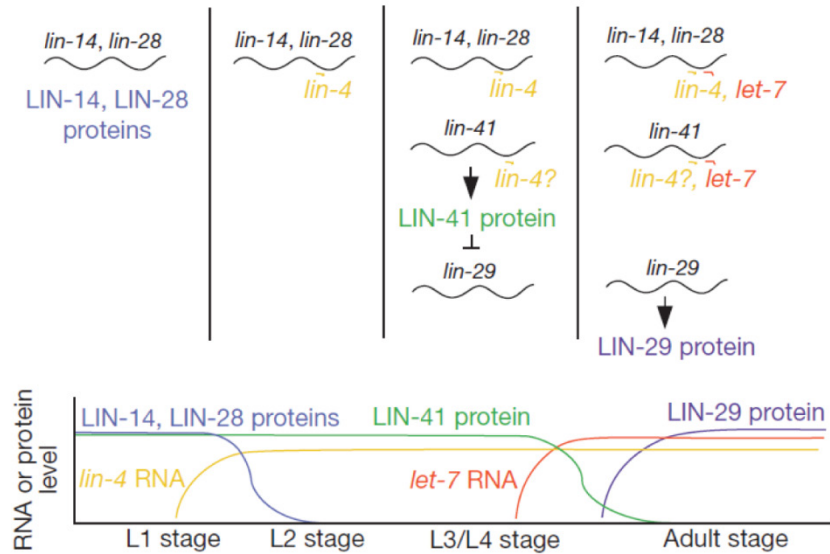


Figure 1-1. Modèle de la régulation successive des gènes hétérochroniques par les miARN *lin-4* et *let-7* chez *C. elegans*. Provenant de (4).

Plus d'une décennie après la fondation de cette classe de petits ARN non codant, de nouveaux miARN ont été découverts chez les animaux (bilatériens, cnidaires et les éponges), les plantes, les algues vertes ainsi que chez certains virus. Actuellement, plus de 2000 miARN différents ont été identifiés chez l'humain et près de 400 sont connus chez *C. elegans* ou chez *Arabidopsis thaliana*.

1.1.2 Transcription des microARN

Tel que mentionné précédemment, les microARN matures sont le produit d'une biogenèse à multiples étapes. Suite à sa transcription dans le noyau, le transcrit primaire (pre-miARN) est clivé une première fois par une endonucléase nommée Drosha. Le microARN précurseur (pre-miARN) produit par ce clivage est exporté dans le cytoplasme pour y être clivé par Dicer, une autre endonucléase. Le duplexe d'ARN, d'environ 22 paires de bases, formé par ce dernier clivage est séparé et l'un des deux brins est chargé sur le complexe inhibant l'expression par l'ARN (*RNA-induced silencing complex*, RISC) pour inhiber

spécifiquement l'expression du gène cible. À noter que les deux brins formant le duplex peuvent être chargés sur RISC. Puisque cette thèse porte sur un mécanisme de régulation post-transcriptionnelle des miARN, une attention particulière sera portée à la biogenèse de ces derniers.

Au cours de la dernière décennie, plusieurs travaux ont porté sur la biogenèse des miARN et nous comprenons mieux maintenant, comment un pri-miARN de quelques milliers de nucléotides peut, suite à plusieurs interventions enzymatiques, former un miARN mature d'environ 22 nucléotides. La transcription des miARN ainsi que le contrôle de celle-ci reste, quant à elle, encore mystérieuse à bien des égards. Il a d'abord été proposé que la polymérase III (pol III) soit responsable de la transcription des gènes de miARN, puisque celle-ci transcrit la majorité des petits ARN tels que les ARN de transfert (ARNt) et le *small nucleolar RNA U6*. Pourtant, certains transcrits primaires ont une longueur de plusieurs kilobases (kb), ce qui ne correspond pas à l'activité de pol III (9). De plus, les pri-miARN possèdent une coiffe ainsi qu'une queue de poly(A) et leur production est sensible à l' α -amanitine (10). Puisque cela sont des caractéristiques des transcrits de la polymérase II (pol II), il est maintenant reconnu que les plus grands pri-miARN sont transcrits par la même ARN polymérase que les ARNm.

Toutefois, on retrouve des répétitions Alu en amont de certains gènes de miARN et certaines régions contenant ces gènes sont enrichies lors d'immunoprécipitation de la chromatine (ChIP) avec des anticorps contre pol III (11). Bref, tous les miARN ne sont pas transcrits par la même polymérase. Le consensus actuel est que tous les miARN transcrits comme intron ou exons d'un gène codant pour une protéine sont transcrits par la pol II, alors que les miARN possédant leur propre promoteur sont transcrits soit par la pol II ou la pol III (12). Une récente étude démontre qu'il peut y avoir des exceptions au consensus. Il a été démontré que certains gènes de miARN ne contenant aucun promoteur peuvent être transcrits malgré la présence d'inhibiteurs contre pol I, pol II et pol III (13). La machinerie nécessaire à cette transcription atypique reste inconnue.

Il est estimé qu'environ 50% des gènes de miARN sont localisés à proximité de d'autres miARN (14). Ces polycistrons (*clusters*) de miARN sont transcrits ensemble pour former un long pri-miARN contenant plusieurs pre-miARN différents (6,7,9) appelé unité de transcription (*polycistronic transcription unit*, UT). Dépendant du même promoteur, cette

organisation permet une transcription concertée de plusieurs miARN différents. Le positionnement du polycistron de miARN dans le transcrite est variable (voir Figure 1-2) et influence sa régulation. Certains polycistrons sont retrouvés dans des régions introniques ou exoniques de gènes non codant, alors que d'autres sont intégrés à des gènes codant pour des protéines. Ces derniers sont donc exprimés avec la protéine et répondent aux mêmes stimuli ou inhibitions de transcription.

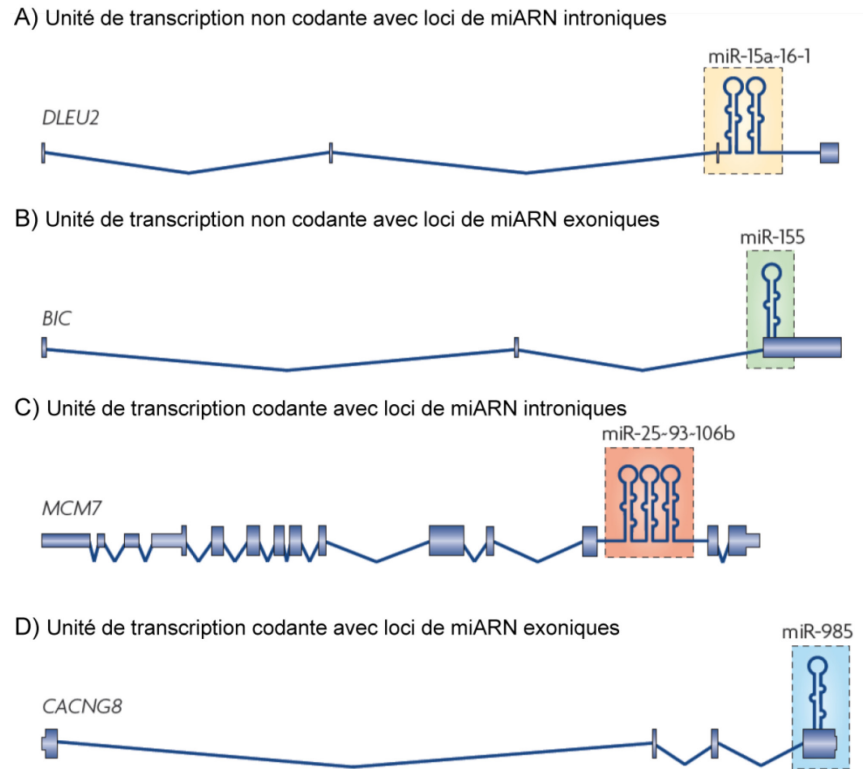


Figure 1-2. Organisation des structures possibles des unités de transcription (UT) contenant des loci des miARN. (A) Unité de transcription non codante avec loci de miARN introniques. (B) Unité de transcription non codante avec loci de miARN exoniques. (C) Unité de transcription codante avec loci de miARN introniques. (D) Unité de transcription codante avec loci de miARN exoniques. Adaptée de (14).

Cette régulation de la transcription des miARN par différents facteurs fait fréquemment partie de boucles de rétroaction. Par exemple, l'expression de miR-145 dans des cellules de cancer du sein provoque l'activation de p53, ce qui inhibe la prolifération et induit

l'apoptose. De plus, cette activation de p53 induit la transcription de miR-145 dans les cellules cancéreuses. Le microARN mir-145 et la protéine p53 forment donc une boucle régulatrice de rétroaction positive induisant la mort cellulaire (15). À l'opposé, la transcription de certains miARN peut être inhibée par la cible de ces derniers. Par exemple, le polycistron de miARN humain MC-let-7a-1~let-7d incluant les miARN let-7a-1, let-7f-1 et let-7d possède un promoteur contenant deux sites de liaison pour c-Myc. Alors que la liaison de c-Myc à ce promoteur inhibe la transcription du polycistron de miARN (16), l'expression de ce dernier inhibe la traduction de c-Myc en permettant aux miARN de se fixer à l'extrémité 3'-UTR de son ARNm (17). Ce type de boucle de régulation de rétroaction doublement négative permet d'amplifier de faibles changements de concentrations des facteurs impliqués.

1.1.3 Clivage du pri-miARN par Drosha et exportation au cytoplasme

Suite à leur transcription dans le noyau, les pri-miARN ont fréquemment une taille de plusieurs milliers de nucléotides. Ce pri-miARN est habituellement formé d'une tige de 33 paires de bases (pb), d'une boucle terminale (TL) et de régions voisines pouvant parfois mesurer quelques milliers de nucléotides. La première étape, dans la voie de maturation d'un miARN, est le clivage de ce transcrit primaire par une endonucléase nommée Drosha (18) (voir Figure 1-3 et Figure 1-4).

Drosha est une endonucléase de plus de 150 kDa contenant deux domaines RNase III (RIIID) ainsi qu'un domaine de liaison à l'ARN double brin (dsRBD) [voir Figure 1-3, (19,20)]. Ces deux types de domaines sont retrouvés chez toutes les RNase III, mais leur organisation variable est utilisée pour les classer. Les RNase III de classe I que l'on retrouve chez les bactéries et les levures ne possèdent qu'un domaine RNase III et un dsRBD. La protéine Drosha qui est conservée chez les animaux constitue la classe II, alors que la protéine Dicer contenant deux RIIID, un dsRBD, un domaine hélicase ainsi qu'un domaine PAZ constitue la classe III (20,21) (voir section 1.1.4 pour plus de détails). Pour cliver un double brin d'ARN, deux domaines RNase III sont requis. Pour les RNase III de classe I retrouvées chez les bactéries et les levures, deux protéines dimérisent pour associer leur RIIID (22). Par contre, puisque Drosha et Dicer possèdent deux domaines RNase III par protéine, ces derniers

forment un dimère intramoléculaire. Par conséquent, une seule protéine est apte à cliver un double brin d'ARN (23).

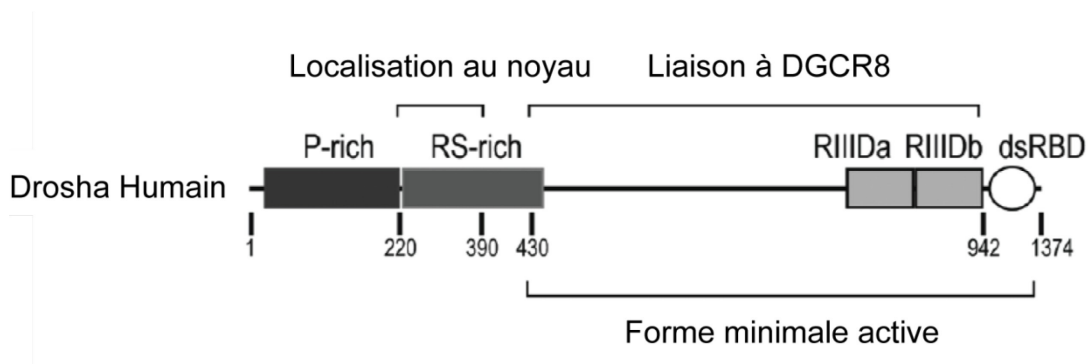


Figure 1-3. Organisation et rôles des différents domaines de Drosha. RIIID indique les domaines catalytiques RNase III, dsRBD est le domaine de liaison à l'ARN double brin, P-rich indique une région riche en proline et RS-rich est une région riche en arginine et en sérine. Le signal de localisation nucléaire est localisé dans le domain RS-rich. Adaptée de (21).

Si Drosha avec ses deux RIIID possède le site catalytique nécessaire au clivage du pri-miARN, la protéine n'est pas suffisante pour cette étape de maturation d'un miARN. Drosha fait partie d'un grand complexe protéique nommé le Microprocesseur d'une taille d'environ 500 kDa chez la drosophile (24) et d'environ 650 kDa chez l'humain (25). Parmi les partenaires protéiques de Drosha identifiés, on retrouve une protéine conservée nommée DGCR8 (DiGeorge Syndrome Critical Region Gene 8) ou Pasha (partner of Drosha) qui est essentielle au clivage du pri-miARN par Drosha (23,25,26). DGCR8 est une protéine de 86 kDa (chez l'humain) possédant deux dsRBD essentiels au clivage des pri-miARN (27) ainsi qu'un motif WW (21). Il est maintenant connu que DGCR8 peut lier l'hème ce qui provoque la formation d'un homodimère (28) là où le motif WW est au centre de l'interface des deux protéines (29). Contrairement à Drosha (30), DGCR8 est capable, via ses deux dsRBD, de lier directement le pri-miARN (27,30). Une récente étude propose que cette reconnaissance spécifique du pri-miARN est rendue possible par la formation d'un complexe multimérique où

plusieurs dimères de DGCR8 lient le même ARN de façon hautement coopérative (31). Cette habileté permet à DGCR8 d'agir comme ancre moléculaire en liant la jonction entre la base de la tige de 33 pb et les régions voisines simples brins du pri-miARN. Une fois le complexe DGCR8/pri-miARN formé, Drosha peut interagir à la fois avec DGCR8 et le pri-miARN. La formation de ce complexe, stabilisé par le domaine C-terminal de DGCR8 (27), positionne précisément le site actif de Drosha 11 pb au-dessus de la jonction. Suite au clivage du double brin d'ARN, le pre-miARN nouvellement formé est relâché (30) (voir Figure 1-4). De récentes études démontrent que tout ce processus s'effectue possiblement de façon co-transcriptionnelle (32-34). Le pre-miARN relâché est constitué d'une tige de 22 pb ainsi que d'une boucle terminale. Il possède une extrémité 5'-monophosphorylée qui sera celle du miARN mature ainsi qu'une extrémité 3'-OH simple brin de 2 nt, produit typique du clivage par une RNase III (20).

Il existe une voie de biogenèse empruntée par quelques miARN qui ne demande pas l'utilisation de Drosha/DGCR8. Ces miARN sont d'abord transcrits sous forme d'intron et possèdent toutes les caractéristiques d'un pre-miARN une fois correctement épissés, débranchés et repliés. Étant à la fois un intron et un microARN, cette classe de miRNA a été nommée *mirtron* (35). Cette voie d'abord décrite chez *C. elegans* et chez la Drosophile (35,36) est également présente chez les mammifères (37). Même si la maturation du pri-miARN est différente pour les mirtrons, leur exportation au cytoplasme et leur clivage par Dicer s'effectue par la voie de biogenèse canonique des miARN (voir Figure 1-4).

La protéine responsable de l'exportation des pre-miARN dans le cytoplasme est l'Exportin-5, une protéine RanGTP dépendante, liant l'ARN double brin (38,39). Cette protéine, qui est également responsable de l'export de certains ARN de transfert (40), reconnaît ses cibles par les nucléotides simple brin présents à l'extrémité 3' (39,41). Par cette spécificité envers une caractéristique des pre-miARN correctement clivés par Drosha, l'Exportin-5 agit comme un filtre lors d'une étape clé de la biogenèse des miARN. Cette étape d'exportation des pre-miARN peut être limitante et une surexpression de l'Exportin-5 permet d'améliorer l'activité des miARN (42).

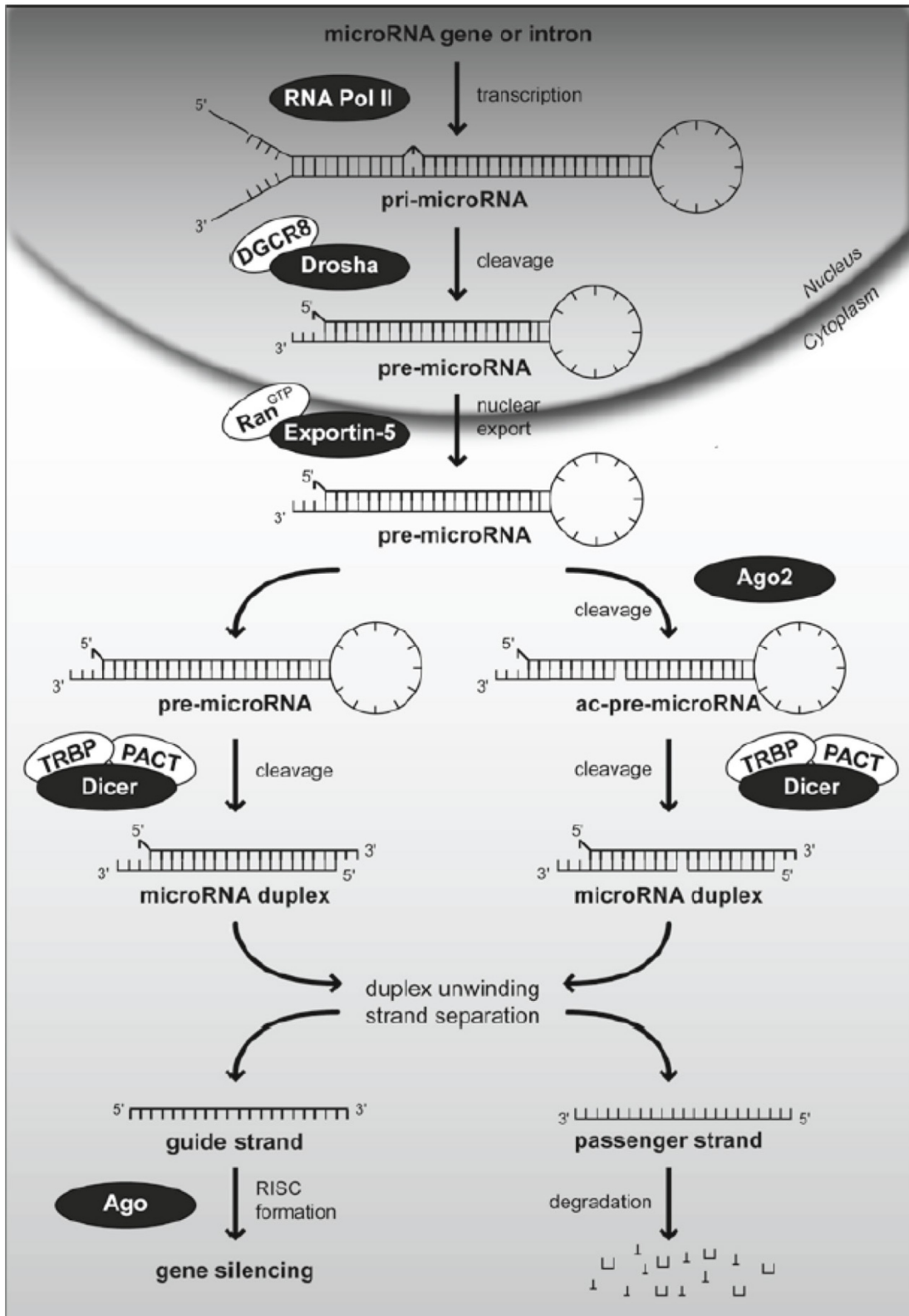


Figure 1-4. Voie de biogenèse canonique des microARN. Provenant de (43).

1.1.4 Clivage par Dicer du pre-miARN

Dans le cytoplasme, le pre-miARN est clivé une deuxième fois par une endonucléase nommée Dicer (voir Figure 1-4) pour former un duplex d'environ 22 pb (44-48). Pratiquement tous les eucaryotes possèdent au moins un homologue de cette protéine essentielle au développement (49). Par exemple, *D. melanogaster* possède l'homologue Dicer 1 qui est requis pour la biogenèse des miARN ainsi que Dicer 2 impliqué dans la voie des siARN (14,50,51). Dicer est une RNase III de classe III, le sous-groupe le plus structurellement complexe de cette famille. La protéine humaine Dicer comporte un domaine DEAD-Helicase de fonction mal définie à l'extrémité N-terminale, un domaine DUF283 (*domain of unknown function*) adoptant un repliement semblable aux dsRBD (52), un domaine PAZ, deux domaines RNase III et un domaine de liaison à l'ARN double brin (voir Figure 1-5). La forme minimale de Dicer pour observer une activité est constituée du domaine PAZ ainsi que des deux RIIID. La protéine Dicer de *Giardia intestinalis* (GI) est d'ailleurs limitée à ces domaines, mais est tout à fait fonctionnelle. La structure tridimensionnelle de GI-Dicer en complexe avec un double brin d'ARN [voir Figure 1-5B (53)] ainsi qu'une structure à basse résolution de la protéine humaine (54) sont connues et permettent de mieux comprendre le fonctionnement de cette enzyme.

Le domaine PAZ est conservé chez Dicer ainsi que chez les protéines Ago (55,56), deux familles de protéines impliquées dans l'inhibition des gènes par de petits ARN. Ce domaine, indispensable pour l'activité de Dicer (22), est responsable de lier l'extrémité du pre-miARN. Il est connu depuis quelques années que l'extrémité simple brin de 2 nucléotides en 3' des pre-miARN est reconnue par le domaine PAZ. Différentes études structurales (53,57-59) ont démontré que ce domaine possède une pochette hydrophobe fortement conservée qui sert d'ancrage à ces deux nucléotides simple brin en 3'. La face des nucléotides insérés dans le domaine PAZ est exposée au solvant. Cette absence de spécificité de séquence, en plus d'un espace suffisant dans la pochette, permet au domaine PAZ de lier toutes les combinaisons de nucléotides possibles et donc tous les pre-miARN (57). Récemment, il a été démontré que le domaine PAZ de Dicer possède également une pochette reconnaissant l'extrémité 5' des pre-miARN (60). Cette pochette reconnaît le phosphate de l'extrémité 5' du pre-miARN à l'aide

de plusieurs arginines et permet le bon positionnement du pre-miARN pour son clivage par les domaines RNase III.

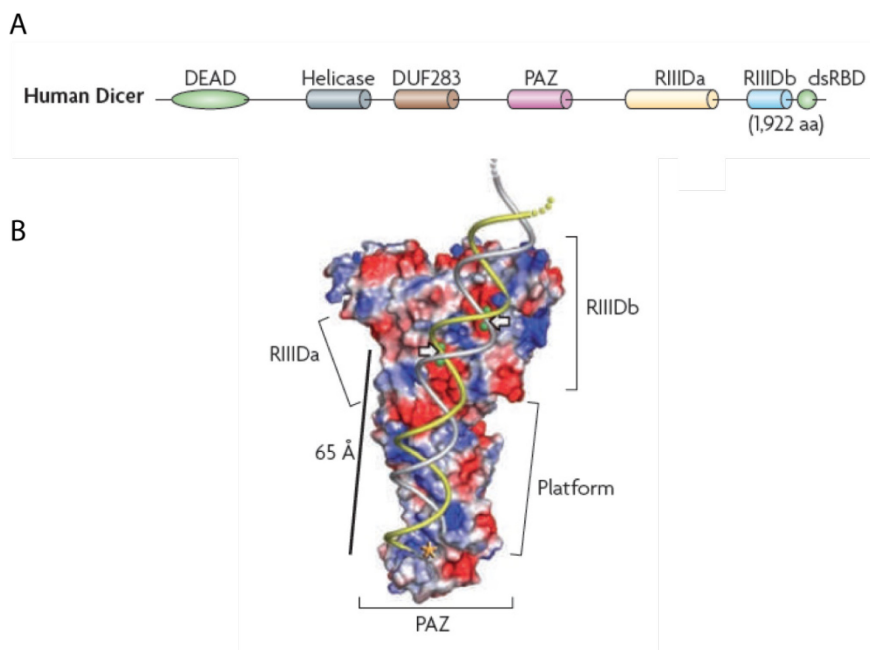


Figure 1-5. Structures primaire et tertiaire de la protéine Dicer. (A) Les différents domaines de la protéine humaine Dicer et (B) la structure tridimensionnelle des domaines RIIID et PAZ de *Giardia intestinalis* en complexe avec de l'ARN double brin. Adaptée de (14,53).

Le domaine PAZ est joint au premier domaine RNase III par une hélice basique servant de plateforme stabilisatrice longeant la tige du pre-miARN (53). Tout comme Drosha, les deux RIIID de Dicer forment un dimer intramoléculaire de part et d'autre du site de clivage de l'ARN (voir Figure 1-5B). Par contre, Dicer n'utilise pas de partenaires protéiques jouant le rôle de règle moléculaire (*molecular ruler*) pour localiser ce site de clivage. La liaison de l'extrémité du pre-miARN au domaine PAZ permet plutôt la localisation idéale du duplex intramoléculaire de RIIID. Chez les eucaryotes inférieurs, cette orientation s'effectue à partir de l'extrémité 3' (53), alors que la pochette en 5' est celle utilisée chez les eucaryotes supérieurs (60). Dans tous les cas, le clivage de la boucle terminale par Dicer produit une nouvelle extrémité asymétrique avec deux nucléotides simple brin en 3'. Chez l'humain, la protéine Dicer possède également un domaine DEAD-Helicase à l'extrémité N-terminale. Ce

domaine adoptant une structure en forme de pince à la base du miARN précurseur (54) est impliqué dans la reconnaissance et la liaison du pre-miARN (61-63).

1.1.5 Protéines interagissant avec Dicer

Même si la protéine Dicer isolée est apte à cliver la boucle terminale d'un pre-miARN, cette enzyme est, *in vivo*, associée à différents partenaires protéiques. Chez l'humain, Dicer interagit avec la protéine liant l'ARN TAR (TRBP) (64,65) et la protéine activatrice de PKR (PACT) (66) deux protéines contribuant à la formation du *RNA-induced silencing complex* (RISC; voir ci-dessous pour détails). Le rôle exact de ces protéines reste à être déterminé, mais il a été démontré que la suppression de ces dernières nuit à la production de miARN (66). L'un des modèles actuellement proposé (43,66-68) est que Dicer, TRBP et/ou PACT et possiblement Ago2 (voir ci-dessous) soient préalablement assemblés avant la liaison et le clivage du pre-miARN. Cet assemblage rendu possible par l'affinité qu'ont TRBP et PACT envers Dicer et Ago2 permettrait d'accélérer le clivage de la boucle terminale et la séparation des brins du duplex.

1.1.6 Chargement du miARN sur RISC

Suite au clivage de la boucle terminale par Dicer, le duplex de miARN est transféré à l'intérieur du RISC sur la protéine Ago2. La protéine Ago2 fait partie de la famille des argonautes et est composée des domaines N-terminal, PAZ, Mid et PIWI. Parmi ces derniers, les domaines PAZ et PIWI sont fonctionnellement les plus importants. Le domaine PAZ, retrouvé chez les argonautes, est similaire à celui de Dicer et permet aussi de lier une des extrémités 3' simple brin de deux nucléotides présentes sur le duplex de miARN (58,59,69,70). Le domaine PIWI peut lier l'extrémité 5' du brin guide (71,72), adopte une structure similaire à celle de la RNase H (70) et possède une activité nucléase dépendante de la présence d'un brin guide (73). Cette activité *slicer* permet à Ago2 de cliver le brin passager du duplex de miARN si l'appariement au brin guide est parfait ce qui en facilite la dissociation. Ce pre-

miARN avec un brin passager clivé est nommé *Ago2-cleaved pre-microRNA* [ac-pre-miRNA, voir Figure 1-4 (74,75)].

En plus d'Ago2, l'humain possède trois autres protéines argonautes bien caractérisées (Ago1, Ago3 et Ago4). Ces dernières ne possèdent pas d'activité slicer, mais sont tout de même utilisées dans les RISC pour l'inhibition de l'expression génique par les miARN (76-78). Ne pouvant cliver le brin passager pour dissocier le duplex de miARN et conserver seulement le brin guide, les autres membres de la famille argonaute utilisent sans doute une hélicase dont l'identité reste encore inconnue (79). Une hélicase est aussi probablement utilisée par Ago2 lorsque l'appariement du duplex n'est pas parfait, ce qui est souvent le cas pour les miARN (80). Plusieurs hélicases pouvant s'associer avec RISC ont été identifiées (p68, p72, Gemin3/4, Mov10), mais aucune d'entre elles n'est une hélicase universelle (43,81). La séparation des brins du duplex se fait sans doute lors du transfert de Dicer à l'une des protéines argonautes. Après le clivage de la boucle terminale par Dicer, l'extrémité la plus stable du duplex est liée par TRBP, alors que l'autre extrémité est liée par Ago (82,83). Par la suite, le brin avec les bases les moins stables à son extrémité 5' est sélectionné comme brin guide et l'autre est dégradé (84,85). Le brin guide, étant correctement chargé sur le RISC, celui-ci peut spécifiquement inhiber l'expression de son ARNm cible par différents mécanismes.

1.1.7 Reconnaissance de l'ARNm cible

Pour la reconnaissance de l'ARNm cible, tous les nucléotides du miARN inclus dans le RISC n'ont pas la même importance. Il est connu depuis la découverte de *lin-4* que les nucléotides présents à l'extrémité 5' du miARN sont les plus conservés en plus d'être complémentaires à ceux retrouvés sur l'ARNm (1,2,86). À l'aide d'études bioinformatiques, il a été précisé que les nucléotides importants à la spécificité des miARN sont localisés aux positions 2 à 7 du petit ARN (87). L'importance de cette région en 5' a également été confirmée expérimentalement par des études de mutagenèse (88). De plus, le chargement du miARN sur Ago force cette région à adopter une conformation similaire à celle d'une hélice de type A, tout en exposant les faces Watson-Crick au solvant (89). Cette structure de l'ARN

permet de faciliter la liaison du RISC à l'ARNm cible. Le site présent sur l'ARNm s'appariant parfaitement à cette séquence du miARN de six nucléotides nommé *seed* est soit précédé d'une adénosine, suivi d'un septième nucléotide apparié au miARN ou possède à la fois l'adénosine et le septième nucléotide supplémentaire [voir Figure 1-6 (90)]. Ces trois types de sites cibles canoniques sont, non seulement celles les plus fréquemment retrouvés chez les miARN connus (91), ils sont aussi les plus efficaces pour inhiber l'expression du gène cible (92).

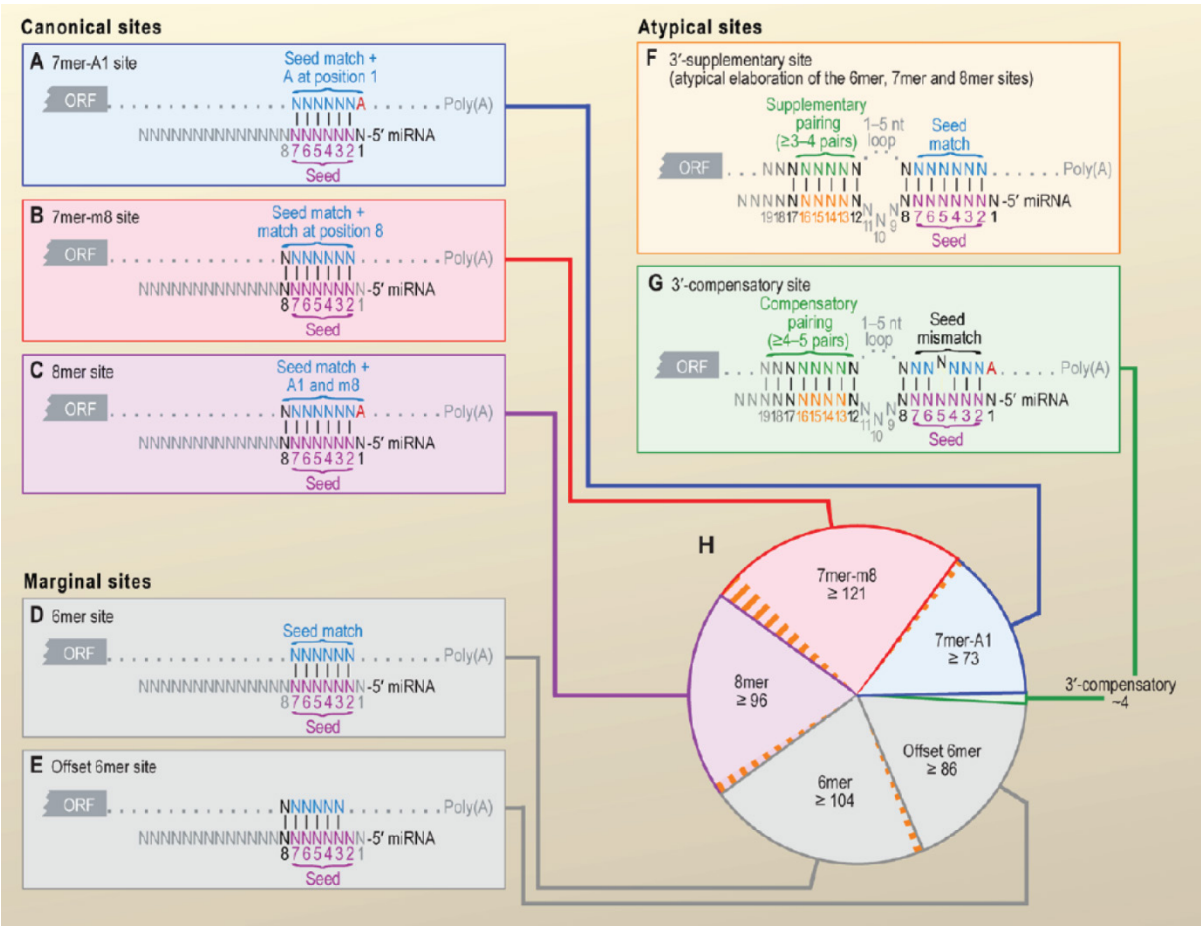


Figure 1-6. Les différents types de séquences *seed* des microARN et leur site d'appariement sur l'ARNm. Les barres verticales indiquent des paires de base Watson-Crick, le graphique (H) indique la fréquence des différents types de sites cibles. Provenant de (91).

Dans de rares cas (voir Figure 1-6F), une deuxième région du miARN s'apparie à l'ARNm cible. Cette région nommée *site supplémentaire en 3'* permet à environ quatre nucléotides supplémentaires de s'apparier à l'ARNm. L'effet de ce site supplémentaire sur la reconnaissance de l'ARNm cible pour un miARN liant un site canonique est faible et ce type de miARN est rare (91,93). Par contre, si un miARN dont la région *seed* lie un site non canonique possède une telle région appariée en 3', cette dernière peut jouer un rôle de régulateur important. Ces régions nommées *site compensatoire en 3'* (voir Figure 1-6G) permettent de compenser pour un appariement plus faible de la séquence *seed*. Par exemple, *C. elegans* possède plusieurs membres de la famille *let-7* qui présentent tous la même séquence *seed* (*let-7*, miR-48, miR-84 et miR-241) alors que le reste de la séquence du miARN est variable (86). Étant exprimé avant *let-7* dans le cycle de développement de *C. elegans* (94), miR-48/84/241 pourraient inhiber, de façon précoce, l'expression de *lin-41* (4,5). Pour éviter cette inhibition, les sites *let-7* présents sur l'ARNm *lin-41* possèdent une séquence *seed* imparfaite pour la famille *let-7* et un site compensatoire en 3' spécifique à *let-7*. Ces deux caractéristiques empêchent les miARN partageant la séquence *seed* de *let-7* d'inhiber l'expression de *lin-41* tout en permettant l'inhibition spécifique par *let-7* (90,91). Malgré cette fonction régulatrice efficace, les sites compensatoires en 3' restent peu fréquents (voir Figure 1-6H).

Depuis la découverte des premiers miARN, il est connu qu'un miARN peut avoir plusieurs sites de liaison présents sur le même 3'-UTR d'un ARNm (1,2,4). Dans la majorité des cas, ces sites agissent de façon indépendante l'un de l'autre et leur effet est additif (93). Par contre, une coopérativité positive est observée lorsque deux sites d'un même miARN sont localisés de 8 à 40 nt l'un de l'autre (93). De plus, il a été démontré que cette coopérativité est également possible entre deux sites de miARN de séquences différentes exprimés dans une même cellule. Cet effet coopératif permet une réponse plus efficace à un changement de concentration des miARN et serait la conséquence d'un effet stabilisateur qu'ont les complexes RISC les uns sur les autres (95). La localisation d'un site de miARN à l'intérieur du 3'-UTR n'influence pas seulement son potentiel de coopérativité. En effet, d'autres règles concernant l'environnement du miARN régissent son efficacité d'inhibition. Par exemple, les sites de miARN les plus efficaces sont habituellement localisés dans les régions riches en

nucléotides A et U, près du codon stop et dans une région pauvre en structures secondaires (93). Pourtant, certains ARNm tels que Nanog, OCT4 et SOX2 possèdent des sites pour miARN fonctionnels localisés dans leur cadre de lecture (96). En résumé, certaines règles générales utilisent le positionnement des sites de miARN pour prédire leur efficacité, mais certains miARN demeurent efficaces tout en ne correspondant pas à ce consensus.

1.1.8 Actions du RISC sur les ARNm

La tâche du miARN mature incorporé dans RISC est de guider celui-ci vers un ARNm cible. Une fois que le RISC a lié cet ARNm, l'expression de ce dernier est inhibée par différents mécanismes qui sont toujours à l'étude aujourd'hui. Ce qui permet au RISC d'inhiber l'expression d'un ARNm, c'est sa capacité d'interaction avec plusieurs cofacteurs différents. La composition exacte du RISC est dynamique et la taille de ce complexe protéique varie de 150 kDa à plus de 3MDa (97-99), indiquant que les cofacteurs présents varient selon les besoins du RISC. Les composants minimaux du RISC sont une protéine argonaute et un brin d'ARN guide. Il est intéressant de noter que ce brin d'ARN n'est pas essentiel pour l'inhibition de l'expression et que son rôle est plutôt de déterminer la spécificité de l'interaction entre Ago et l'ARNm (100). Dans cette section, les mécanismes utilisés par le RISC et ses cofacteurs pour inhiber l'expression d'un ARNm seront décrits. Ce champ de recherche étant en pleine évolution, plusieurs de ces mécanismes sont encore des hypothèses de travail et certains mécanismes peuvent sembler contradictoires.

Un partenaire essentiel du RISC pour l'inhibition de l'expression est la protéine GW182. Autant chez la drosophile que chez l'humain, la suppression de GW182 nuit à la répression de l'expression par les miARN (101). Le domaine N-terminal de GW182 contient plusieurs répétitions GW jouant un rôle dans l'interaction avec les protéines Ago (voir Figure 1-7B) (102,103). L'autre domaine important de GW182 est le *silencing domain* situé à l'extrémité C-terminale. Ce domaine sert de plateforme d'interaction pour une multitude de cofacteurs participant à la répression de l'expression génique. Par exemple, le domaine PAM2 est responsable de l'interaction avec la *poly(A) binding protein* (PABP, voir détails ci-dessous). Le *silencing domain* de GW182 pourrait également interagir directement avec les

complexes de déadénylation CCR4-NOT et PAN2-PAN3 [voir détails ci-dessous (104-106)]. En liant Ago, puis en recrutant divers cofacteurs, GW182 est la plateforme qui permet au RISC d'inhiber l'expression génique en réprimant la traduction et/ou en déstabilisant l'ARNm par des mécanismes qui ne sont pas encore complètement élucidés (voir Figure 1-7A). La protéine GW182 est également responsable de diriger le complexe Ago/miARN/ARNm vers les *P-bodies* (107) où l'on retrouve les protéines nécessaires à l'inhibition de l'expression des ARNm (108).

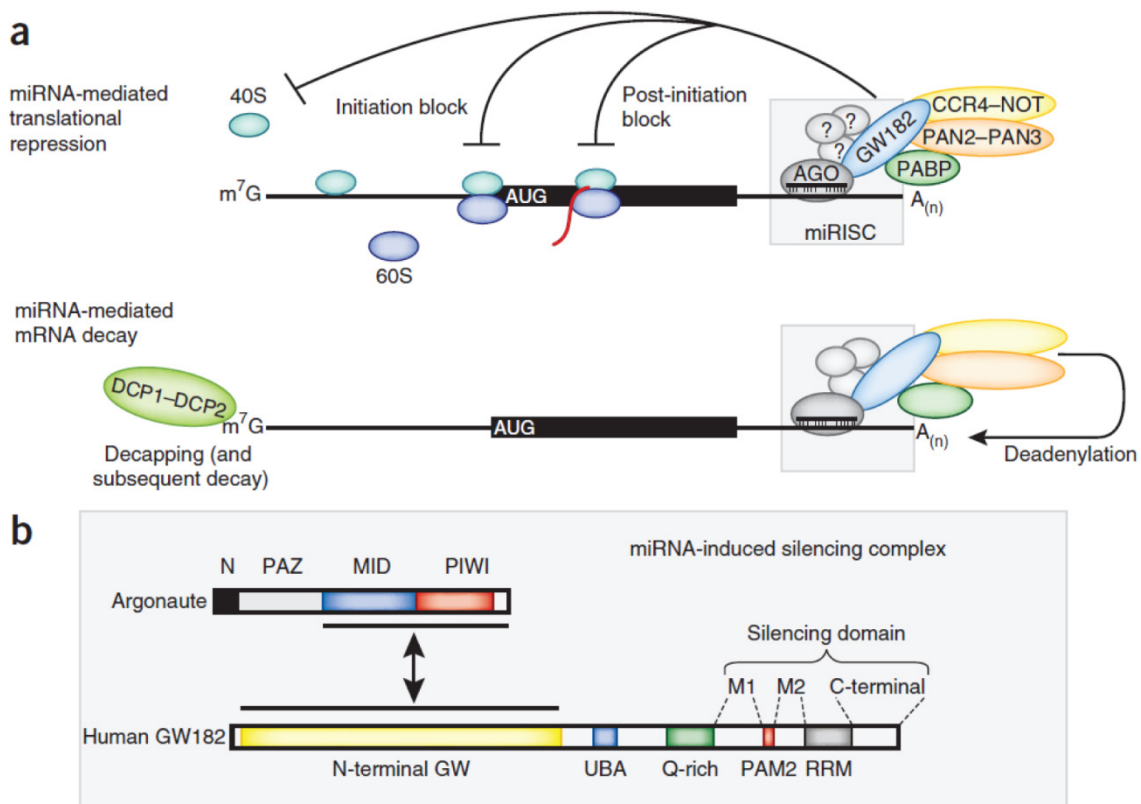


Figure 1-7. Rôle de la protéine GW182 dans l'inhibition de l'expression génique par les microARN. (A) Représentation schématique de la répression de la traduction et de la déadénylation induit par le complexe RISC. (B) Différents domaines d'interaction entre les protéines Argonautes et GW182. Provenant de (109).

Il est connu que les ARNm, dont la traduction est dépendante de la coiffe, sont sensibles à la répression de l'expression par les miARN, alors que des ARNm dont la traduction est indépendante de la coiffe possèdent une résistance à ce type de répression (110-112). Cette dépendance à la présence de la coiffe, pour une répression de la traduction efficace par les miARN, suggère un lien entre les protéines interagissant avec la coiffe et le RISC (voir Figure 1-7A). Jusqu'à maintenant, les mécanismes exacts expliquant cette inhibition sont inconnus. Différents modèles controversés telle que la liaison directe d'Ago2 à eIF4E [*eukaryotic initiation factor 4E*, (113-115)] ou le blocage de la circularisation des ARNm par GW182 et PABP (116,117) apportent des pistes de solutions, mais sont contradictoires avec certaines études (109,118). De plus, il a récemment été démontré que la répression de l'initiation de la traduction par le RISC est possible même en absence de GW182 (119). Un modèle global expliquant les relations entre le RISC et l'initiation de la traduction dépendant de la coiffe reste donc à définir.

Il a également été suggéré que la répression de la traduction se produise à l'étape de l'élongation, alors que les ribosomes sont déjà engagés sur l'ARNm (120). Ce modèle, tout en étant appuyé par la présence de protéines Ago avec des polysomes purifiés (120-122), n'explique pas comment le RISC pourrait réprimer la traduction à l'étape d'élongation. Il a été proposé que le RISC pourrait recruter différentes protéases responsables de la dégradation des protéines en cours de synthèse (120,123), mais ce modèle n'explique pas comment une protéine synthétisée directement au niveau du réticulum endoplasmique peut être régulée par les miARN (110,124). Encore une fois, quoique plusieurs études impliquent le RISC dans la régulation de l'élongation lors de la traduction, aucun mécanisme précis n'explique encore ce phénomène.

L'autre mécanisme majeur, expliquant l'inhibition de l'expression génique par le miARN, est la déstabilisation des ARNm par la déadénylation. Tel que mentionné précédemment, la protéine GW182 interagit avec plusieurs cofacteurs dont les complexes de déadénylation CCR4-NOT et PAN2-PAN3 (104-106). En liant à la fois ces complexes protéiques et PABP, GW182 permet le positionnement idéal de la queue poly(A) pour induire la déadénylation (116). La déadénylation d'un ARNm est habituellement suivie du retrait de la coiffe en 5' par des enzymes tels que DCP1-DCP2, puis survient la dégradation de l'ARNm

par des exonucléases tels que Xrn1 (125). La déstabilisation des ARNm par leur déadénylation est donc un mécanisme menant à la destruction des ARNm contrairement à la répression de la traduction qui bloque seulement leur expression.

De récentes études cinétiques (126,127) ont avancé un modèle à deux étapes expliquant l'inhibition provenant des miARN par plusieurs mécanismes. Dans ce nouveau modèle, la traduction est dans un premier temps réprimée, puis l'ARNm est déadénylé ce qui propose le retrait de la coiffe et éventuellement la dégradation de l'ARNm. Il est aussi possible que tous les mécanismes décrits ci-dessus soient indépendants les uns des autres et possiblement redondants. Le choix du mécanisme utilisé dépendrait donc de différents facteurs tels que la phase de développement de l'organisme ou celle du cycle cellulaire (124). Afin de compliquer davantage les effets potentiels du RISC sur l'expression des ARNm, il est intéressant de mentionner que certains miARN ont la capacité de stimuler l'expression de leur ARNm cible (128-131). Par contre, cet effet des complexes RISC est rare et semble pour l'instant limité aux cellules non prolifératives. En résumé, malgré l'obtention de nombreuses informations au cours des dernières années, beaucoup de travail reste à faire afin de connaître les mécanismes régissant la répression de l'expression génique par le RISC.

1.1.9 Voies de biogenèse alternatives

Bien que la voie de biogenèse canonique décrite précédemment génère la majorité de miARN chez les animaux (voir Figure 1-4 et Figure 1-8A), il existe des voies de biogenèse alternatives indépendantes de Drosha ou de Dicer. Par exemple, les mirtrons sont des miARN dont la forme précurseur est le produit d'un épissage plutôt que de provenir de Drosha (35-37). Dans certains cas d'épissage, les mirtrons possèdent une longue extrémité simple brin qui n'est pas compatible avec la machinerie d'exportation au cytoplasme. Ces mirtrons non conventionnels subissent donc une étape de maturation additionnelle par une exonucléase qui retire ces nucléotides supplémentaires [voir encadré de la Figure 1-8B (132)].

En fait, l'enzyme Drosha est celle qui est le plus souvent remplacée dans les voies non canoniques (voir Figure 1-8B). Par exemple, certains miARN précurseurs proviennent des ARN de transfert (ARNt). Chez la souris, l'ARNt de leucine peut adopter deux conformations

différentes. En plus du repliement classique en trèfle reconnu par la RNase P, cet ARNt peut adopter une structure en tige-boucle. Cette tige possède deux nucléotides simple brin en 3' et peut être reconnue directement par l'Exportin-5 et Dicer (133). De plus, le génome du virus MHV68 (murine γ -herpes-virus 68) contient une tige boucle en aval d'une structure semblable à un ARNt précurseur. Suite à son clivage par la tRNase Z lors du processus de maturation, la tige-boucle produite possède toutes les caractéristiques d'un pre-miARN (134). Cette stratégie permet donc au virus de produire des miARN indépendamment de Drosha.

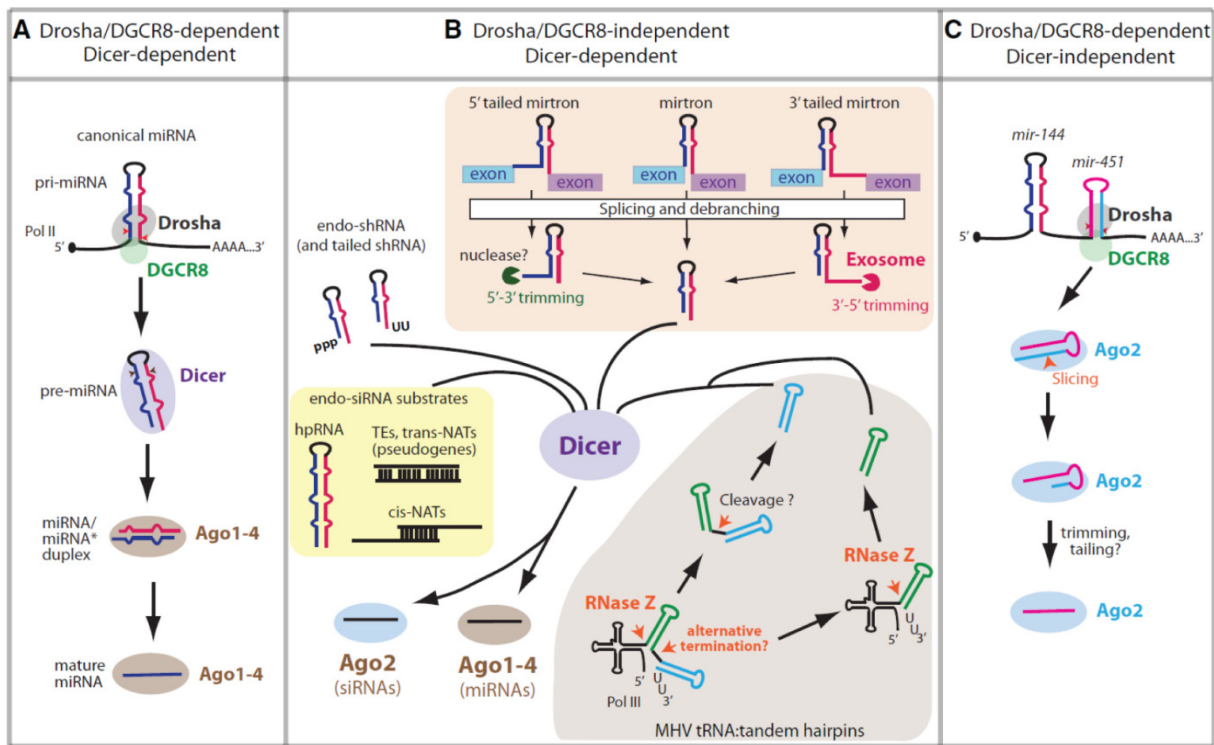


Figure 1-8. Comparaison des voies de biogénèses canonique et alternatives des miARN chez les animaux. (A) Voie de biogénèse canonique des miARN par Drosha et Dicer. (B) Plusieurs voies de biogénèses alternatives sont indépendantes de Drosha pour la génération d'un pre-miARN. (C) Certains produits de Drosha peuvent être incorporés directement sur Ago2. Provenant de (135).

Il existe aussi une catégorie de petits ARN dont l'une des deux extrémités, analogue à celle d'un pre-miARN, est définie lors de la transcription (voir Figure 1-8B). Cette classe d'ARN nommée court ARN en tige-boucle endogène ou endo-shRNA possède parfois une

extrémité simple brin qui doit être retirée avant l'exportation au cytoplasme (133). Cette étape de maturation supplémentaire est possiblement la même que celle utilisée par les mirtrons. Finalement, la dernière voie permettant d'éviter le clivage par Drosha est celle des petits ARN interférant endogènes ou endo-siARN. En plus d'être utilisé par la réponse immunitaire (136), le clivage d'un ARN double brin endogène suivi de son incorporation dans RISC est un phénomène retrouvé dans certains types de cellules. Les ovocytes et les embryons par exemple, sont tolérants à l'ARN double brin et utilisent le système endo-siARN pour réprimer l'expression de certains gènes (133,137,138).

L'endonucléase Dicer semble plus difficile à remplacer lors de la maturation des miARN. Pourtant, il existe un cas de miARN pouvant passer outre le clivage par Dicer (voir Figure 1-8C). Après son exportation dans le cytoplasme, pre-miR-451 est directement chargé sur une protéine Argonaute possédant une activité *slicer* qui retire le brin complémentaire en 3'. Les quelques nucléotides supplémentaires sont par la suite retirés par une nucléase inconnue, pour produire un RISC fonctionnel. Malgré l'abondance de voies de biogenèses alternatives, il est important de souligner que la voie canonique de maturation est responsable de la production de la majorité des miARN fonctionnels. C'est donc sur cette dernière que ce concentrera cette thèse.

1.1.10 Comparaison avec la biogenèse des miARN chez les plantes

Bien que ce travail porte sur la voie de biogenèse des miARN chez les animaux, il convient de mentionner que les plantes possèdent également un système de régulation par de petits ARN. La voie de biogenèse et les effets des miARN chez les plantes sont différents de ceux retrouvés chez les animaux. En fait, il est possible que ces deux règnes du vivant aient développé leur système de régulation par les miARN indépendamment l'un de l'autre. Ces deux différentes voies de miARN proviendrait d'un système ancestral commun de siARN, utile pour la défense contre les invasions virales (139).

La plupart des étapes de la biogenèse des miARN chez les plantes sont différentes de celles retrouvées chez les animaux. Tout d'abord, les *loci* de miARN chez les plantes ont habituellement leur propre unité de transcription et leur expression est donc indépendante des

protéines. Aussi, si les végétaux possèdent quelques polycistrons de miARN ces derniers contiennent habituellement une répétition du même miARN plutôt que plusieurs miARN différents tels que retrouvés chez les animaux (140). Ce transcrit primaire de quelques centaines de nucléotides est par la suite clivé une première fois par une endonucléase nommée *Dicer-like 1* (DCL1). Le pre-miARN produit est immédiatement clivé une deuxième fois par DCL1 qui retire la boucle terminale générant ainsi le duplexe de miARN (141). En plus d'utiliser la même enzyme pour cliver le pri-miARN et le pre-miARN, les plantes effectuent ces premières étapes de biogenèse dans le noyau, à l'intérieur de structures nommées *dicing-bodies* (142,143). Avant l'exportation au cytoplasme par HASTY, un homologue d'Exportin-5 (142), les duplexes de miARN sont 2'-O-méthylés par la protéine nucléaire HEN1 (144). Cette modification post-transcriptionnelle a pour but de protéger l'extrémité 3' du duplexe contre l'uridylation qui entrainerait la dégradation de l'ARN (145).

Le mode d'action des miARN de plantes sur les ARNm est similaire à celui des animaux. Suite à leur exportation au cytoplasme, l'un des brins du duplex est chargé sur une protéine Ago (142). Chez *A. thaliana*, la protéine Ago1 est la plus utilisée (146), mais contrairement aux animaux, il semble que toutes les protéines Ago de cette espèce possèdent une activité *slicer* (139,147). Lors de la reconnaissance de l'ARNm cible, les RISC de plante se lient avec une complémentarité presque parfaite à l'extrémité 5'-UTR, au cadre de lecture ou à l'extrémité 3'-UTR de cet ARN messager (148). Cet appariement presque parfait permet le clivage de la cible par Ago, mais la répression directe de la traduction par RISC a également été observée (149). En résumé, si les principes de bases régissant l'inhibition de l'expression par les miARN sont les mêmes chez les plantes et les animaux, la biogenèse et le mode d'action de ces derniers sont différents. L'importance de la régulation par les miARN est également différente. Si on estime que 30% des gènes humains codant pour des protéines sont régulés par des miARN, seulement 1% des gènes de plantes possèdent une telle régulation (139).

1.2 Régulation post-transcriptionnelle des microARN

La voie de biogenèse des microARN, avec ses multiples étapes de maturation successives, est une cible potentielle pour des mécanismes de régulation post-transcriptionnelle. Effectivement, divers facteurs cellulaires interagissent spécifiquement avec les formes immatures de certains miARN afin de moduler leur biogenèse. Dans cette section, différentes protéines permettant une telle régulation post-transcriptionnelle des miARN seront décrites.

1.2.1 Édition des miARN par ADAR

Parmi les facteurs pouvant réguler les miARN de façon post-transcriptionnelle, on retrouve une classe d'enzymes capable d'éditer la séquence de ces ARN. Les membres de la famille ADAR (*Adenosine deaminases acting on RNA*) peuvent lier l'ARN double brin et désaminer les adénosines en inosine. Les vertébrés comptent trois membres de la famille ADAR (ADAR1-3) tous composés de multiples dsRBD et d'un domaine désaminase (150-152). Bien que son domaine désaminase semble fonctionnel, aucune activité n'a été observée chez ADAR3 (153,154). Les fonctions régulatrices semblent donc réservées aux deux membres restant de la famille. Le changement d'adénosine pour inosine a plusieurs conséquences pour l'ARN. D'abord, la nouvelle paire de base I•U adopte une conformation *wobble* qui modifie la structure locale de l'ARN double brin. L'adénosine ainsi modifiée sera reconnue comme étant une guanosine lors de l'appariement d'un brin complémentaire. La modification A→I a donc le potentiel de changer la structure de l'ARN ainsi que le message contenu dans ce dernier.

Le pouvoir régulateur d'ADAR dans la biogenèse des miARN a été démontré pour la première fois avec miR-142 (155). L'édition du pri-miR-142 par ADAR1 et ADAR2 modifie 11 adénosines présentes sur ce transcrit primaire. La désamination de ces nucléosides, en particulier ceux localisés à la base de la tige, inhibe le clivage par Drosha. Il est possible que cette inhibition soit causée par un changement de structure de l'ARN suite à ces multiples modifications (155). Par contre, la répression de Drosha pourrait également être expliquée par

l'encombrement stérique causé par la liaison d'ADAR sur le pri-miARN. En effet, en utilisant des dsRBD non spécifiques pour lier les pri-miARN il est donc concevable que Drosha et ADAR soient en compétition pour lier les mêmes cibles (156). Une régulation similaire a également été observée chez miR-151 (157). Dans ce cas, ADAR1 désamine quelques adénosines sur le transcrit primaire ainsi que sur le microARN précurseur. Contrairement à miR-142, ces modifications ne bloquent pas le clivage par Drosha, mais plutôt celui de Dicer. Dans les deux cas, l'introduction de paires de bases I•U dans les formes immatures du miARN induit le recrutement de Tudor-SN (158), une nucléase clivant l'ARN double brin. La répression de la biogenèse par ADAR semble donc flexible, étant apte à inhiber le clivage de Drosha ou de Dicer tout en induisant la dégradation du miARN immature cible.

L'édition d'un miARN par ADAR, lors de sa biogenèse, n'est pas toujours synonyme de répression et de dégradation. Chez l'humain, le polycistron miR-376 traverse la voie de biogenèse canonique normalement malgré la présence de plusieurs sites de désamination sur cinq des sept miARN composant ce polycistron (159). Si ces changements A→I n'affectent pas la biogenèse de ces miARN, l'effet régulateur reste important. En effet, la majorité de ces sites sont localisés dans les séquences *seed* des miARN matures. Il a été démontré que le remplacement d'une adénosine par une inosine dans ces séquences *seed* est suffisant pour changer la spécificité du miARN (159). Par conséquent, l'interaction des protéines ADAR avec les miARN du polycistron miR-376 au cours de leur biogenèse suffit à les rediriger vers une nouvelle cible.

L'importance exacte de la régulation des miARN par ADAR reste encore à déterminer. Il est estimé que 6% des pri-miARN sont modifiés par ADAR (160), mais ceci n'est qu'une supposition, peut-être sous-estimée, puisque de nombreux pri-miRNA peuvent être désaminés *in vitro* (155). De plus, puisqu'il est connu qu'une désamination n'est pas suffisante pour prédire la répression de la biogenèse (156) l'impact réel d'ADAR sur ces microARN cibles reste à déterminer.

Ce qui module la spécificité d'ADAR reste également à découvrir. Il est connu que l'expression d'ADAR1 varie selon les tissus grâce à sa répression par le miARN miR-1 (161). Il a également été démontré qu'ADAR1 et ADAR2 ne modifient pas les mêmes adénosines et que ces deux enzymes sont en compétition (159). Il est donc concevable que l'expression de

chacune des enzymes soit modulée par des miARN afin qu'elles soient exprimées seulement dans les tissus qui demandent une répression ou une redirection de certains miARN. Une fois exprimée dans la bonne cellule, ADAR reconnaît sans doute son adénosine cible par la présence de renflements (*bulges*) sur la tige du pri-miARN (162), ainsi que par le voisinage immédiat du nucléotide (163-165). Ces règles sur les préférences d'ADAR ont été déterminées avec des ARNm avant que le rôle de cette classe d'enzyme dans la biogenèse des miARN soit connu. À ce jour, il n'a pas été confirmé que la reconnaissance d'un pri-miARN par ADAR s'effectue de la même façon.

1.2.2 Mono-uridylation des miRNA du groupe II de la famille let-7

L'importance du miARN let-7 a déjà été mentionnée dans cette thèse. En plus de son rôle critique dans le développement temporel de plusieurs organismes, une mauvaise régulation de ce miARN est fréquemment retrouvée dans certains cancers. Chez les mammifères, le miARN let-7 est présent en plusieurs copies différentes [voir détails dans la section 1.4 (166,167)]. Ces différentes copies de let-7 (let-7a, let-7b, etc.) possèdent toutes une séquence *seed* identique, alors que le reste du miARN mature est légèrement variable. Parmi les membres de la famille let-7, certains miARN différents peuvent contenir le même miARN mature associé à des boucles terminales variables. Les miARN de ce type sont différenciés à l'aide d'un chiffre. Par exemple, chez l'humain, le miARN let-7a mature est le produit de pre-let-7a-1, pre-let-7a-2 et pre-let-7a-3. Malgré le fait que les membres de cette famille partagent des séquences et des fonctions semblables, ces derniers ne subissent pas tous les mêmes modes de régulation post-transcriptionnelle.

Les membres de la famille let-7 sont sous-divisés en deux groupes (I et II) selon leurs extrémités suite au clivage par Drosha (voir Figure 1-9). En effet, si les membres du groupe I possèdent une extrémité 3' présentant 2 nt simple brin, les pre-miARN du groupe II ont un seul nucléotide à cette extrémité suite à l'action de Drosha (166). Ce produit atypique de Drosha est sans doute causé par la présence d'un renflement d'une uridine, en 5', près du site de clivage de l'enzyme chez les membres du groupe II. Ce nucléotide, potentiellement positionné à l'extérieur de l'hélice, est ignoré par l'endonucléase lors du clivage et permet au pre-miARN

de former une paire de base U•U supplémentaire (voir Figure 1-9B). Cette extrémité simple brin à un seul nucléotide des miARN du groupe II ne peut être liée efficacement par Dicer (voir section 1.1.4 sur la spécificité de cette enzyme). La maturation des membres du groupe II dépend donc d'une classe d'enzymes responsable de l'ajout d'un nucléotide supplémentaire à l'extrémité 3' de ces pre-miARN.

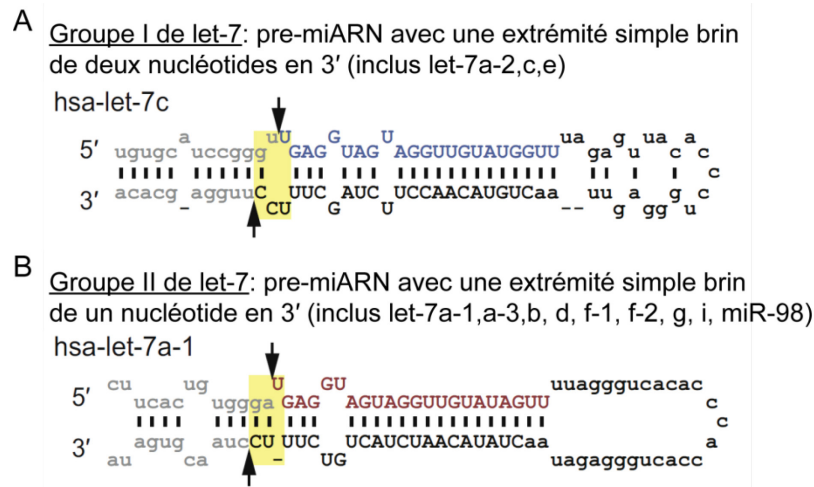


Figure 1-9. Représentation des structures adoptées par les pri-miARN des groupes I et II de la famille let-7. (A) Les membres du groupe I possèdent une extrémité de 2 nt simple brin en 3' après le clivage par Drosha. (B) Ceux du groupe II ont un seul nucléotide simple brin 3' après l'action de Drosha. Ces derniers ont donc besoin d'une mono-uridylation afin de permettre leur clivage par Dicer. Les séquences des miARN matures en 5' sont en lettres majuscules bleues ou rouges, alors que celles des brins complémentaires en 3' sont en lettres majuscules noires. Les sites de clivage de Drosha sont mis en évidence par les flèches noires en plus d'être encadrés en jaune. Adaptée de (166).

Les enzymes responsables de cet ajout sont des tutases (TUT) capables de monouridyler (ajouter une uridine) spécifiquement l'extrémité 3' des pre-miARN du groupe II (166). Trois tutases différentes (TUT2/4/7) peuvent lier transitoirement ces miARN précurseurs afin d'ajouter le nucléotide manquant à leur reconnaissance par Dicer (166,168). Les caractéristiques moléculaires permettant cette liaison spécifique par les tutases ne sont toujours par connues. Bien que cette interaction soit essentielle à la maturation de ce sous-

groupe de miARN, aucun rôle de régulation de l'expression génique n'a encore été attribué à ce mécanisme de mono-uridylation. Pourtant, 90% des miARN de la famille let-7 exprimés chez les vertébrés font partis du groupe II (166). La mono-uridylation est donc une modification post-transcriptionnelle avec un grand potentiel régulateur.

1.2.3 Régulation par SF2/ASF

Contrairement à ADAR et aux tutases, certains facteurs protéiques n'ont pas à modifier la séquence des miARN immatures pour influencer leur biogenèse. La protéine *splicing factor 2/alternative splicing factor* (SF2/ASF) est connue pour son implication dans l'épissage constitutif et alternatif des ARNm (169,170). Par un épissage alternatif, cette protéine est capable d'autorégulation en produisant divers isoformes inactifs de son propre ARNm (171). En plus de son rôle dans l'épissage des ARNm, SF2/ASF est impliquée dans la régulation post-transcriptionnelle de certains miARN. L'un des miARN les plus affectés par cette protéine est miR-7 (172), un miARN conservé et considéré comme suppresseur de tumeurs (173). Bien que SF2/ASF ne soit pas essentielle à l'expression de miR-7, la présence de la protéine améliore la production de ce miARN (172).

Cette fonction régulatrice de SF2/ASF est indépendante de son rôle dans l'épissage des ARNm. Il a plutôt été démontré *in vivo* et *in vitro* que cette protéine est capable de lier directement le transcrit primaire de miR-7 (172). La région d'interaction de la protéine est localisée sur la tige du pri-miARN, sous le site de clivage de Drosha. Cette interaction est spécifique, puisque la mutation de la séquence d'ARN reconnue par SF2/ASF retire tous les effets régulateurs de la protéine (172). La formation de ce complexe permet de stimuler le clivage du pri-miR-7 par Drosha, mais les mécanismes expliquant les effets de SF2/ASF sur l'endonucléase restent inconnus. La stimulation de la biogenèse de miR-7 par SF2/ASF fait partie d'une boucle de rétroaction négative. En effet, ce miARN mature cible l'ARNm de SF2/ASF. En stimulant la maturation de miR-7, SF2/ASF induit donc sa propre répression (172).

Près de quarante miARN différents sont possiblement régulés par cette protéine. Parmi ces derniers, il a été démontré expérimentalement que miR-29b-1, miR-221 et miR-222 voient

leur biogenèse stimulée lors de l'expression de SF2/ASF. De plus, le rôle de régulateur post-transcriptionnel de cette protéine est spécifique à un groupe de miARN, puisque la biogenèse de miR-29b-2 n'est pas affectée par SF2/ASF. En définissant mieux quels sont les miARN régulés par cette protéine, il sera sans doute possible de comprendre les mécanismes par lesquels SF2/ASF améliore la production de pre-miARN par Drosha.

1.2.4 Régulation par KSRP

D'autres protéines, d'abord décrites comme des facteurs impliqués dans l'épissage des ARNm, jouent un rôle dans la régulation post-transcriptionnelle des miARN. C'est le cas de *KH-type splicing regulatory protein* (KSRP) une protéine conservée qui est impliquée dans plusieurs étapes de la production et de la localisation des ARNm (174,175). Cette protéine contient quatre domaines KH tous capables de lier l'ARN (voir Figure 1-10A). De plus, l'arrangement particulier des domaines KH2 et KH3 est propice aux interactions protéine-protéine (176).

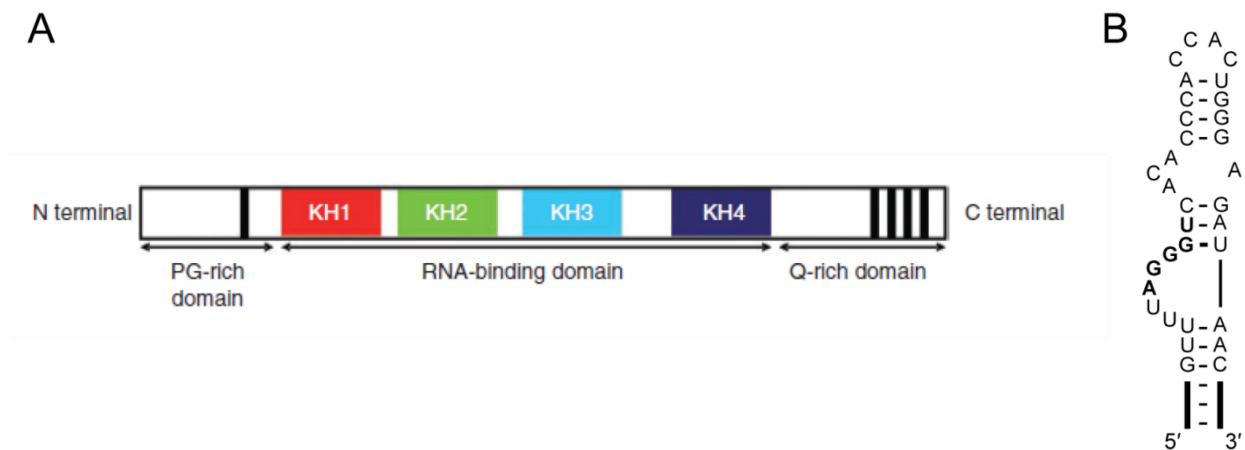


Figure 1-10. Interaction de KSRP avec la boucle terminale de let-7a-1. (A) Représentation schématique de KSRP, les quatre domaines KH liant l'ARN sont mis en évidence. Provenant de (177). (B) Structure secondaire possible de la boucle terminale du miARN let-7a-1. Les nucléotides en gras mettent en évidence le site de liaison de KSRP identifié par Nicastro *et al.* (177).

Ces caractéristiques permettent à KSRP de stimuler la biogenèse d'un groupe de miARN. Il a en effet été démontré que la présence de KSRP est importante à la biogenèse de certains microARN dont let-7a-1 (178). En liant directement le miARN, KSRP permet l'assemblage des complexes protéiques contenant Drosha et Dicer, ce qui entraîne la maturation normale du microARN. La présence de KSRP semble essentielle à plusieurs miARN puisque la répression de cette protéine bloque l'interaction de Drosha à pri-let-7a-1 et pri-miR-21 tout en réduisant la liaison de Dicer à ces miARN (178). Conséquemment, on observe une forte expression de Myc et Ras, deux cibles de let-7, dans les cellules où KSRP est réprimée.

Cet effet de KSRP est spécifique à certains miARN puisque sa répression n'affecte pas tous les miARN. La protéine reconnaît spécifiquement un site présent dans la boucle terminale de let-7a-1 (TL-let-7a-1) et la présence de cette dernière suffit à rendre la biogenèse de ce miARN dépendant de KSRP (178). Une étude récente a démontré que l'interaction spécifique est causée par la liaison du domaine KH3 à une région riche en guanines présente et conservée dans la boucle terminale de let-7 [voir Figure 1-10B (177)]. Les domaines KH2 et KH4 contribuent faiblement à l'interaction en augmentant l'affinité globale de la protéine. L'interaction de KSRP avec let-7 est un bon exemple de régulation post-transcriptionnelle spécifique d'un miARN par une protéine ne faisant pas normalement partie de la voie de biogenèse canonique.

1.2.5 Régulation par hnRNP A1

Les effets de certaines protéines modulant la biogenèse des miARN sont parfois plus complexes. La protéine hnRNP A1 a d'abord été caractérisée comme un facteur impliqué dans l'épissage alternatif (179-182) et l'exportation des ARNm (183,184). Puis, cette protéine a été associée à la traduction (185) et à la stabilité des ARNm (186). Récemment, hnRNP A1 a été définie comme un facteur pouvant réguler de façon post-transcriptionnelle la biogenèse de quelques miARN (187-189). Cette protéine a une taille de 320 acides aminés et comporte deux domaines principaux. Le domaine N-terminal est composé de deux motifs de reconnaissance de l'ARN (RRM) repliés indépendamment qui s'associent pour reconnaître une seule

séquence d'ARN (190-192). Ce domaine, appelé *unwinding protein 1* (UP1), a la capacité de dérouler les hélices d'ARN et peut donc agir comme une chaperonne à ARN (193,194). Le domaine C-terminal de hnRNP A1 est riche en glycines, peut également lier les acides nucléiques et permet à la protéine de se répandre de façon coopérative sur son ARN cible (195).

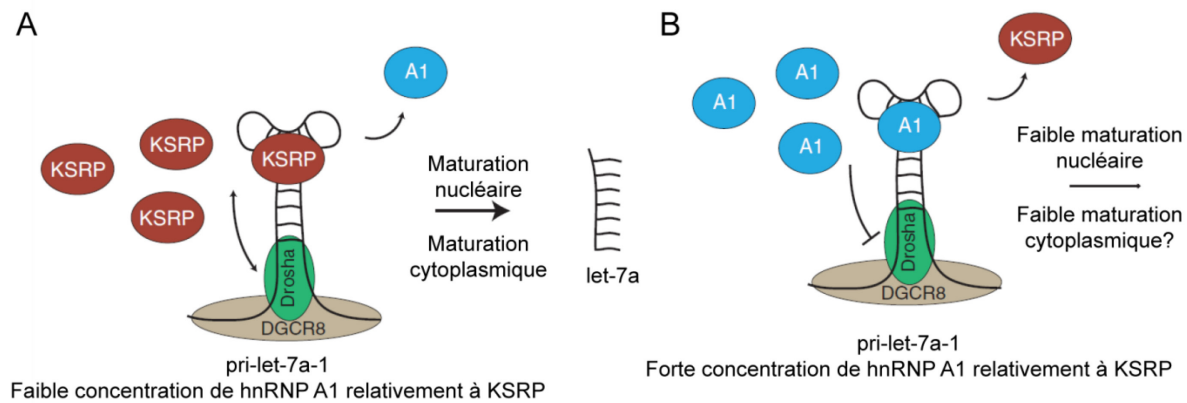


Figure 1-11. Mécanisme démontrant l'importance de la boucle terminale de let-7a-1 dans sa régulation par KSRP et hnRNP A1. (A) Lorsque la concentration de KSRP est plus forte que celle de hnRNP A1, KSRP lie la boucle terminale de let-7a-1 et stimule le recrutement de Drosha. (B) Par contre lorsque hnRNP A1 est abondante, celle-ci déplace KSRP et inhibe le clivage par l'endonucléase. Adaptée de (187)

En plus de ses rôles dans la biogenèse des ARNm, il est maintenant connu que hnRNP A1 est également impliquée dans la voie de maturation de quelques miARN. Il a d'abord été démontré que cette protéine peut lier spécifiquement le transcrit primaire de miR-18a afin de permettre son clivage par Drosha (189). Ce miARN fait partie du polycistron miR-17-92 décrit précédemment dans cette thèse comme un oncogène bien caractérisé (voir section 1.3.2). Ce transcrit primaire comporte donc plusieurs microARN précurseurs, mais seul miR-18a est dépendant de hnRNP A1 pour son clivage par l'endonucléase Drosha. Les séquences voisines de pre-miR-18a dans ce polycistron sont également essentielles. En effet, ce pre-miARN devient indépendant de hnRNP A1 suite à son introduction dans un nouveau polycistron (189). Il a été proposé que la liaison de la protéine à pri-miR-18a permette à

l'ARN d'adopter une structure optimale pour sa reconnaissance et son clivage par Drosha (188,189). En effet, la liaison de hnRNP A1 permet une relaxation du bas de la tige de pri-miR-18a, dans la région du site de clivage par Drosha. L'introduction de mutations contraignant le transcrit primaire à adopter cette structure relaxée permet d'ailleurs à miR-18a d'être indépendant de cette protéine (188). Cette action de la protéine est sans doute rendue possible par l'activité chaperonne de son domaine UP1 (193,194,196).

Le rôle de hnRNP A1 dans la biogenèse des miARN est complexe. Si la liaison de cette protéine est essentielle à la biogenèse de miR-18a, celle-ci inhibe la maturation du miRNA let-7a-1 (187,188). En effet, la présence de hnRNP A1 dans certains types cellulaires différenciés corrèle bien avec une faible expression de ce miARN. De plus, une surexpression de la protéine a pour conséquence l'inhibition de la biogenèse de ce microARN (187). Des études de protection aux ribonucléases ont démontré que la liaison de hnRNP A1 modifie la structure de la boucle terminale de let-7a-1 suite à la reconnaissance de la séquence consensus 5'-UAGGGA/U-3'. Cette séquence est présente chez plusieurs membres de la famille let-7 et contient la région riche en guanines reconnue par la protéine KSRP (voir Figure 1-10B). C'est cette compétition pour lier le même site qui explique le rôle inhibiteur de hnRNP A1 (voir Figure 1-11). Lorsque la concentration de hnRNP A1 est faible comparée à celle de KSRP, cette dernière peut lier TL-let-7a-1 et recruter Drosha. Par contre, lorsque la concentration de hnRNP A1 est élevée, cette dernière lie la boucle terminale de pri-let-7a-1 et empêche de liaison de KSRP. La protéine KSRP étant essentielle au clivage de pri-let-7a-1 par Drosha, cette compétition inhibe la maturation de ce miARN (187). Puisque hnRNP A1 est présente dans le cytoplasme (183), il est possible que cette dernière inhibe le clivage de Dicer grâce à cette compétition avec KSRP. Pour l'instant, une telle inhibition de la maturation des pre-miRNA let-7 par hnRNP A1 n'a pas été observée.

Il est intéressant de constater que hnRNP A1 permet l'expression d'un miRNA oncogène alors qu'elle inhibe la biogenèse d'un autre miARN suppresseur de tumeur. Puisque trop peu de cibles de cette protéine sont connues, il est trop tôt pour affirmer que hnRNP A1 joue le rôle d'un oncogène dans la régulation de la biogenèse des miARN. Par contre, étant donné que cette protéine n'est pas exprimée dans tous les tissus (197), on peut avancer

l'hypothèse que l'expression de hnRNP A1 induit des changements permettant la prolifération cellulaire (198).

Que la biogenèse de let-7a-1 dans les cellules différenciées soit contrôlée par la compétition entre KSRP et hnRNP A1 pour la liaison de boucle terminale met en évidence l'importance de cette dernière. Il est estimé que 14% des miARN possèdent une boucle terminale conservée au travers des espèces (188). Cette pression sélective indique sans doute qu'elles sont essentielles à la fonction de leur miARN, possiblement par leur capacité à lier spécifiquement divers facteurs cellulaires.

1.2.6 Introduction au complexe Lin28 et let-7

La régulation de la biogenèse des miARN de la famille let-7 par la protéine Lin28 est probablement le meilleur exemple en termes de régulation spécifique de la biogenèse de miARN par une protéine. La protéine Lin28 est connue pour jouer un rôle dans la régulation temporelle du développement à l'aide de différents mécanismes de régulation de l'expression génique. Cette thèse s'intéresse plus précisément à sa capacité d'inhiber la maturation des miARN de la famille let-7. Par conséquent, le reste de cette introduction détaille principalement l'état actuel des connaissances sur la famille de miARN let-7 ainsi que sur la protéine Lin28. En dernier lieu, la structure et les fonctions du complexe formé par ces deux molécules seront décrits.

Avant d'étudier en détail le complexe Lin28/let-7, il est bon de résumer comment Lin28 s'intègre aux différentes protéines permettant une régulation post-transcriptionnelle de la famille let-7. Dans cette section (1.2), trois protéines différentes contrôlant la biogenèse de ces miARN ont été présentées. Premièrement, trois différentes tutases permettent la mono-uridylation des miARN du groupe II afin de rendre possible leur clivage par Dicer. Ce mécanisme de mono-uridylation est vrai en l'absence de Lin28. Si la protéine est présente, l'interaction des tutases avec le pre-miARN est stabilisée et plusieurs uridines sont ajoutées. Cette poly-uridylation envoie le miARN immature dans une voie de dégradation [voir détails section 1.6.1 (199-204)]. La présence de Lin28 inverse donc complètement la fonction des tutases.

Deuxièmement, il y a compétition entre KSRP, qui stimule la biogenèse des miARN let-7, et Lin28 qui inhibe leur maturation (177,178). Jusqu'à présent, les mécanismes expliquant cette compétition sont incertains, puisque les sites de liaison principaux des deux protéines semblent parfois différents. De plus, il n'a pas été déterminé si cette compétition avait lieu dans le noyau ou le cytoplasme.

Troisièmement, il n'y a pas de compétition entre hnRNP A1 et Lin28 puisque hnRNP A1 est exprimée dans les cellules somatiques et Lin28 dans les cellules non différenciées (187). Au contraire, hnRNP A1 permet de réprimer let-7 dans certains tissus où Lin28 est absent. Toutes ces différentes interactions de Lin28 seront approfondies dans le reste de cette thèse, mais il était important de les mettre en contexte avec d'autres régulateurs des miARN afin d'apprécier les implications majeures qu'a l'interaction entre Lin28 et les formes immatures de let-7.

1.3 Importance des microARN

En contrôlant l'expression de près de 30% des gènes codant pour des protéines chez les vertébrés (90), les miARN jouent un rôle majeur dans la cellule. Les miARN sont impliqués dans de multiples processus biologiques tels que l'apoptose, la sécrétion d'insuline, la différenciation musculaire et l'hématopoïèse (205,206). Cette thèse portant sur la régulation de la biogenèse des miARN, un simple survol de deux fonctions majeures des miARN sera effectué. Dans cette section, le rôle des miARN dans la régulation temporelle du développement sera abordé. De plus, l'implication de quelques microARN dans différents cas de cancer sera discutée.

1.3.1 Rôle des microARN dans la régulation temporelle

La fonction des miARN dans la régulation temporelle du développement a principalement été étudiée chez *C. elegans*. C'est d'ailleurs chez cet organisme modèle qu'ont été découverts les deux premiers miARN importants dans la régulation du développement. La maturation de *C. elegans* passe par quatre stades larvaires (L1 à L4) avant d'être considéré adulte (voir Figure 1-12). La transition d'une phase à l'autre est contrôlée par les facteurs de transcription LIN-14, HBL-1 et LIN-29 qui coordonnent respectivement les passages L1/L2, L2/L3 et L4/adulte (207). Le miARN lin-4, exprimé à la fin du stade L1 (voir Figure 1-1), permet l'inhibition de l'expression de LIN-14 et donc le passage à L2 (208,209). Le passage au stade L3 est permis par la répression du facteur de transcription HBL-1 présent aux stades L1 et L2. Cette répression est effectuée par plusieurs miARN différents (let-7, mir-58, mir-84 et mir-241) faisant tous partie de la famille let-7. L'activité de ces miARN est redondante et leur mutation simultanée est nécessaire pour empêcher la répression de HBL-1 (94). Cette inhibition par la famille let-7 est modulée positivement par NHL-2, une protéine s'associant au RISC (210). Finalement, la famille let-7 est impliquée dans une boucle de rétroaction avec le facteur de transcription DAF-12. Cette interaction permet d'intégrer des signaux de l'environnement tels que la température ou la disponibilité de la nourriture à la régulation temporelle du développement de *C. elegans* (211). La dernière étape de maturation larvaire est

contrôlée par LIN-29, un facteur de transcription réprimé par LIN-41 (voir Figure 1-12). Lors de l'expression du miARN let-7, la répression de LIN-41 permet l'activité de LIN-29 et donc le passage L4/adulte (4). Tel que discuté dans la section 1.1.7 de cette thèse, cette action spécifique de let-7 est rendue possible par le site compensatoire en 3' propre à ce miARN.

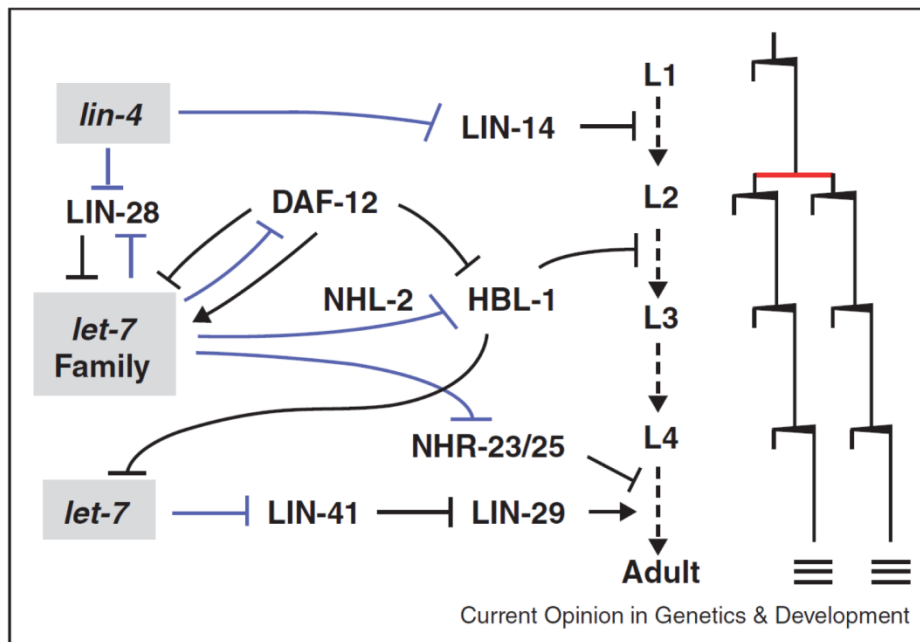


Figure 1-12. Rôles des miARN dans le développement de *C.elegans*. Différents miARN (dans les boîtes grises) permettent de contrôler l'expression de LIN-14, HBL-1 et LIN-29 en interagissant directement ou indirectement avec leur ARNm (ligne bleue). Ces trois facteurs de transcription sont responsables du passage en stade L2, L3 et adulte de *C. elegans*. Chaque stade de développement est caractérisé par la division asymétrique d'une cellule (pointe noire) permettant de conserver des caractéristiques similaires à une cellule souche (*stem cell-like self-renewal division*). Tout juste avant le stade L2, les cellules effectuent une division symétrique illustrée par la ligne rouge. Finalement, à fin du stage L4, les cellules se différencient et entrent dans le stade adulte (trois lignes horizontales superposées). Provenant de (207).

Cette importance des miARN dans la régulation temporelle du développement est conservée chez les animaux. Tel que mentionné précédemment, la séquence et la fonction de *let-7* sont conservées jusqu'aux mammifères (5). Tout comme chez *C. elegans*, *let-7* y est seulement exprimé dans les cellules différenciées. Il a également été démontré que *let-7*

contrôle l'expression de lin-41 dans le développement de la souris (212). Le miARN lin-4 possède également un homologue chez les mammifères, miR-125b, régulant la prolifération hématopoïétique (213). D'autres miARN sont importants pour le développement normal des mammifères. D'ailleurs, la suppression de Dicer chez la souris cause un arrêt du développement au début du stade embryonnaire (49). Bref, les miARN ont un rôle majeur dans la régulation temporelle du développement. Si la description de cette fonction s'est limitée aux animaux pour les besoins de cette thèse, il convient de mentionner que les plantes utilisent également des miARN pour contrôler leur développement (214). Que ces deux règnes du vivant adoptent indépendamment ce système de régulation temporelle démontre bien l'importance des miARN dans le développement.

1.3.2 Les microARN et le cancer

En plus de leurs rôles dans le développement normal des cellules, les miARN sont étudiés pour leurs fonctions de plus en plus établies dans certains types de cancer. Ayant décrit le pouvoir de contrôle qu'ont les miARN, il est peu surprenant qu'une dérégulation de ces derniers puisse avoir un effet marqué sur la croissance cellulaire et sur l'apparition de tumeurs. Cette section est un court résumé de l'état des connaissances actuelles sur le lien entre quelques miARN et le cancer.

La tumorigenèse est fréquemment accompagnée d'une baisse marquée et générale des miARN matures, alors que les niveaux de pri- ou pre-miARN restent constants (215,216). Ce déséquilibre peut être causé par une mauvaise expression ou par l'introduction de mutations dans les différents facteurs responsables de la maturation des miARN (217). Par exemple, une mauvaise régulation de Drosha est liée à certains cancers (218). Aussi, une dérégulation de Dicer est fréquemment retrouvée dans les cellules cancéreuses (219,220) et une réduction des niveaux de cette enzyme est associée à un mauvais pronostic chez les patients avec un cancer du poumon (221). Les autres partenaires impliqués dans la maturation des miARN peuvent également être dérégulés lors de la tumorigenèse. En effet, le gène TRBP des cellules cancéreuses contient parfois des mutations diminuant son expression (222). Une telle dérégulation de la voie de la biogenèse des miARN affecte tous les miARN produits par la

cellule. Puisque 30% des gènes codant pour des protéines sont réprimés par les miARN, il est peu surprenant qu'une telle dérégulation ait un effet marqué sur la croissance cellulaire. Pourtant, il existe des miARN dont la mauvaise régulation individuelle provoque des conséquences importantes pour la cellule.

La première évidence publiée qu'un groupe ciblé de miARN peut jouer un rôle majeur dans la tumorigenèse est associée au polycistron miR-15a-16. Il est connu que l'absence de ce polycistron dans le génome est un indice d'une grande susceptibilité à la leucémie (223). Lorsqu'ils sont correctement exprimés, les miARN de ce polycistron ciblent le gène anti-apoptotique BCL-2 (224). En permettant ainsi l'apoptose cellulaire, miR-15a et miR-16 sont des miARN suppresseurs de tumeurs. À l'opposé, certains miARN qualifiés d'*oncomiR* sont surexprimés en cas de cancer. On retrouve dans cette catégorie le polycistron miR-17-92 qui est surexprimé dans plusieurs carcinomes (225). La fonction des miARN contenus dans ce polycistron n'est pas complètement connue, mais ils sont impliqués dans la régulation de la protéine suppresseuse de tumeur PTEN (87), ainsi que celle de E2F1 impliquée dans la régulation du cycle cellulaire (226). Un autre oncomiR bien caractérisé est miR-21. Ce dernier est surexprimé dans plusieurs types de cancers (227-229) et agit en ciblant le suppresseur de tumeur *Programmed Cell Death 4* [PDCD4, (230)]. De plus, certains miARN sont associés à la formation de métastases. Ces métastamiR tels que miR-10b, miR-21 ou miR-373 favorisent la migration des cellules cancéreuses (231-233).

L'un des miARN suppresseur de tumeur le plus connu et le mieux caractérisé est sans doute let-7. En plus d'être essentiel au développement cellulaire et de promouvoir la différenciation telle que mentionné précédemment, let-7 permet la répression de la prolifération. En ciblant des oncogènes tels que Ras (234), Myc (17) et HMGA2 (235), let-7 limite la prolifération des cellules différenciées. Il est donc peu surprenant que l'on retrouve de faibles niveaux de let-7 dans certains cancers (236). La fonction régulatrice de let-7 est si importante pour le contrôle de la prolifération qu'il a été publié que la présence d'une mutation ponctuelle dans les séquences cibles de let-7 sur l'ARNm de Ras augmente les risques de cancer du poumon (237).

Plusieurs facteurs peuvent expliquer la dérégulation d'un groupe particulier de miARN. Divers facteurs de transcription peuvent influencer la production de pri-miARN et

par le fait même, la quantité de miARN matures présente dans la cellule. En effet, la transcription de certains miARN est régulée par des oncosuppresseurs tel que p53 (238), alors que d'autres sont plutôt contrôlés par des oncogènes comme MYC (239). De plus, certains miARN font partie d'un réseau complexe de régulation et sont indirectement sous le contrôle d'autres miARN. Par exemple, le polycistron miR-103-107 peut provoquer une baisse globale de la production de microARN, tout en ciblant particulièrement la famille miR-200 (240). Cette répression entraîne une production plus agressive de métastases. Finalement, la biogenèse de plusieurs miARN est contrôlée de façon post-transcriptionnelle (voir section 1.2). Un dérèglement de cette régulation peut également expliquer des niveaux variables de miARN matures observés dans certains cancers (167).

1.4 Famille let-7 de microARN

Let-7 est le premier microARN découvert chez l'humain suite à sa caractérisation chez *C. elegans*. Ce qui rend ce miARN particulier, c'est sa séquence et ses fonctions bien conservées à travers les espèces. En effet, tous les animaux bilatéraux possèdent au moins une copie de ce miARN. Chez les mammifères, les oiseaux, les reptiles et les poissons, plusieurs copies semblables de ce miARN sont regroupées sous l'appellation *famille let-7* (241). Tous les membres de cette famille possèdent la même séquence *seed*, alors que le reste du miARN mature est légèrement variable. Un alignement de séquences de tous les membres de la famille let-7 présents chez l'humain (voir Figure 1-13) illustre bien cette conservation. En effet, les nucléotides 2 à 8 formant la séquence *seed* sont parfaitement conservés. Toutefois, le reste de la séquence mature est légèrement variable et les boucles terminales sont différentes pour chacun des membres. Dans certains cas, le même miARN mature est exprimé à partir de différents loci. Chez l'humain, par exemple, le miARN let-7a provient de trois différents loci (let-7a-1, let-7a-2 et let-7a-3, voir Figure 1-13). Tous les organismes ne possèdent pas les mêmes membres de la famille let-7. Effectivement, l'humain exprime neuf miARN let-7 matures provenant de douze précurseurs différents, alors que la souris possède dix miARN let-7 matures exprimés par treize précurseurs (166,167,241). Le peu de variabilité entre les multiples copies de let-7 dans les cellules mammifères rend difficile l'attribution de rôles individuels à chacune de ces copies (242). C'est pourquoi, le terme plus global *let-7* est régulièrement employé lorsque les fonctions des miARN de cette famille sont étudiées.

L'expression de let-7 est temporellement régulée chez tous les organismes où celui-ci est exprimé [voir section 1.3.1 (4,5,243)]. Ce contrôle est principalement causé par des mécanismes de régulation post-transcriptionnelle, puisque les niveaux de pri-let-7 sont constants au cours du développement alors que l'expression de let-7 mature y est variable (244-246). De plus, une augmentation de la production de pri-let-7 n'est pas nécessairement accompagnée d'une hausse de let-7 mature (247). Divers facteurs cellulaires impliqués dans la régulation post-transcriptionnelle de let-7 ont déjà été décrits (voir section 1.2), mais la protéine régulatrice clé de let-7 est sans doute Lin28 (199,201,245,248-250). En effet, Lin28

est la protéine responsable de l'inhibition de la maturation de let-7 dans les cellules non différenciées ainsi que dans certains types de cancer (voir section 1.5 et 1.6 pour détails).

```

hsa-let-7a-1  UGAGGUAGUAGGUUGUAUAGUUUU-AGGGUC----AC---A--CCCACCACUGGGAGAUAACUAUACAAAUCUACUGUCUUUC
hsa-let-7a-2  UGAGGUAGUAGGUUGUAUAGUUU--AGAAUU----AC---A--UCA-----AGGGAGAUAACUGUACAGCCUCCUAGCUUUC
hsa-let-7a-3  UGAGGUAGUAGGUUGUAUAGUUUG--GGGCU----CU---G--CCC--UGCUAUGGGAUAACUAUACAAAUCUACUGUCUUUC
hsa-let-7b   UGAGGUAGUAGGUUGUGUGUUUC-AGGGCA-GUGAUGUUG--CCC--CUCGGAAGAUAACUAUACAAAUCUACUGCCUUC
hsa-let-7c   UGAGGUAGUAGGUUGUAUAGUUU--AGAGUU----AC---A--CCC-----UGGGAGUUAACUGUACAAAUCUACUGCUUUC
hsa-let-7d   AGAGGUAGUAGGUUGCAUAGUUU-AGGGCA-GGGAUUUUG--CCC--ACAAGGAGUAACUAUACAGCCUCCUAGCCUUC
hsa-let-7e   UGAGGUAGGAGGUUGUAUAGUUG--AGGAGG----AC---A--CCC-----AAGGAGAUCACUAUACGCCUCCUAGCUUUC
hsa-let-7f-1  UGAGGUAGUAGAUUGUAUAGUUGU-GGGGUA-GUGAUUUUA--CCC-UGUUCAGGAGAUAACUAUACAAAUCUAUUGCCUUC
hsa-let-7f-2  UGAGGUAGUAGAUUGUAUAGUUUU-AGGGUC----AU---A--CCC-CAUCUUGGAGAUAACUAUACAGCUCUACUGUCUUUC
hsa-let-7g   UGAGGUAGUAGUUUGUACAGUUUG-AGGGUCUAGAUACCA--CCC-GGUACAGGAGAUAACUGUACAGCCACUGCCUUGC
hsa-let-7i   UGAGGUAGUAGUUUGUCUGUUUGGUCGGGU-UGUGACAUUG--CCC-GC-UGUGGAGAUAACUGCGCAAGCUACUGCCUUGC
hsa-mir-98   UGAGGUAGUAAGUUGUAUUGUUUGU-GGGGUA-GGGAUUUAGGCC--CAAUUAGAAGAUAACUAUACAAAUCUA-CUUUC
mmu-let-7g   UGAGGUAGUAGUUUGUACAGUUUG-AGGGUCUAGAUACCA--CCC-GGUACAGGAGAUAACUGUACAGCCACUGCCUUGC
***** *  ***      ***      *                               *                               * * *** *

```

Figure 1-13. Alignement de séquences des pre-miARN let-7 retrouvés chez l'humain. Les séquences des miARN matures en 5' sont en rouge, leur séquence complémentaire en bleue et la séquence de la boucle terminale (TL) en noire. Cette boucle terminale est la portion du pre-miARN clivée par Dicer. Les astérisques indiquent les nucléotides parfaitement conservés parmi les séquences présentées. Les 13 pre-miARN de la famille let-7 chez humain (hsa) sont comparés au pre-let-7g de la souris (mmu). Les séquences provenant de miRBase (251) ont été alignées avec T-coffee (252).

1.4.1 Rôle physiologique

Le rôle de let-7 dans la régulation temporelle du développement chez *C. elegans* a déjà été décrit (voir section 1.3.1). Cette fonction est bien conservée au travers les espèces. Par exemple, chez *D. melanogaster*, let-7 est considéré comme un gène hétérochronique pour son contrôle de la prolifération cellulaire dans le développement des ailes durant la métamorphose (253,254). De façon plus générale, le rôle de let-7 est de promouvoir la différenciation cellulaire.

Dans les cellules souches embryonnaires (cellules ES), let-7 est essentiellement absent alors que le polycistron miR-290-295 compte pour 60 à 70% de tous les miARN exprimés (255-257). Les miARN de ce polycistron permettent la transition rapide de la phase G1 à la phase S du cycle cellulaire en ciblant p21 et LATS2, des inhibiteurs de la *Cyclin-dependent*

Kinase 2 [Cdk2 (258)]. Ces miARN ciblent également d'autres régulateurs du cycle cellulaire tels que Wee1 et Fbx15, ce qui permet au polycistron d'induire la prolifération et d'empêcher la différenciation cellulaire (259). Par un mécanisme indirect encore mal compris, miR-290-295 maintient l'expression de Lin28 dans les cellules ES (260). Ce polycistron agit donc comme un inhibiteur indirect pour la maturation de la famille let-7. Lors de la différenciation, l'expression de miR-290-295 diminue suite à la répression de facteurs de transcription associés à la pluripotence tels que Nanog, Sox2 et Oct4 (257,260). Ceci permet l'initiation de l'expression du miARN let-7. Cette production augmente avec le processus de différenciation, puisque la production de let-7 fait partie de doubles boucles de rétroaction négatives avec Lin28 et Myc. Les membres de la famille let-7 ont plusieurs cibles permettant d'inhiber la prolifération et d'induire la différenciation cellulaire. Par exemple, let-7 réprime l'expression de Cdk4, Cdk6, *Cell Division Cycle 25 Homolog A* (CDC25A) et de Cyclin D1/2, ce qui ralentit la transition G1-S (242,261). La présence de let-7 inhibe également l'expression de Myc, Sal4 et Lin28, des facteurs associés à la pluripotence cellulaire (260,262). Bref, chez les mammifères tout comme chez *C. elegans*, l'expression de let-7 est essentielle à l'arrêt de la prolifération et permet la différenciation cellulaire.

1.4.2 Suppresseur de tumeur

L'activité supprimeuse de tumeurs de let-7 a déjà été brièvement abordée (voir section 1.3.2). Par sa capacité à cibler les oncogènes Ras, Myc et HMGA2, let-7 limite la prolifération des cellules différenciées (17,234,235). Plus précisément, le miARN let-7 peut inhiber l'expression des trois gènes Ras (K-Ras, N-Ras and H-Ras) que possèdent les humains (234). Une dérégulation de ces gènes étant fortement associée au cancer du poumon (263,264), on peut considérer let-7 comme étant un gène régulateur de ce type de cancer (173,236,265). Le facteur de transcription Myc est un autre oncogène ciblé par let-7 (17). Myc et let-7 font partie d'une double boucle de rétroaction négative où le miARN inhibe l'expression de l'ARNm de Myc et où ce facteur de transcription bloque la production de pri-let-7a-1/f-1/d/g (266). De plus, let-7 inhibe la traduction d'IMP1, une protéine protégeant certains ARNm de la dégradation, et déstabilise ainsi le transcrit de Myc (267). Par ce système de régulation

complexe, une faible réduction de l'expression d'un des deux facteurs peut être rapidement amplifiée. Finalement, let-7 cible également HMGA2 une protéine impliquée dans le renouvellement (*self renewal*) des cellules souches (235,268). Par ailleurs, lorsque l'expression de HMGA2 est mal régulée dans certains cas de cancer, l'extrémité 3'-UTR de l'ARNm est fréquemment écourtée. Ces ARNm plus courts ne sont pas inhibés par let-7, permettant ainsi l'expression de la protéine oncogène (269).

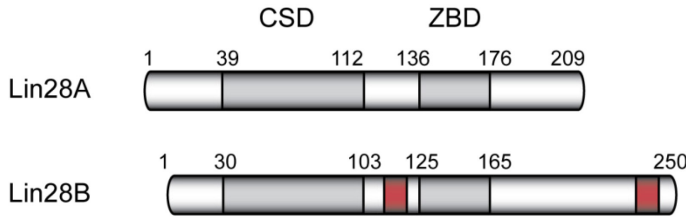
Il est maintenant reconnu que la faible expression de let-7 mature est un marqueur de cancers avancés pour lesquels le taux de survie est faible (270-272). Cette valeur pronostique de let-7 s'accorde avec le modèle des cellules initiateuses du cancer (*tumor initiating cells*, TIC). Dans ce modèle, un tissu cancéreux est composé de trois types cellulaires : des cellules cancéreuses différenciées, des cellules se divisant rapidement et de rares cellules initiateuses du cancer se divisant lentement et ayant des caractéristiques de cellules souches. Les TIC ont la capacité de se diviser à l'infini, peuvent se différencier en plusieurs types cellulaires, sont responsables des rechutes après les traitements et peuvent se métastaser (273,274). Il a été démontré que l'absence de let-7 mature est nécessaire pour que les TIC possèdent ces caractéristiques propres aux cellules souches (271). La faible expression de let-7 mature est donc un indicatif de la présence de TIC et, par conséquent, de l'agressivité de la tumeur.

1.5 Protéine Lin28

La protéine Lin28 joue un rôle majeur dans la régulation temporelle de *C. elegans* où elle bloque l'expression des miARN de la famille let-7 en empêchant ainsi le passage L2/L3 (voir Figure 1-12). Au stade de maturation L2, la traduction de Lin28 est inhibée par le miARN lin-4 permettant ainsi la production de let-7 [voir Figure 1-1, (3)]. Ce miARN étant impliqué dans une double boucle de rétroaction négative avec Lin28, l'expression de la protéine cesse complètement à partir de ce stade de maturation (4). L'expression de cette protéine chez les mammifères est également temporellement contrôlée par le miARN miR-125 (un homologue de lin-4) et let-7 (275,276). En fait, Lin28 est présente et bien conservée chez plusieurs espèces incluant les mammifères, les insectes, les poissons, les oiseaux et les reptiles (voir Figure 1-14B). Dans tous ces cas, l'expression de la protéine varie avec le développement et est absente des cellules différenciées (276,277).

Sachant que Lin28 inhibe la maturation des miARN let-7, il est peu surprenant que cette protéine soit associée à la prolifération et à la différenciation cellulaire. Par exemple, cette protéine a récemment été identifiée comme l'un des quatre facteurs (Lin28, Oct4, Sox2 et Nanog) nécessaires pour reprogrammer une cellule somatique en cellule souche pluripotente (278). Lin28 est également associée à certains attributs physiques tels que la taille et à la régulation de la puberté chez l'homme et la souris (279-281). Comme d'autres facteurs de pluripotence, cette protéine est requise pour le développement normal des organismes (277,282), tandis qu'une mauvaise régulation de cette dernière est liée au cancer. En effet, l'expression de Lin28 est activée dans 15% des tumeurs (283), corrèle avec un faible taux de survie des patients (284) et augmente la susceptibilité de certains cancers (285,286). Toutes ces fonctions de Lin28 pourraient être directement liées à sa capacité d'inhiber la maturation de let-7 [voir détails de cette inhibition dans la section 1.6 (167,287)]. Pourtant, Lin28 possède également certaines fonctions indépendantes de let-7 qui lui permettent de stimuler directement la traduction de certains ARNm (voir section 1.5.2). Par ce double rôle, la protéine Lin28 est un régulateur très puissant de la croissance et de la prolifération cellulaire.

A



B

	RNP-1			RNP-2		
hsa	40-GAGICKWFNV	RMGFGFLSMT	ARAGVALDPP	VDVFNHQS	HMEGFRSLKE	GEAVEFTFKK
mmu	40-GAGICKWFNV	RMGFGFLSMT	ARAGVALDPP	VDVFNHQS	HMEGFRSLKE	GEAVEFTFKK
gga	34-GSGICKWFNV	RMGFGFLSMT	AKGGAMLDSP	VDVFNHQS	HMEGFRSLKE	GEAVEFTFKK
dre	34-GSGVCKWFNV	RMGFGFLSMT	HREGICLDSP	VDVFNHQS	HMEGFRSLKE	GEAVEFTFKR
xla	34-GSGVCKWFNV	RMGFGFLTMT	KKEGTDLETP	LDVFNHQS	HMEGFRSLKE	GESVEFTFKK
	*.:*****	*****:.*	: * *:	:*****	*****	*:*****:
				ZBD-1		
hsa	100-SAKGLESIRV	TGPGGVFCIG	SERRPKGKSM	QKRRSKGDR	YNCGGLDHHA	KECKLPPQPK
mmu	100-SAKGLESIRV	TGPGGVFCIG	SERRPKGKSM	QKRRSKGDR	YNCGGLDHHA	KECKLPPQPK
gga	94-SSKLESIRV	TGPGGVFCIG	SERRPKSKSL	QKRRSKGDR	YNCGGLDHHA	KECKLPPQPK
dre	94-SSKLESIQV	TGPGGAPCVG	SEKKPKG--T	QKRRSKGDR	FNCGGPNHHA	KECQLPPQPK
xla	94-SSKLESTQV	TGPGGAPCIG	SERRPKVKGQ	QKRRRQGDRC	YNCGGLDHHA	KECKLPPQPK
	*:*****	.:****	*:.*	**::**	****.:****	:**** :*** ***:*****
				ZBD-2		
hsa	160-KCHFCQSISH	MVASCPLKAQ	Q-180			
mmu	160-KCHFCQSINH	MVASCPLKAQ	Q-180			
gga	154-KCHFCQSISH	MVANCPAKAQ	Q-174			
dre	152-KCHFCQSISH	MVANCPIKAQ	Q-172			
xla	154-KCHFCQNPNH	MVAQCPEKAM	Q-174			
	*****.	.:***	.** ** *			

Figure 1-14. Conservation des protéines Lin28A et Lin28B chez les vertébrés (A) Les protéines humaines Lin28A et Lin28B possèdent un domaine *cold shock* (CSD) et deux domaines de liaison au zinc rapprochés (ZBD). Les séquences de ces domaines sont très bien conservées entre les deux paralogues. Les séquences de localisation au noyau de Lin28B sont identifiées en rouge. (B) Les séquences de Lin28A provenant de *Homo sapiens* (hsa), *Mus musculus* (mmu), *Gallus gallus* (gga), *Danio rerio* (dre) et *Xenopus laevis* (xla), ont été alignées en utilisant Clustal Omega (288). Le positionnement des motifs RNP et ZBD est indiqué à l'aide de lignes noires, les acides aminés coordonnant les atomes de zinc des ZBD sont en rouge, un astérisque indique une conservation parfaite, le deux-points et le point sont respectivement signes d'une forte ou d'une faible similarité.

1.5.1 Rôles des différents domaines de Lin28

Lin28 est une petite protéine d'environ 23 kDa composée de deux régions principales. L'extrémité N-terminale de la protéine contient un domaine *cold shock* (CSD), alors que l'on retrouve deux domaines de liaison au zinc (ZBD) à l'extrémité C-terminale [voir Figure 1-14A (3,276)]. Chez les animaux, Lin28 est l'unique protéine à posséder cette combinaison de domaines. Les seules autres protéines connues avec cette rare combinaison sont les protéines CSDP1 et CSDP2 retrouvées chez les plantes (289). Ces protéines végétales sont des chaperonnes permettant l'adaptation au froid d'*Arabidopsis thaliana* par la liaison et la déstabilisation de certains ARNm (290).

Plusieurs organismes possèdent un paralogue à Lin28. Lin28B a une taille similaire à Lin28 (parfois identifié comme Lin28A pour clarification) et possède également un CSD en plus de deux ZBD [voir Figure 1-14A, (291)]. Les séquences des deux homologues sont globalement similaires et sont pratiquement identiques à l'intérieur des domaines *cold shock* et de liaison au zinc. Ces similitudes expliquent sans doute pourquoi ces deux protéines ont la capacité d'inhiber la maturation des miARN let-7. Aussi, l'ARNm de Lin28B possède plusieurs sites de liaison pour let-7 sur son 3'-UTR (291). Ce paralogue est donc impliqué dans le même type de double boucle de rétroaction négative que Lin28A. Ce qui démarque Lin28B de son paralogue, est la présence de deux signaux de localisation au noyau (NLS) (292,293). Ces deux NLS, localisés au milieu et à l'extrémité C-terminale de la protéine (voir Figure 1-14A), permettent à Lin28B de s'accumuler dans le noyau. Cette caractéristique permet à la protéine d'inhiber le clivage des pri-miARN par Drosha, alors que Lin28(A) agit dans le cytoplasme (voir détails dans la section 1.6.1).

Le domaine *cold shock* de Lin28 est homologue à celui présent chez les protéines Y-box des vertébrés impliquées dans la régulation de la traduction (294,295), mais ressemble davantage aux CSD des protéines *cold shock* bactériennes [Csps, (3)]. Les Csps interagissent avec l'ARNm, ont la capacité d'en réguler la traduction et possèdent une faible spécificité de séquence (296). Par exemple, la chaperonne à ARN CspA peut coopérativement lier un ARN simple brin de séquence variable et déstabiliser les structures secondaires voisines (297). Ce type de domaine contient deux motifs de liaison à l'ARN (RNP-1 et RNP-2) composés d'acides aminés aromatiques (298). Chez Lin28, les deux motifs RNP sont parfaitement

conservés au travers les espèces (Figure 1-14B). La pression sélective sur ces motifs est expliquée par l'importance du CSD dans les fonctions de Lin28. En effet, la mutation d'acides aminés conservés dans ce domaine est associée à des problèmes de régulation temporelle du développement (3). Également, l'expression d'un isoforme de Lin28B ne possédant pas un domaine *cold shock* complet (Lin28BS), ne stimule pas la croissance contrairement à Lin28A et Lin28B (291). La présence du CSD est donc vitale au bon fonctionnement de la protéine.

L'extrémité C-terminale de Lin28 est composée de deux domaines de liaison au zinc de type CCHC (3). Ce type de domaine est habituellement retrouvé dans les protéines nucléocapsides de rétrovirus comme le VIH où ils lient spécifiquement des guanines exposées (299-303). Ces deux domaines ainsi qu'une région riche en lysines et arginines forment un plus grand ensemble nommé le *NCp7-like domain* [voir section 2.0 pour plus de détails, (304)]. Le *NCp7-like domain* de Lin28 est essentiel pour la localisation de la protéine aux *P-bodies* (292), des sites de dégradation de l'ARNm et de régulation des miARN (305,306). Par contre, la présence des deux domaines de liaison à l'ARN de Lin28 est nécessaire pour une forte localisation aux polysomes (292).

1.5.2 Liaison aux ARNm

Cette localisation de Lin28 aux polysomes (292) est la première évidence que Lin28 possède une activité indépendante de let-7. En effet, cette protéine s'associant aux formes immatures du miARN, cette présence aux polysomes est probablement liée à la traduction des ARNm. La protéine Lin28 possède effectivement la capacité de lier directement les ARNm matures en cours de traduction (307,308). Cette association permet de stimuler l'expression d'un vaste groupe d'ARN messagers. Lin28 a d'abord été impliquée dans l'expression de l'*insulin-like growth factor* (IGF-2) lors de la différenciation des cellules musculaires (307). Alors, il a été proposé que Lin28 stimulait la traduction grâce à son interaction avec le facteur d'initiation de la traduction eIF3 β . Jusqu'à aujourd'hui, l'existence et le rôle de cette interaction entre cette protéine et Lin28 n'ont pas été confirmés (309). Plusieurs autres ARNm cibles de Lin28 ont par la suite été identifiés. Il est maintenant connu que Lin28 peut stimuler la traduction d'ARNm associés à la régulation du cycle cellulaire

[CyclinA, Cyclin B, Cdk4 (310)], à la croissance [Histone H2a, Oct4 (311,312)] ou à l'épissage alternatif [hnRNP, TIA-1, FUS/TLS et TDP-43 (313)]. Dans la majorité de ces cas, une région de quelques centaines de nucléotides a été identifiée comme étant responsable de la stimulation par Lin28. Cet élément sensible à Lin28 (*Lin28-responsive element*, LRE) peut être présent dans la région codante ou dans le 3'-UTR de l'ARNm cible. Finalement, des études génomiques ont identifié des milliers d'autres transcrits pouvant être liés par Lin28 (314). Il est maintenant estimé qu'entre 5 et 25% des ARNm sont des cibles potentielles de Lin28 (313,315). Puisque cette activité de Lin28 contrôle divers facteurs de croissance et différents régulateurs du cycle cellulaire, il est peu surprenant que cette stimulation de la traduction par Lin28 ait été impliquée dans différents cancers (316).

L'un des rôles de Lin28 dans la régulation de la traduction est de servir d'ancrage sur l'ARN à la *RNA Helicase A* (RHA). La RHA est une enzyme pouvant améliorer la traduction de certains ARNm dont le 5'-UTR contient des structures secondaires nuisant au déplacement des ribosomes lors de l'initiation de la traduction (317,318). Les raisons pour lesquelles le recrutement par Lin28 d'une telle enzyme dans la région codante ou dans le 3'-UTR permet la stimulation de la traduction sont encore inconnues. Il a été proposé que la RHA pourrait modifier des éléments de structure habituellement reconnus par des protéines inhibitrices de la traduction (309,319), mais cette hypothèse reste à vérifier. Pour servir d'ancrage à la RHA sur l'ARNm, Lin28 doit évidemment posséder ses deux domaines de liaison à l'ARN afin de faire contact avec la cible. De plus, les 35 derniers acides aminés de l'extrémité C-terminale de la protéine sont essentiels afin de former un complexe avec la RHA. Inversement, la RHA utilise chacune de ses extrémités pour faire contact avec Lin28. Ces expériences provenant d'essais Co-IP et de *GST pull-down* (319), une cartographie précise de l'interaction reste à être déterminée. Par ailleurs, on ignore si la RHA reste en complexe avec Lin28 lors de la déstabilisation des structures secondaires par l'hélicase.

En plus de sa capacité à stimuler la traduction de certains facteurs de croissance et de différents régulateurs du cycle cellulaire, une récente publication indique que Lin28 est responsable de la répression de la traduction d'un sous-groupe d'ARNm (320). En effet, la liaison de Lin28 bloquerait la traduction de certains ARNm épissés dans les cellules souches embryonnaires. Cette inhibition de la traduction est spécifique aux protéines destinées au

réticulum endoplasmique (RE). La présence de Lin28 diminue donc la quantité de protéines dans le RE et dans l'appareil de Golgi en plus de réduire le nombre de protéines transmembranaires, ainsi que la quantité de protéines sécrétées. Le mécanisme permettant cette répression par Lin28 est inconnu, mais débiterait encore une fois par la liaison directe de Lin28 à ces ARNm.

Plusieurs groupes se sont intéressés aux motifs ou aux structures reconnues par Lin28 sur les ARNm. À partir des différents LRE connus, un motif contenant un renflement (*bulge*) d'une adénosine possédant deux paires de base GC comme voisines a été identifié (321). Ce motif contenu dans une tige stable est surprenant, puisque les deux domaines de liaison à l'ARN de Lin28 sont connus pour lier l'ARN simple brin. En utilisant des expériences de *crosslinking* et d'immunoprécipitation (CLIP) de courtes séquences présentes sur les ARNm liées par Lin28 ont été mises en évidence. D'abord, Wilbert *et al.* (313) ont avancé que la séquence consensus 5'-GGAGA-3' est nécessaire pour la stimulation de la traduction par Lin28. Ensuite, Cho *et al.* (320) ont identifié les séquences 5'-AAGNNG-3', 5'-AAGNG-3' et 5'-UGUG-3' comme étant importantes pour la répression de l'expression des protéines destinées au ER. De plus, les séquences de Wilbert et Cho doivent être présentes dans une boucle, ce qui va à l'encontre du modèle proposé précédemment. Puis, Hafner *et al.* (314) ont conclu que Lin28 reconnaît des séquences d'ARN simple brin riches en uridines et ayant quelques guanosines comme voisinage immédiat. Pour terminer, Graf *et al.* (322) ont proposé la séquence consensus RGGSWG (R = G ou A, S = G ou C et W= A ou T) pour la reconnaissance d'ARN par Lin28. Bref, la séquence consensus permettant à Lin28 de reconnaître spécifiquement ses ARNm cibles semble difficile à définir. Par contre, les séquences 5'-GxG-3' ou 5'-GxxG-3' sont celles qui ressortent le plus souvent lors d'expériences de CLIP.

1.6 Inhibition de la biogenèse de let-7 par Lin28

Quoique la protéine Lin28 exerce un contrôle direct sur un groupe d'ARNm, son rôle le mieux caractérisé jusqu'à présent est celui du contrôle de la biogenèse des miARN de la famille let-7. Il est connu, depuis près de dix ans, que les miARN let-7 possèdent un mode de régulation post-transcriptionnelle. En effet, si l'expression de let-7 mature est variable au cours du développement (voir section 1.4), la production de pri-let-7 reste constante (244-246). De plus, une augmentation de la production de pri-let-7 n'est pas nécessairement accompagnée d'une hausse de let-7 mature (247) ce qui est un indicateur de la présence d'un contrôle post-transcriptionnel.

En 2008, plusieurs groupes ont simultanément identifié la protéine Lin28 comme l'inhibiteur spécifique à la maturation des miARN de la famille let-7 (199,245,248). La présence de Lin28 dans les cellules non différenciées et dans certaines cellules cancéreuses permet effectivement d'expliquer l'inhibition de la biogenèse de let-7 dans ces types cellulaires. En effet, la surexpression de Lin28A ou de Lin28B dans des cellules différenciées est suffisante pour inhiber la maturation de let-7. Également, l'utilisation d'interférence à l'ARN (*RNAi mediated knockdown*) contre Lin28 permet d'augmenter spécifiquement la concentration de let-7 mature dans des cellules non différenciées ou cancéreuses (199,248,250,287). De plus, des études *in vitro* ont démontré que l'extrait nucléaire de cellules P19, dérivées d'un carcinome embryonnaire, est suffisant pour bloquer le clivage de pri-let-7g par Drosha. Cette inhibition est spécifique, puisque l'extrait nucléaire de P19 est sans effet pour pri-miR-17 (248). Finalement, un contact direct entre Lin28 et la boucle terminale de pri-let-7 permet de confirmer que la liaison de cette protéine aux formes immatures de let-7 est responsable de l'inhibition de la maturation de ces miARN (248).

1.6.1 Modèles de régulation de pri-let-7 par Lin28

Différents mécanismes permettant à Lin28 d'inhiber la biogenèse de let-7 ont été proposés dans les dernières années (voir Figure 1-15). Le consensus actuel est que Lin28 possède plusieurs moyens pour mener à bien une inhibition efficace de la biogenèse de let-7.

Par conséquent, les différents mécanismes identifiés jusqu'à présent agissent sans doute en parallèle de façon à maintenir une forte répression de la biogenèse de cette famille de miARN.

Puisque les niveaux de pri-let-7 restent constants au cours du développement, il a d'abord été proposé que Lin28 permet l'inhibition du clivage par Drosha dans le noyau (voir Figure 1-15). L'une des premières études sur ce sujet a d'ailleurs démontré que la présence de Lin28 est nécessaire et suffisante pour bloquer le clivage de pri-let-7g par le complexe *Microprocessor* contenant l'enzyme Drosha (245). Récemment, il a été précisé que la protéine Lin28B était responsable de cette inhibition au noyau (293). Avec ses deux signaux de localisation nucléaire (voir section 1.5 et Figure 1-14A), Lin28B a la capacité de lier pri-let-7g dans le noyau et de localiser cet ARN dans le nucléole. L'endonucléase Drosha étant absente du nucléole, l'inhibition de la maturation par Lin28B s'effectuerait donc par séquestration plutôt que par encombrement stérique tel qu'il était initialement suggéré (293). Il est maintenant proposé que tous les pri-let-7 subissent ce type de répression par Lin28B. Il est à noter que seuls les deux signaux de localisation nucléaire différencient Lin28B de Lin28(A). En effet, le mode de reconnaissance de l'ARN par ces deux protéines est prédit pour être identique (323), c'est pourquoi le reste de cette thèse se concentrera principalement sur Lin28.

Dans le cytoplasme, Lin28 inhibe la biogenèse de pre-let-7 à l'aide de plusieurs mécanismes différents. En liant le miARN précurseur de let-7g (293), Lin28 empêche la liaison de Dicer à ce pre-miARN [voir Figure 1-15, (250)]. Structuellement, la liaison de Lin28 à pre-let-7g cause un changement de conformation dans les sites de clivage de Dicer. Ces sites de clivage, correspondant aux liaisons phosphodiester localisées entre les nucléotides U6-U7 et U41-C42 de la représentation présente à la Figure 1-16B (251), sont positionnés à l'extrémité de la tige du pre-miARN. Ce sont les quelques paires de base composant le site de clivage qui se dissocient suite à la liaison de Lin28 (324). C'est ce changement de conformation de la tige causé par liaison de Lin28 ou l'encombrement stérique provoqué par la protéine qui inhibe l'activité de Dicer.

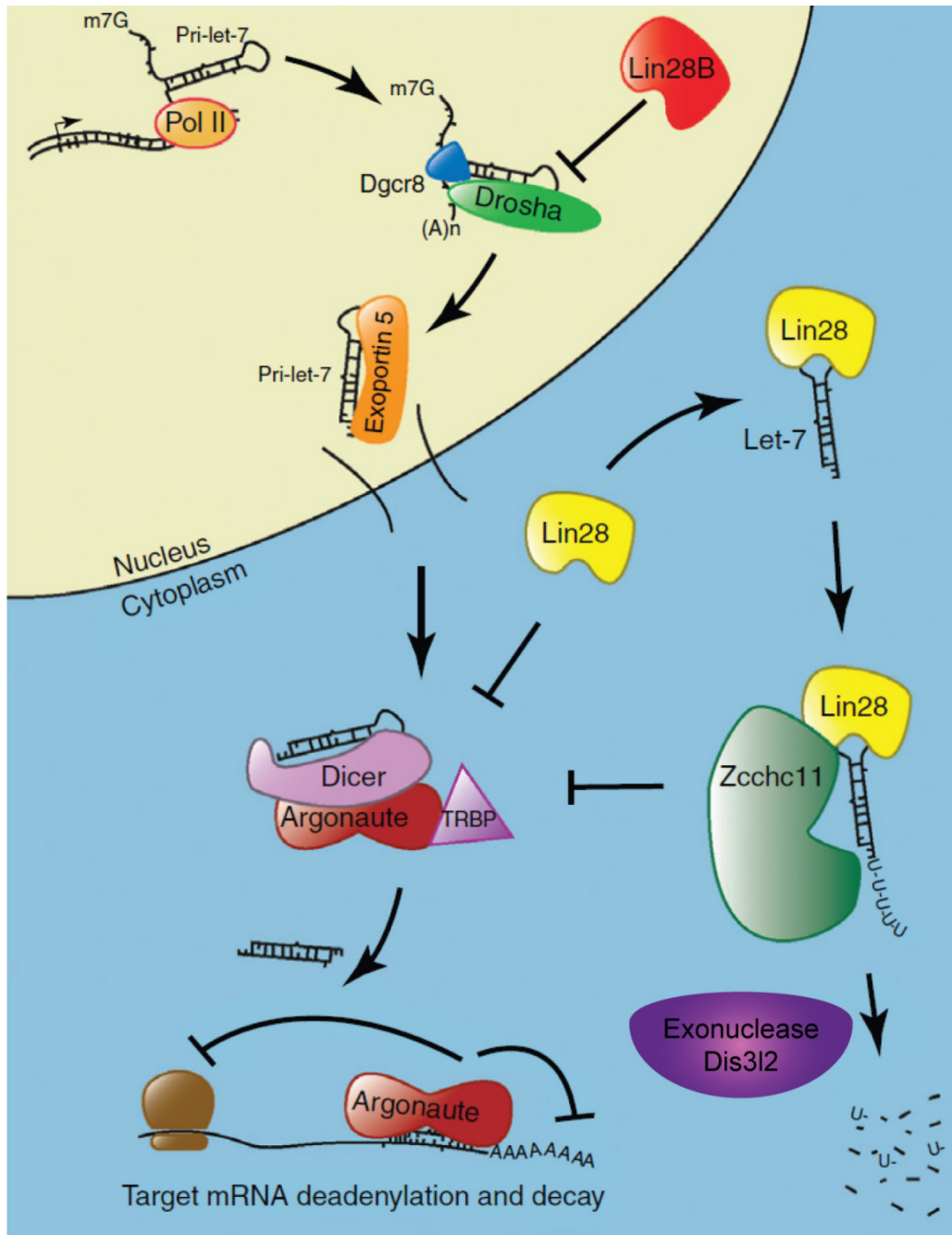


Figure 1-15. Régulation de la biogenèse du miARN let-7 par Lin28. Dans le noyau, la protéine Lin28B permet l'inhibition du clivage de pri-let-7 par Drosha. La protéine Lin28 (Lin28A) dans le cytoplasme inhibe directement le clivage par Dicer en plus de permettre le recrutement du Zcchc11 (Tut4). Cette enzyme poly-uridyle l'extrémité 3' de pre-let-7 ce qui empêche son clivage par Dicer et induit sa dégradation par l'exonucléase Dis3L2. Adaptée de (167).

La liaison de Lin28 au pre-miARN induit également le recrutement de l'enzyme Tut4 (aussi appelée Zcchc11) qui ajoute une queue d'environ 14 uridines (queue poly-U) à l'extrémité 3' du miARN précurseur (199). Contrairement à la mono-uridylation décrite précédemment (voir section 1.2.2), cette poly-uridylation est dépendante de la présence de Lin28. En effet, la formation du complexe Lin28/pre-let-7 est requise pour la poly-uridylation par Tut4 dans le cytoplasme (200). Les éléments structuraux permettant au complexe Lin28/pre-let-7 d'être reconnu par la tutase restent à définir. Il est possible que la liaison de Lin28 modifie la structure de pre-let-7, permettant ainsi sa reconnaissance par Tut4. Le scénario inverse où la formation du complexe altérerait la structure de Lin28, rendant ainsi possible une interaction Lin28/Tut4, est également envisageable. Pour l'instant, il est seulement connu que *NCp7-like domain* de Lin28 est essentiel à l'ajout de cette queue de poly-U (199). La modification de l'extrémité 3' du miARN précurseur a deux effets sur la biogenèse de ce dernier. D'abord, une telle extrémité simple brin poly-uridylée est incompatible avec le mode de reconnaissance de Dicer (voir section 1.1.4) ce qui bloque la biogenèse du pre-miARN (199). Ensuite, la queue poly-U agit comme un signal de dégradation (199) en recrutant une exonucléase 3'→5' nommée Dis3L2 (204). L'activité de cette nucléase est spécifique à la présence de la queue de poly-U sur les précurseurs de la famille let-7 et est indépendante de Lin28. Une queue de 10 U est minimale pour détecter une activité nucléase, mais 14 U est la longueur optimale (204). Dis3L2 est spécifique à la présence de ces uridines, puisque l'enzyme est active si une queue de 14 U sont ajoutées à l'extrémité 3' de miR-21 (204). La poly-uridylation par Tut4 entraîne donc la dégradation du pre-miARN en plus de bloquer son clivage par Dicer (voir Figure 1-15).

Le recrutement de cette tutase par Lin28 a donc un fort pouvoir régulateur sur la biogenèse de let-7. En effet, l'inhibition de Tut4 augmente les niveaux de let-7 mature sans affecter ceux de miR-16 ou pri-let-7 (200), ce qui induit la répression des cibles de let-7 (201). La tutase Tut4 est une grande protéine de 184 kDa de la famille des poly(A) polymérases non canoniques. L'extrémité N-terminale de la protéine est composée d'un domaine de fonction inconnue (pneumo-G) suivi d'un domaine de liaison au zinc de type C2H2 et d'un domaine catalytique *Nucleotidyl Transferase* (Ntr) inactif, similaire aux protéines de levure TRF4, lié à un domaine associé aux poly(A) polymérases (PAP). Le reste de la protéine est composé d'un

domaine de liaison au zinc de type CCHC, du domaine catalytique Ntr, d'un deuxième PAP, de deux autres domaines CCHC et finalement d'un domaine de fonction inconnue similaire à la protéine atrophine-1 à l'extrémité C-terminale. À l'exception des deux domaines de fonctions inconnues et du dernier ZBD, tous les domaines de Tut4 sont requis pour la polyuridylation des pre-let-7 (203). Également, le motif de liaison au zinc C2H2 à l'extrémité N-terminale est essentiel pour observer un effet stimulateur par Lin28. Puisque ce type de domaine est connu pour lier l'ADN, l'ARN et les protéines, cette fonction stimulatrice ne permet pas de proposer un mécanisme moléculaire qui expliquerait la reconnaissance du complexe Lin28/pre-let-7 par Tut4. La tutase Tut7 (Zcchc6) peut également ajouter une queue de poly-U aux pre-let-7 en absence de Tut4 (203). Ce mécanisme alternatif s'explique par la ressemblance entre ces deux enzymes. En effet, Tut7 est la seule autre tutase possédant un ZBD de type C2H2 à son extrémité N-terminale (166,203). Afin de comprendre ce qui permet à Tut4/7 de reconnaître spécifiquement le complexe Lin28/pre-let-7, il devient important de bien cartographier les interactions moléculaires entre Lin28, pre-let-7 et Tut4/7 pour en comprendre le fonctionnement.

1.6.2 Cartographie de l'interaction entre pre-let-7 et Lin28

La liaison de Lin28 aux formes immatures de let-7 étant l'élément déclencheur de plusieurs mécanismes responsables de l'inhibition de la biogenèse de ce miARN, plusieurs études se sont intéressées au mode reconnaissance de Lin28 envers ces ARN. La boucle terminale (TL) commune au transcrit primaire et au miARN précurseur contient plusieurs éléments importants pour la liaison de Lin28. En effet, cette boucle terminale est suffisante pour la liaison de Lin28 et la protéine possède une affinité similaire envers TL-let-7g et pre-let-7g soit un K_d d'environ 2 μ M lorsque des ARN compétiteurs sont utilisés (249,324). Par contre, Lin28 est incapable de lier un duplex d'ARN formé par le miARN mature et son brin complémentaire (249). Encore une fois, Lin28B présente des caractéristiques semblables, puisqu'un ARN de 24 nt correspondant à la séquence des nucléotides 14 à 37 de TL-let-7g (voir Figure 1-16B) est suffisant pour lier ce paralogue (323).

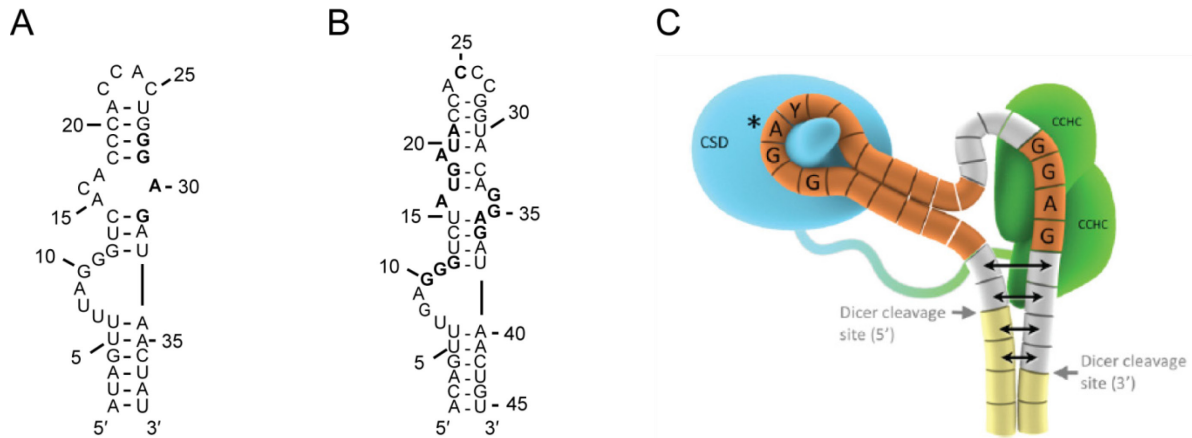


Figure 1-16. Sites de liaison de Lin28 présents sur TL-let-7a-1 et TL-let-7g. Structures secondaires des boucles terminales de pre-let-7a-1 (A) et de pre-let-7g (B) trouvées chez l'humain. Les différents sites de liaison de Lin28 identifiés sur ces deux ARN sont mis en évidence à l'aide de caractères gras. (C) Modèle de l'interaction de Lin28 avec la boucle terminale d'un miARN précurseur de la famille let-7. L'ARN simplifié utilisé pour l'obtention de la structure cristalline est seulement composé des nucléotides en orange. Les nucléotides gris et jaunes ont été ajoutés à l'illustration pour compléter le modèle (325).

Puisque tous les membres de la famille let-7 sont inhibés par Lin28 et que les boucles terminales de ces derniers n'ont pas de rôles dans la répression de la traduction par les miARN, il est probable que les motifs communs à ces TL soient importants dans l'inhibition par Lin28. En utilisant des alignements de séquences (voir Figure 1-13), plusieurs régions conservées chez la famille let-7 ont été identifiées. Une cytosine conservée (C25) présente dans la boucle apicale de quatre nucléotides de TL-let-7g (voir Figure 1-16B) a d'abord été identifiée (249). L'importance de ce nucléotide pour la liaison par Lin28 est controversée. En effet, si la mutation de cette cytosine dans l'étude originale a provoqué une diminution de l'affinité d'environ 20X (249), la même mutation a seulement changé le K_d d'un facteur de quatre lors de la reproduction de l'expérience par une autre équipe (326). Trois autres nucléotides conservés de TL-let-7g (A19, G34 et G35 voir Figure 1-16B) ont été identifiés comme importants pour l'inhibition de la biogenèse de let-7 par Lin28. En effet, la mutation de ces trois nucléotides de la boucle terminale permet à pre-let-7g d'être insensible à

l'inhibition par Lin28 lors d'essais avec des extraits nucléaires de cellules P19 (248). Puisque ce type d'essai ne permet pas d'observer la liaison directe entre la protéine et le miARN précurseur, il est impossible d'affirmer que ces mutations empêchent la liaison de la protéine. Leurs effets pourraient effectivement être le fruit de cofacteurs n'ayant pas été identifiés.

L'importance de la région riche en guanine en 3' a également été observée chez TL-let-7a-1. Tout comme TL-let-7g, cette boucle terminale possède la séquence bien conservée 5'-GGAG-3' (voir Figure 1-16A) et cette dernière est importante pour la liaison et l'inhibition par Lin28. En effet, l'introduction des mutations 5'-GuAu-3' empêche la liaison de Lin28 à TL-let-7a-1 et abolit complètement la poly-uridylation par Tut4. Toutefois, certaines mutations préservant la dernière guanine (5'-GagG-3' et 5'-aaAG-3') sont compatibles avec la liaison de Lin28, mais bloquent la poly-uridylation (200). Ce résultat met en évidence la complexité de l'assemblage; la liaison de Lin28 au pre-miARN est nécessaire sans toutefois être suffisante pour observer une poly-uridylation. Puisque la séquence 5'-GGAG-3' retrouvée chez TL-let-7g (voir Figure 1-16B) est aussi nécessaire pour sa liaison par Lin28B (323), il est possible d'affirmer que cette séquence conservée est cruciale pour la régulation de la famille let-7 par Lin28. Finalement, des essais de protection aux ribonucléases ont permis d'identifier le motif 5'-GGG-3' (voir Figure 1-16B) localisé dans le renflement en 5' de TL-let-7g comme un site de liaison de Lin28 (324). Cette même étude a également mis en évidence plusieurs nucléotides dont la sensibilité aux nucléases diminue lors de l'ajout de Lin28. Ces derniers sont principalement situés entre les positions 15 et 22 (voir Figure 1-16B) et sont indicatifs d'une protection directe par Lin28 ou d'un changement de conformation de l'ARN suite à la liaison de la protéine.

La multiplicité des séquences d'ARN qualifiées d'importantes pour la liaison de Lin28 s'explique peut-être par la présence de deux domaines de liaison à l'ARN sur la protéine. Il est en effet probable que chacun d'entre eux reconnaissent différentes séquences et/ou éléments structuraux présents sur les miARN immatures. Selon les méthodes utilisées, les deux domaines de liaison à l'ARN de Lin28 sont généralement essentiels pour former un complexe avec les boucles terminales de let-7 et en inhiber la biogenèse (293). Dans d'autres cas, le retrait d'un des deux domaines de liaison à l'ARN réduit seulement l'affinité de la protéine Lin28 envers pre-let-7 (325). De plus, il est possible de retirer les extrémités N-terminale et C-

terminale de Lin28 ne contenant pas de domaines de liaison à l'ARN sans modifier le K_d du complexe (325).

L'introduction de mutations dans le *NCp7-like domain* réduit significativement l'affinité de la protéine envers TL-let-7a-1 (200) et TL-let-7g (326). De plus, une telle protéine mutée est incapable de stimuler la poly-uridylation d'un pre-miARN en présence de Tut4 (200). Ces caractéristiques ressemblent aux effets des mutations introduites dans la séquence conservée 5'-GGAG-3' et pourraient indiquer que cette séquence est le site de liaison du *NCp7-like domain*. En effet, les domaines de liaison au zinc de type CCHC sont connus pour lier des ARN simple brin riches en guanines (299-303). Récemment, une étude de spécificité utilisant un fragment de Lin28B provenant de *Xenopus tropicalis* (Xtr) a démontré que le ZBD a une affinité maximale pour les séquences 5'-GGAG-3' ou 5'-GGUG-3' (327). Conformément à ce qui a été observé lors de la liaison de Lin28 aux boucles terminales entières, la mutation de la dernière guanine de cette séquence bloque toute interaction (327).

Contrairement aux domaines de liaison au zinc, le domaine *cold shock* (CSD) de Lin28 ne semble pas avoir une forte spécificité de séquence. En effet, Mayr *et al.* (327) ont étudié l'interaction entre ce domaine et différents acides nucléiques. Ils ont conclu que le CSD de Lin28 peut lier différents sites sur Xtr-pre-let-7g avec une faible spécificité de séquence. Leurs études de spécificité ont tout de même permis de proposer 5'-GUNNUNN-3' et 5'-GTNNTNN-3' comme séquences consensus, puisque ce domaine lie à la fois l'ADN et l'ARN. Finalement, le CSD de Lin28 possède une activité chaperonne à l'ARN qui lui permet de modifier la structure d'une tige-boucle d'ARN ressemblant à TL-let-7g. Plus précisément, la liaison du CSD à cet ARN permet la séparation des brins situés à la base de cette structure ce qui éloigne les extrémités 5' et 3' de la tige-boucle d'ARN (327). En résumé, Lin28 possède deux domaines de liaison à l'ARN possédant des fonctions spécifiques. Le domaine CSD a une faible spécificité, mais permet le remodelage de l'ARN grâce à son activité chaperonne. À l'opposé, les domaines de liaison au zinc sont importants pour la poly-uridylation et possèdent une forte affinité pour les séquences consensus 5'-GNNG-3' ou 5'-GNG-3'.

1.6.3 Études structurales de Lin28

Récemment, trois études structurales ont permis de confirmer plusieurs données, obtenues à l'aide des études biochimiques décrites précédemment. D'abord, une structure résolue par RMN (326) démontre que le *NCp7-like domain* de Lin28 est suffisant pour former un complexe stable avec un court brin d'ARN. La séquence d'ARN utilisée (5'-AGGAGAU-3') provenant de TL-let-7g (voir Figure 1-16B) peut être liée par ce fragment de protéine avec un K_d de 0.4 μ M et correspond bien au consensus décrit dans la section 1.6.2. Néanmoins, Loughlin *et al.* (326) soulignent dans leur conclusion que tous les membres de la famille let-7 ne possèdent pas cette séquence consensus et que, chez plusieurs membres, cette région de l'ARN forme une double hélice. Puisque cette structure confirme que le *NCp7-like domain* de Lin28 lie l'ARN simple brin, ce site consensus bien conservé demeure inaccessible chez plusieurs pre-let-7.

En plus de la structure en solution du complexe *NCp7-like domain*/ARN, les structures cristallines des domaines *cold shock* isolés provenant d'*H. sapiens* et de *X. tropicalis* sont connues. Les domaines *cold shock* des deux organismes sont similaires; un tonneau de cinq brins bêta antiparallèles similaires aux protéines *cold shock* bactériennes. Un côté du domaine globulaire contient des acides aminés chargés négativement alors que l'autre est une plateforme hydrophobe constituée de plusieurs résidus aromatiques. (327). Dans la structure de Xtr-CSD en complexe avec un brin d'ADN composé de six thymines, l'interaction avec l'ADN se fait par l'empilement (*stacking*) des nucléobases avec les chaînes latérales aromatiques (327). Ce type de liaison explique le peu de spécificité observé précédemment.

Finalement, en utilisant une construction minimale de Lin28 et un ARN contenant quelques éléments structuraux retrouvés chez les TL-let-7, Nam *et al.* (325) ont obtenu une structure cristalline incluant les deux domaines de liaison à l'ARN de la protéine. La construction de Lin28 utilisée contient le CSD et le *NCp7-like domain* reliés par quelques acides aminés. De son côté, l'ARN est une tige-boucle à laquelle la séquence consensus 5'-GGAG-3' est fusionnée (voir nucléotides oranges à la Figure 1-16C). Dans cette structure le CSD lie une séquence consensus 5'-NGNGAYNNN-3' qui doit idéalement être fermée par une tige qui encercle le domaine *cold shock* (325). Ce domaine interagit avec l'ARN grâce à

l'empilement des chaînes latérales hydrophobes avec les nucléotides. Dans ce cas-ci, une certaine spécificité est obtenue grâce à des ponts hydrogènes et à de l'exclusion stérique qui favorise des pyrimidines à certaines positions. En accord avec l'étude précédente (326), le domaine *NCp7-like domain* lie la séquence 5'-GGAG-3' fusionnée à la tige-boucle. Cette région est localisée en 3' du site de liaison du CSD et l'interaction est spécifique au consensus 5'-NGNNG-3' (325).

En combinant ces différentes études, un modèle a été proposé où le domaine *cold shock* est le premier à faire contact avec l'ARN (327). Puis, grâce à son activité chaperonne, le CSD modifie la structure de la boucle terminale et rend le site de liaison du *NCp7-like domain* disponible. Puisque de la coopérativité a été observée lors de la liaison de Lin28 aux boucles terminales, il a été proposé que le phénomène décrit s'effectue possiblement en *trans* (327). Dans ce cas, plusieurs protéines Lin28 seraient responsables de modifier la structure de l'ARN afin de permettre la liaison du *NCp7-like domain*. De plus, le changement de la structure des boucles terminales pourrait être le signal permettant à Tut4 de reconnaître ses ARN cibles. Ceci expliquerait pourquoi la tutase reconnaît seulement le complexe Lin28/pre-let-7 et non chacune des molécules individuellement (203).

1.7 Objectifs de la thèse

La liaison par Lin28 de pri-let-7 ou de pre-let-7 est l'élément déclencheur de plusieurs mécanismes permettant l'inhibition de la maturation de cette famille de microARN. L'objectif global de cette thèse est de cartographier l'interaction entre la protéine Lin28 et le microARN précurseur let-7g. Puisque Lin28 interagit avec les boucles terminales des formes immatures de let-7, l'étude de pre-let-7g permettra également l'analyse des sites de liaison présents chez pri-let-7g. Nous utiliserons le miARN let-7g comme modèle de la famille let-7, puisque ce dernier est l'un des miARN les plus affectés lors de l'expression ou l'inhibition de Lin28 (245,248).

Nous croyons que les différents domaines de liaison à l'ARN de Lin28 possèdent des fonctions différentes lors de la liaison de pre-let-7g. En effet, le CSD de Lin28 est similaire à ceux retrouvés chez les protéines *cold shock* bactériennes. Ce type de domaine est connu pour lier les acides nucléiques avec une faible spécificité de séquence. De plus, certaines protéines bactériennes de ce type ont une préférence pour les pyrimidines, permettent une liaison coopérative à l'ARN et possèdent une activité chaperonne. Par conséquent, est-ce que le CSD de Lin28 lui confère ces différentes propriétés? Le second domaine de liaison à l'ARN de Lin28 est similaire aux protéines nucléocapsides des rétrovirus. Puisque ces protéines sont connues pour lier spécifiquement les guanines exposées au solvant, est-ce que le domaine ZBD de Lin28 a une telle spécificité?

Pour répondre à ces questions, la suite de cette thèse est divisée en deux chapitres. Le premier chapitre s'intéresse à la reconnaissance spécifique de pre-let-7g par Lin28. L'objectif de cette première publication est de déterminer les éléments minimaux permettant une liaison spécifique à haute affinité de pre-let-7g par Lin28. Nous croyons que l'un des deux domaines de liaison à l'ARN est suffisant pour obtenir une telle liaison. Cette première étude s'est effectuée alors que plusieurs groupes travaillaient sur la définition des sites d'interaction entre Lin28 et pre-let-7g (voir section *Mise en contexte*, p. 66 pour les détails historiques), mais notre étude *in vitro* effectuée avec des molécules purifiées a permis d'obtenir une cartographie très précise du site d'interaction. Après avoir défini ces éléments minimaux et pris connaissance des nouvelles études publiées, le deuxième chapitre caractérise l'assemblage

séquentiel de multiples protéines Lin28 sur une même boucle terminale de let-7g. Cet assemblage s'effectue suite à la liaison de la première protéine tel que décrit dans la première étude. L'objectif de ce deuxième manuscrit est d'identifier les différents sites de liaison présents sur l'ARN en plus de définir les caractéristiques permettant à Lin28 cet assemblage séquentiel. Cette dernière étude nous a permis de proposer un nouveau modèle pour la formation du complexe Lin28/pre-let-7g impliquant la liaison de plusieurs protéines sur le même ARN. Pour mener à bien ces études, différentes techniques telles que des essais de retard sur gel, de protection aux ribonucléases, de *in-line probing*, de spectrométrie RMN et de fluorimétrie ont été utilisées.

1.8 Méthodologie utilisée

Les différentes expériences présentées dans les deux prochains chapitres s'intéressent à différents aspects des interactions protéine-ARN tels que l'affinité, la définition des sites de liaison et la déstabilisation des structures secondaires d'ARN. Dans cette dernière section de l'introduction, un bref survol de différentes méthodes utilisées sera effectué.

1.8.1 Purification des macromolécules

Puisque, dans cette thèse, le complexe Lin28-miARN est caractérisé par des méthodes *in vitro*, la totalité des molécules utilisées ont d'abord dû être produites et purifiées. La protéine Lin28 et ses dérivées ont été exprimées dans la bactérie *E. coli*, puis purifiées par une combinaison de chromatographies d'affinité, d'échange d'ions et si nécessaire, de phase inverse. La majorité des ARN ont été produits par transcription *in vitro*, puis purifiés par affinité selon une méthode développée dans notre laboratoire (328,329).

1.8.2 Détermination de l'affinité d'un complexe

L'affinité d'un complexe macromoléculaire se définit comme la force d'une interaction entre deux molécules (330). Dans les différentes expériences présentées dans cette thèse, la constante de dissociation (K_d) est le paramètre quantitatif utilisé pour représenter l'affinité. Le K_d se définit comme le ratio entre le produit de la concentration de chacune des molécules libres ($[A] \times [B]$) sur la concentration du complexe ($[AB]$). Par conséquent, des molécules avec une forte force d'interaction seront majoritairement sous forme de complexe ce qui se traduira par une petite constante de dissociation.

Afin de déterminer l'affinité des différents complexes d'intérêt pour les études présentées dans les deux prochains chapitres, la méthode de retard sur gel natif a été privilégiée. Cette technique permet de visualiser l'effet de la liaison entre la protéine et un ARN dont l'extrémité 5' est marquée par un phosphore radioactif. Lors de la liaison de la

protéine à l'ARN, la taille du complexe sera plus élevée que celle de l'ARN libre ce qui affectera sa migration dans un gel non dénaturant. L'utilisation de radioactivité rend possible l'utilisation de très faibles quantités d'ARN, ce qui est essentiel lors de la détermination de K_d de haute affinité. En effet, pour être valable, cette technique demande que la concentration d'ARN utilisée soit d'au moins 50 fois inférieure au K_d mesuré (330). Il existe d'autres méthodes pour déterminer l'affinité d'un complexe macromoléculaire. Par contre, l'affinité élevée des complexes impliquant Lin28 limite l'utilisation de méthodes telles que la calorimétrie par titrage isotherme (ITC) ou la résonance plasmonique de surface (SPR).

1.8.3 Détermination des sites de liaison

Afin de déterminer les différents sites de liaison de la protéine Lin28 sur l'ARN, deux méthodes ont été utilisées en plus de la détermination de l'affinité de Lin28 envers divers ARN. La technique utilisée dans le premier chapitre est l'empreinte aux ribonucléases (*RNA footprinting*). Cette méthode a l'avantage de montrer visuellement l'emplacement qu'occupe la protéine sur l'ARN. En effet, en comparant la sensibilité des différents nucléotides à diverses ribonucléases en présence ou en absence de protéine, il est possible de visualiser les régions de l'ARN étant protégées suite à la formation du complexe. Par contre, cette méthode demande l'optimisation de conditions permettant à la fois la liaison de la protéine à l'ARN d'intérêt et une activité enzymatique efficace.

Deuxièmement, une série d'expérience utilisant des 2-aminopurines (2-AP) ont permis de déterminer l'ordre de liaison de Lin28 sur un ARN. La 2-AP est tout indiquée pour ce genre d'étude, puisque son émission de fluorescence est fortement liée au microenvironnement de la base (331). Une 2-AP libre en solution émet un signal plus intense qu'une 2-AP incorporée à un ARN (ou un ADN). Aussi, le positionnement adopté par la 2-AP dans l'ARN influence grandement son intensité d'émission. Une base empilée avec des purines sera presque entièrement éteinte (*quenched*), alors qu'une base dans une région simple brin accessible au solvant émettra de la fluorescence. En positionnant judicieusement les 2-AP à l'intérieur d'un ARN d'intérêt (en remplacement d'une adénine) on peut déterminer quelles sont les régions de

l'ARN affectées par la liaison de la protéine tout en minimisant l'impact sur la structure de cet ARN (332).

1.8.4 Visualisation de l'activité chaperonne à ARN

La caractérisation d'une activité chaperonne à ARN d'une protéine, c'est-à-dire son habileté à déstabiliser les structures secondaires d'ARN, peut s'effectuer de différentes façons. Les 2-AP (333), le couplage à un ribozyme (334), l'électrophorèse sur gel (335) et différentes applications de la fluorescence (327,336) sont toutes des méthodes ayant permis d'étudier une telle activité.

L'expérience présentée dans le deuxième manuscrit est basée sur le transfert d'énergie par résonance de fluorescence (*fluorescence resonance energy transfer*, FRET). Chaque brin du duplex utilisé dans cette expérience possède un fluorophore (Cy3 ou Cy5) à son extrémité 5'. Lorsque les brins sont appariés l'un à l'autre, l'excitation de Cy3 provoque une émission qui est absorbée par le fluorophore Cy5. Il devient alors possible d'observer l'émission spécifique de ce dernier. Puisque l'émission provenant de Cy5 est dépendante de la distance le séparant de Cy3, il est possible d'étudier la dissociation des brins (337) ou la déformation de l'hélice d'ARN (338) suite à l'ajout de protéines. Le titrage d'une protéine avec une activité chaperonne causera donc une diminution du signal provenant de Cy5 suite à la déstabilisation du duplex.

2.0 Interaction du domaine *NCp7-like* de Lin28 avec TL-let-7g

Mise en contexte

Les travaux nécessaires à cette première publication se sont effectués de 2009 à 2011, soit un an après les premières publications démontrant l'interaction entre Lin28 et pre-let-7g. Lors de la réalisation de ces travaux, aucune structure tridimensionnelle à haute résolution de ce complexe n'avait encore été publiée. Aussi, si certaines publications avaient identifié des nucléotides importants pour l'interaction de l'ARN avec Lin28, aucun consensus n'existait (voir l'introduction de cette publication, p. 69). De même, le rôle de chacun des domaines de liaison à l'ARN de Lin28 dans l'interaction avec pre-let-7g n'était pas bien défini. C'est dans ce contexte que cette première étude a été réalisée et publiée.

Cette première publication, avait pour but de répondre à deux objectifs importants. Nous cherchions, d'abord, à déterminer la région minimale de Lin28 lui permettant d'interagir spécifiquement avec pre-let-7g, tout en conservant une haute affinité. Ensuite, la région de pre-let-7g permettant une telle interaction a été identifiée. Ces objectifs ont été atteints en utilisant une combinaison d'essais de retard sur gel, de protection aux ribonucléases, de *in-line probing* et de spectrométrie RMN. Afin de cartographier précisément les sites d'interactions, plusieurs molécules dérivées de Lin28 ou de pre-let-7g ont dû être purifiées avant de pouvoir être utilisées dans ces différents essais *in vitro*.

Plusieurs auteurs ayant contribué à cette publication, il est important de préciser ma contribution personnelle. La planification et la conception des différentes expériences ont été faites par moi-même (AD), Pascale Legault (PL) et James G. Omichinski (JGO). Tout le clonage et la purification par affinité des ARN nécessaires au projet ont été réalisés par AD, quoique celui-ci ait pu bénéficier de l'aide de Jonathan Bouvette (JB) lors du stage d'été de ce dernier. Le clonage et la purification des protéines non marquées ont été faits par AD, alors que les protéines destinées à la RMN ont été purifiées par Ao Yang (AY) et JB. Les essais de retard sur gel ont été effectués par AD, lequel a été assisté par JB. AD a également été responsable des essais de protection aux ribonucléases, ainsi que du *in-line probing*. Tous les résultats biochimiques ont été analysés par AD, tandis que AY a réalisé et analysé les expériences de RMN. Finalement, la réalisation des figures et la rédaction de l'article ont été faites par AD et AY (pour la partie RMN), puis le manuscrit a été revu par PL et JGO.

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

Nucleic Acids Res. 2012 Feb; 40 (4):1767-77. doi: 10.1093/nar/gkr808. Epub 2011 Oct 19.

Abstract

The pluripotency factor Lin28 is a highly conserved protein comprised of 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 sequence similarity 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 nucleotides (nts) 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 Droscha 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.

MicroRNAs play important roles in cell differentiation (5-7), and, in mammals, several miRNAs have been shown to act as oncogenes and tumour suppressors [reviewed in (8-13)]. Amongst those playing a role as tumour 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 ten 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 amongst the most affected. Several mechanisms have been proposed to explain the Lin28 inhibition of let-7 biogenesis. Lin28 was shown to interfere with the Droscha 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 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 the 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 reduce 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 carcinoma 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 Stratagene 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'-³²P]-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 d₁₈-HEPES at pH 6.4, 50 mM NaCl, 0.05 mM NaN₃ and 10% D₂O).

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 100,000g 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 CaCl₂ 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, 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 ¹⁵N- and ¹⁵N/¹³C labeling was obtained by growing the cells in minimal media containing ¹⁵N-labeled NH₄Cl and ¹³C₆-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 C₄ 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 electrophoretic mobility shift assay (EMSA), the ³²P-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 0.02X to 50X 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 2-I) by non-linear regression analysis within the Origin 7 SR4 version 7.0552 software (OriginLab, MA). 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 Lin28₁₁₉₋₁₈₀ 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 T₁ ribonuclease from *Aspergillus oryzae* (Sigma) was added and the incubation continued for 15 min at 4°C. The reaction was stopped by addition of

Precipitation/Inactivation buffer (Ambion), incubation for 15 min at -20°C and centrifugation at 16,000g 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 1,900 V for 5 h. The sequencing gel was exposed 2 h to a storage phosphor screen (Bio-Rad). The ³²P-labeled RNA bands were visualized and quantified as for the EMSA assay. For each gel, lanes containing undigested RNA, partial alkaline-digested RNA and T1-digested RNA were included as controls (45).

NMR Spectroscopy

For NMR studies, the following samples were prepared in NMR buffer: 1.0 mM ¹⁵N-labeled Lin28₁₃₆₋₁₈₀; 1.3 mM ¹⁵N-labeled Lin28₁₁₉₋₁₈₀; 1.1 mM ¹³C/¹⁵N-labeled Lin28₁₁₉₋₁₈₀; 0.1 mM TL-let-7g:¹⁵N-labeled Lin28₁₁₉₋₁₈₀; 0.1 mM Δbulge TL-let-7g:¹⁵N-labeled Lin28₁₁₉₋₁₈₀; 0.1 mM G34AG35A TL-let-7g:¹⁵N-labeled Lin28₁₁₉₋₁₈₀; 1.0 mM TL-let-7g:¹⁵N-labeled Lin28₁₁₉₋₁₈₀ and 1.3 mM TL-let-7g:¹³C/¹⁵N-labeled Lin28₁₁₉₋₁₈₀. For the TL-let-7g:Lin28₁₁₉₋₁₈₀ complexes, the samples were prepared by titration of Lin28₁₁₉₋₁₈₀ into a TL-let-7g sample. All NMR experiments were collected on Varian UnityINOVA 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, ¹⁵N and ¹³C) of Lin28₁₁₉₋₁₈₀ in the free and TL-let-7g-bound form were assigned using the following NMR experiments collected at 35°C: two-dimensional (2D) ¹H-¹⁵N HSQC (46), three-dimensional (3D) HNCACB (47-49) and 3D (HB)CBCA(CO)NNH (48,49). ¹H, ¹³C and ¹⁵N 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 was 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 electrophoretic mobility shift assays (EMSA) with purified recombinant proteins and *in vitro* transcribed RNAs (Figure 2-1). 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 (Lin28₁₋₂₀₉) for the terminal loop of pre-let-7g (TL-let-7g; Figure 2-1D). We attempted to fit the data (Figure 2-2A) using the classical one-site binding equation, but the fit was rather poor (Figure 2-2C, red line). Using the Hill equation, a much better fit was obtained (Figure 2-2C, 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 2-I). 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 2-I), 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 ($K_d = 9 \pm 3$ 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 (Lin28₃₉₋₁₁₂) or the two ZBDs (Lin28₁₁₉₋₁₈₀) for pre-let-7g, TL-let-7g and duplex let-7g GNRA (Figure 2-1). For the N-terminal Lin28₃₉₋₁₁₂ fragment comprising the CSD (Figure 2-1A), 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 Lin28₃₉₋₁₁₂ domain detected at concentrations higher than 250 nM. In contrast, the Lin28₁₁₉₋₁₈₀ fragment

containing the ZBDs displays only slightly lower affinities than full-length Lin28 towards 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 2-I, Figure 2-2 2B-C). These binding data for Lin28₁₁₉₋₁₈₀ can be fitted to a classical one-site binding equation (Figure 2-2C and Table 2-I). For the duplex-let-7g GNRA, no specific binding could be observed with Lin28₁₁₉₋₁₈₀. Thus, compared to the full-length protein, the C-terminal Lin28₁₁₉₋₁₈₀ fragment containing the two ZBDs displays similar specificity towards RNAs derived from pre-let-7g. Furthermore, the affinity between Lin28₁₁₉₋₁₈₀ and TL-let-7g is only 10-fold weaker than between the full-length Lin28 protein and pre-let-7g, indicating that Lin28₁₁₉₋₁₈₀ 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 2-3). The percentage of sequence similarity to HIV-1 NCp7 is relatively high for both murine (45%) and human Lin28₁₂₃₋₁₈₀ (45%). All three proteins contain identical zinc-chelating amino acids (CCHC) and spacing between these residues in the two ZBDs. In addition, the spacing between the two ZBDs is similar with 7 residues in NCp7 of HIV-1 and 8 residues in the Lin28 sequences (Figure 2-3). 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 Lin28₁₁₉₋₁₈₀ to TL-let-7g, we performed NMR chemical shift perturbation experiments. In both the free and bound forms, Lin28₁₁₉₋₁₈₀

displays a well-dispersed ^1H - ^{15}N HSQC spectrum (Figure 2-4A), indicating that both adopt a stable homogenous conformation in solution. Analysis of the ^1H 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 2-4B; $\Delta\delta > 0.4$ ppm). When mapped onto the primary structure of Lin28₁₁₉₋₁₈₀, 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 Lin28₁₁₉₋₁₈₀ 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 2-1A). Of the three deletion fragments that were expressed and purified (Lin28₁₃₆₋₁₈₀, Lin28₁₁₉₋₁₄₀ and Lin28₁₁₉₋₁₆₀), none binds TL-let-7g with high affinity ($K_d > 5$ μM ; Table 2-II), and thus Lin28₁₁₉₋₁₈₀ constitutes the minimal domain required for TL-let-7g 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 Lin28₁₃₆₋₁₈₀ is not due to protein misfolding, we compared the ^1H - ^{15}N HSQC spectrum of this fragment with that of Lin28₁₁₉₋₁₈₀ (Supplementary Figure 2-1). The chemical shift similarity between these two spectra indicates that the ZBDs adopt a similar fold in Lin28₁₃₆₋₁₈₀ and Lin28₁₁₉₋₁₈₀. In addition, the importance of the KR-rich domain was further investigated using a mutant of Lin28₁₁₉₋₁₈₀ 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, Lin28₁₁₉₋₁₈₀ (KR⁻) does not bind TL-let-7g with high affinity ($K_d > 5$ μM ; Table 2-II). 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 Lin28₁₁₉₋₁₈₀ (Table 2-II). Thus, in addition to sharing sequence similarity with NCp7, Lin28₁₁₉₋₁₈₀ 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 Lin28₁₁₉₋₁₈₀ 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 2-2) 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 2-5A). For the ribonuclease protection assay, RNase T₁ 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 Lin28₁₁₉₋₁₈₀ (Figure 2-5) 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 Lin28₁₁₉₋₁₈₀. Only one G residue (G18) becomes more accessible in the presence of Lin28₁₁₉₋₁₈₀, most likely resulting from destabilization of the predicted G18-C12 base pair (see Discussion). 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.

Detailed mapping of the let-7g determinants for Lin28₁₁₉₋₁₈₀ 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 Lin28₁₁₉₋₁₈₀ ($K_d = 1.3$ nM; Table 2-III). 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 2-1D) 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 2-1D; Δ iloop) causes a 5.3-fold increase in the K_d compare to the wild-type 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 Lin28₁₁₉₋₁₈₀ binding, but the hairpin loop appears to be replaceable.

Several EMSA results indicate that the G-rich bulge contributes significantly to Lin28₁₁₉₋₁₈₀ binding. First, deletion of the G-rich bulge (Figure 2-1D; Δ bulge) has a substantial effect, since no specific binding could be observed with this mutant at a protein concentration as high as 5 μ M. 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 Lin28₁₁₉₋₁₈₀ binding, each guanine was individually mutated and the K_d was determined by EMSA (Table 2-III). 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 Lin28₁₁₉₋₁₈₀. The G-rich bulge clearly represents the main RNA determinant of the high-affinity interaction with Lin28₁₁₉₋₁₈₀, 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 Lin28₁₁₉₋₁₈₀ in complex with TL-let-7g mutants (Supplementary Figure 2-3). As expected, the ¹H-¹⁵N HSQC spectrum of the G34AG35A/Lin28₁₁₉₋₁₈₀ complex is almost identical to that of the TL-let-7g/Lin28₁₁₉₋₁₈₀ complex, particularly for residues from the two ZBDs. In contrast, the ¹H-¹⁵N HSQC spectrum of the Δ bulge/Lin28₁₁₉₋₁₈₀ complex indicates that, at the high concentration (0.1 mM) used for these NMR studies, Lin28₁₁₉₋₁₈₀ interacts with the Δ bulge mutant, but in a different manner than observed for the wild-type TL-let-7g. Thus, these NMR results confirm that the G-rich bulge is the main determinant for high-affinity binding of Lin28₁₁₉₋₁₈₀ to TL-let-7g.

Given the importance of the G-rich bulge for Lin28₁₁₉₋₁₈₀ binding, we examined the sequences of all known pre-let-7g RNAs (Supplementary Figure 2-4). 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 μ M 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 (20,38). 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 2-2A and 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 2-2A). 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 be needed to 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 Lin28₁₁₉₋₁₈₀ 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 Lin28₁₁₉₋₁₈₀ and the NCp7 protein also extends to their binding affinity and specificity. Indeed, both Lin28₁₁₉₋₁₈₀ 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 Lin28₁₁₉₋₁₈₀ and NCp7 bind TL-let-7g with similar affinities under the same conditions. We find that Lin28₁₁₉₋₁₈₀ 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 2-4), 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 x 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). Lin28₁₁₉₋₁₈₀ 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 Lin28₁₁₉₋₁₈₀ 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 Lin28₁₁₉₋₁₈₀, 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 speculation is inconsistent with a classical evolutionary model that does not include viruses, but it 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 Lin28₁₁₉₋₁₈₀. 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 2-5A) 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.

Acknowledgements

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

Funding

This work was supported by grants from 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: NSERC.

Tables

Table 2-I 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

^aIn these cases, the Hill equation was used to derived the K_d values.

^bNo specific binding observed. The gel mobility shift assays display smearing and multiple shifts, indicating non-specific binding.

Table 2-II. Dissociation constants (K_d in nM) of mutants of the NCp7-like domain of Lin28 (Lin28₁₁₉₋₁₈₀) 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

Table 2-III. Dissociation constants (K_d in nM) of the NCp7-like domain of Lin28 (Lin28₁₁₉₋₁₈₀) for various mutants of the TL-let-7g RNA.

TL-let-7g RNA	K_d (nM)	$K_d / [K_d \text{ (wt)}]^a$
Wild type (wt)	1.3 ± 0.3	1
GNRA tetraloop	4 ± 1	3.1
C25A	2.1 ± 0.5	1.6
Δ iloop	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
Δ bulge	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

^aThe $K_d / [K_d \text{ (wt)}]$ 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.

^bNo specific binding observed. The gel mobility shift assays display smearing and multiple shifts, indicating non-specific binding.

Figures

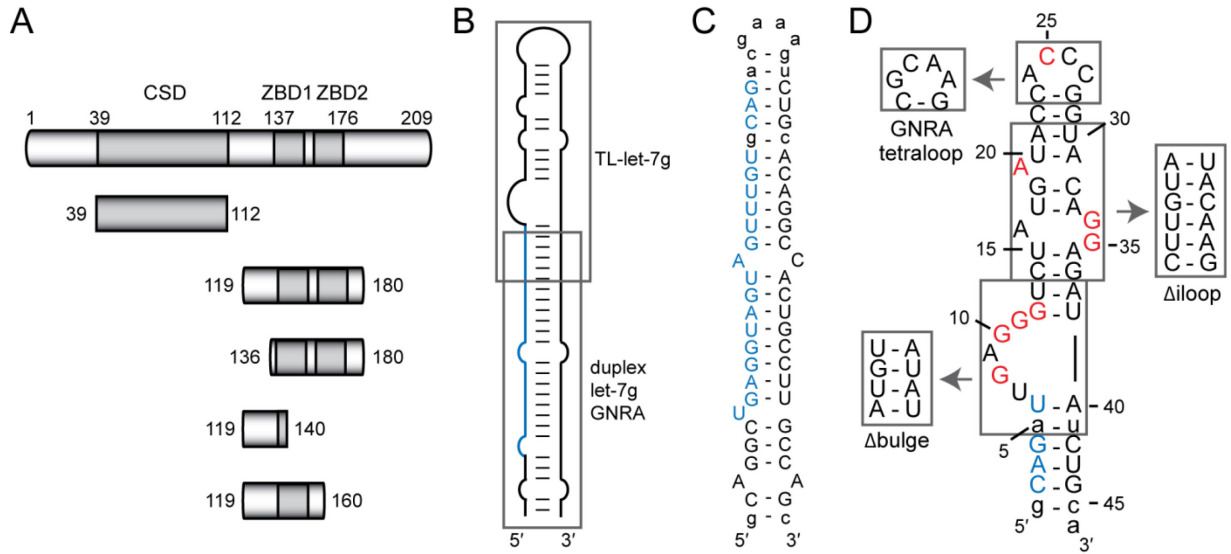


Figure 2-1. The Lin28 protein, pre-let-7g RNA and related sequences used in this study. **(A)** Schematic representation of the primary structures of Lin28 and deletion fragments. The grey boxes delineate sequences of known RNA-binding motifs: a cold shock domain (CSD) and a pair of retroviral-type CCHC zinc-binding domains (ZBD1 and ZBD2). **(B)** Schematic representation of pre-let-7g, indicating the regions (grey boxes) from which TL-let-7g and duplex let-7g GNRA were derived. **(C)(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.

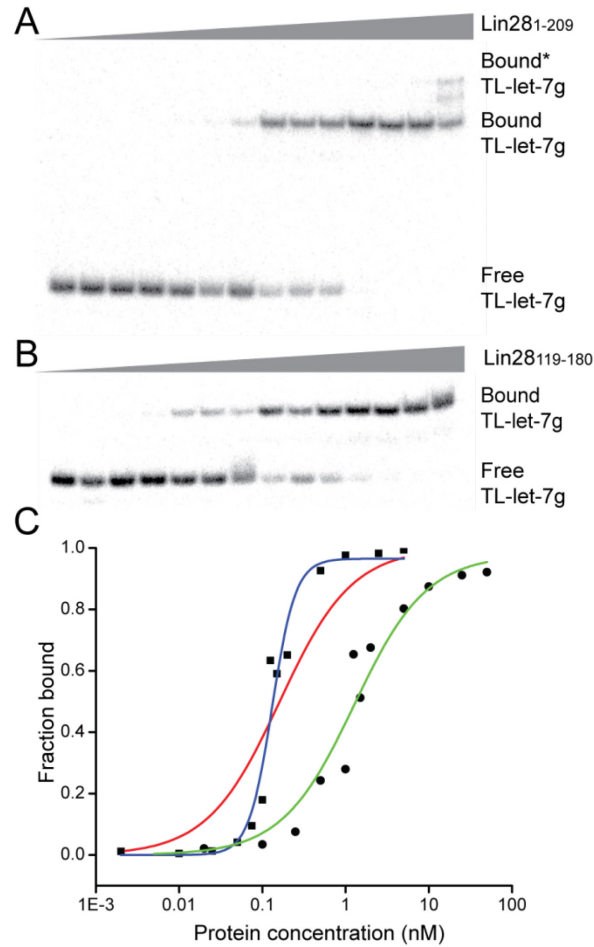


Figure 2-2. Electrophoretic mobility shift assay (EMSA) of TL-let-7g with Lin28₁₋₂₀₉ and Lin28₁₁₉₋₁₈₀. **(A)** Typical EMSA performed with 1 pM of 5'-[³²P]-labeled TL-let-7g and increasing concentrations of Lin28₁₋₂₀₉ (0.0, 0.002, 0.010, 0.025, 0.050, 0.075, 0.10, 0.15, 0.20, 0.50, 1.0, 2.5 and 5.0 nM). **(B)** Typical EMSA for TL-let-7g and increasing concentrations of Lin28₁₁₉₋₁₈₀ (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 Lin28₁₋₂₀₉ (squares) is 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 Lin28₁₁₉₋₁₈₀ (dots) is fitted to the one site binding equation (green line; $K_d = 1.3$ nM).

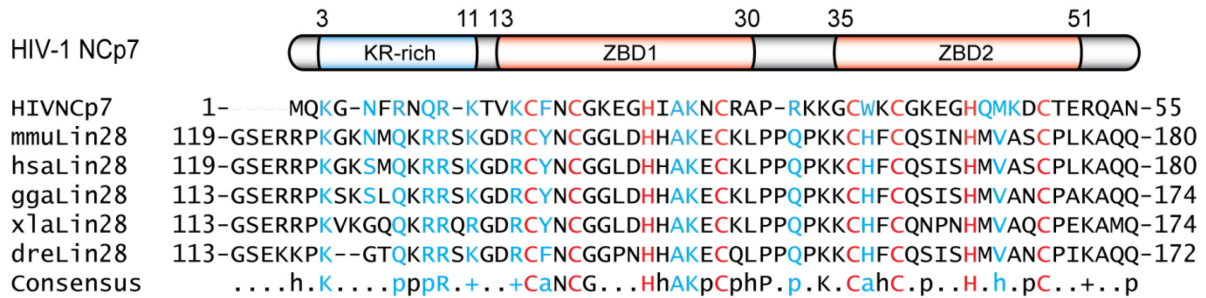


Figure 2-3. Sequence similarity between the HIV-1 NCp7 and the C-terminal domain of Lin28. The sequences of HIV-1 NCp7 and Lin28 from *Mus musculus* (mmu), *Homo sapiens* (hsa), *Gallus gallus* (gga), *Xenopus laevis* (xla), *Danio rerio* (dre) were aligned using ClustalW2 (70). A consensus sequence 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.

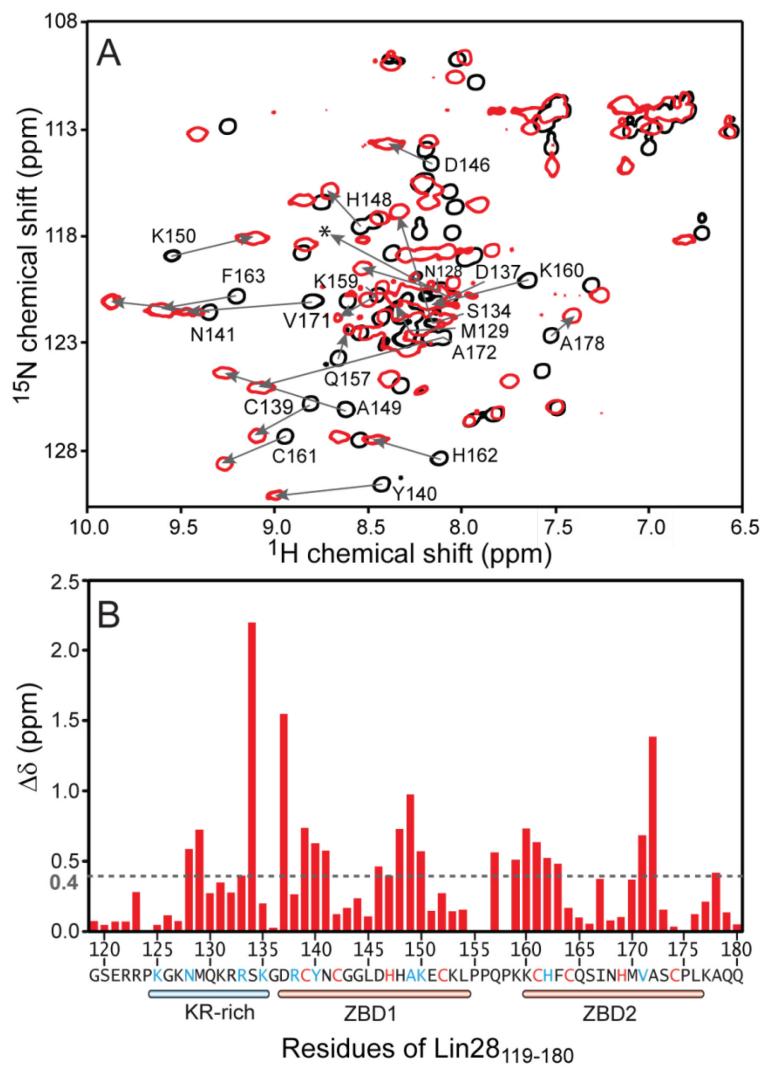


Figure 2-4. Chemical shift mapping of Lin28₁₁₉₋₁₈₀ binding the TL-let-7g RNA (A) Superposition of 2D ¹H-¹⁵N HSQC spectra of 1 mM ¹³C/¹⁵N-labeled Lin28₁₁₉₋₁₈₀ in the free form (black) and bound to 1 mM TL-let-7g (red). The signals from the free form that display a significant chemical shift change ($\Delta\delta > 0.4$ 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 \pm 0.03 ppm) observed after addition of a molar equivalent of TL-let-7g to 1 mM ¹³C/¹⁵N-labeled Lin28₁₁₉₋₁₈₀. The chemical shift differences ($\Delta\delta$) were calculated according to the formula $\Delta\delta = [(\Delta H^N)^2 + (0.17\Delta N^H)^2]^{1/2}$.

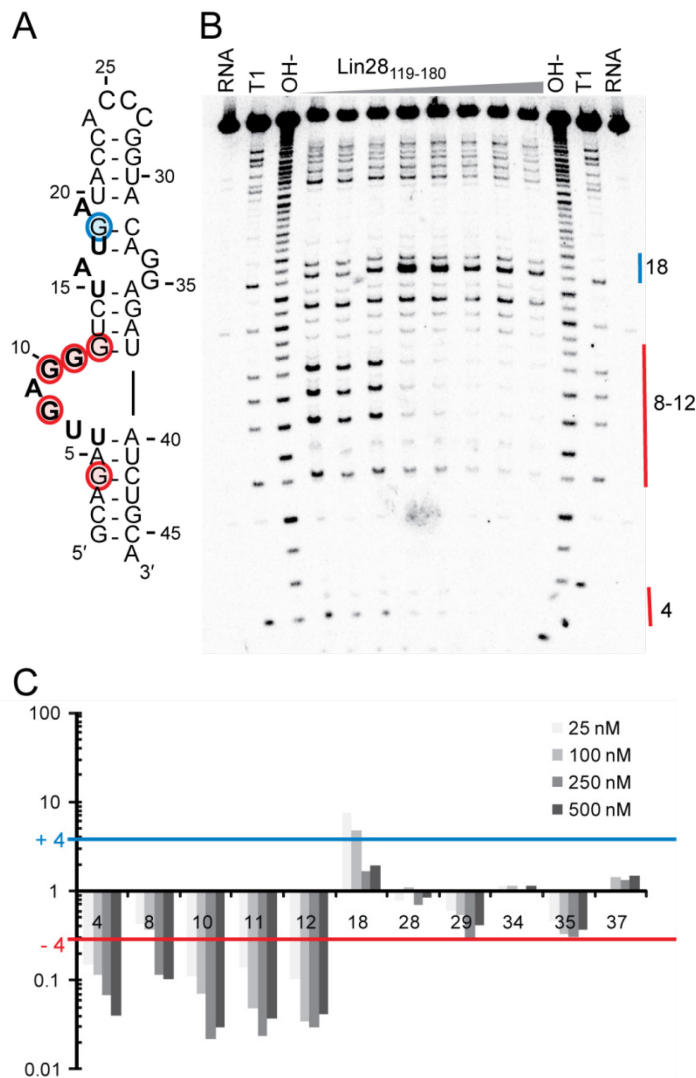


Figure 2-5. Footprint analysis of TL-let-7g with RNase T₁. **(A)** Secondary structure of TL-let-7g with the mapping of in-line probing and T₁ footprinting data. Residues that are the most susceptible to spontaneous cleavage through in-line attack are in bold (Supplementary Figure 2-2), and residues that experience a significant reduction ($I_p/I_0 = -4$) or enhancement ($I_p/I_0 = +4$) of T₁ cleavage in the presence of Lin28₁₁₉₋₁₈₀ are shaded in red and blue, respectively. **(B)** Typical RNA footprinting gel of TL-let-7g in the absence and presence of Lin28₁₁₉₋₁₈₀ (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 T₁ 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 T₁ cleavage of each guanine obtained at 25 nM (blue), 100 nM (gold), 250 nM (green) and 500 nM (violet) of Lin28₁₁₉₋₁₈₀.

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. Poleskaya, 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.

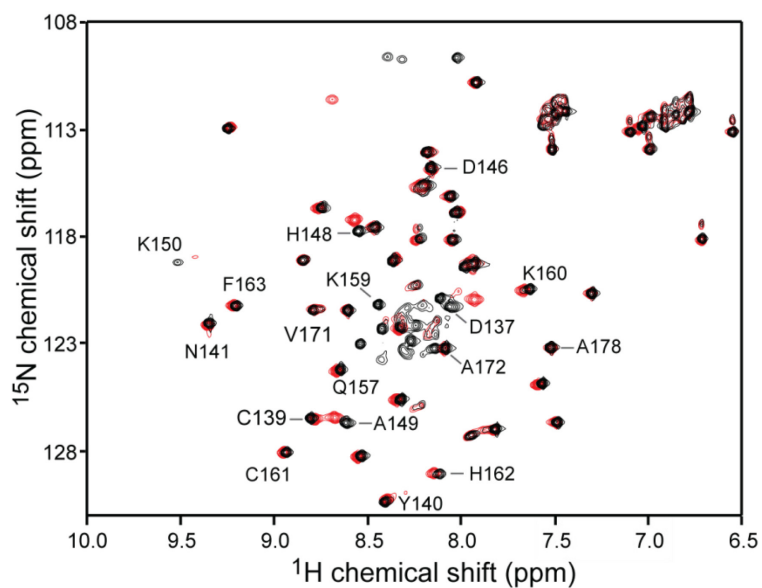
35. Xu, B. and Huang, Y. (2009) Histone H2a mRNA interacts with Lin28 and contains a Lin28-dependent posttranscriptional regulatory element. *Nucleic Acids Res*, **37**, 4256-4263.
36. Xu, B., Zhang, K. and Huang, Y. (2009) Lin28 modulates cell growth and associates with a subset of cell cycle regulator 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) ^1H , ^{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) ^1H , ^{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**, 277-293-293.
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.

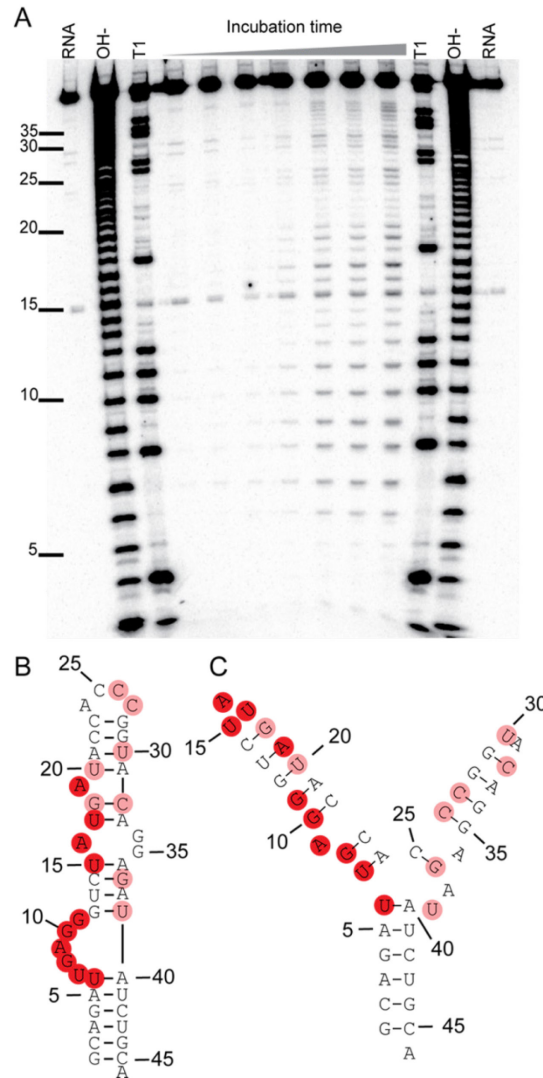
56. Weiss, J.N. (1997) The Hill equation revisited: uses and misuses. *FASEB J*, **11**, 835-841.
57. Ermolenko, D.N. and Makhatadze, G.I. (2002) Bacterial cold-shock proteins. *Cell Mol Life Sci*, **59**, 1902-1913.
58. Jiang, W., Hou, Y. and Inouye, M. (1997) CspA, the major cold-shock protein of *Escherichia 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 x 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. *Nature 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.

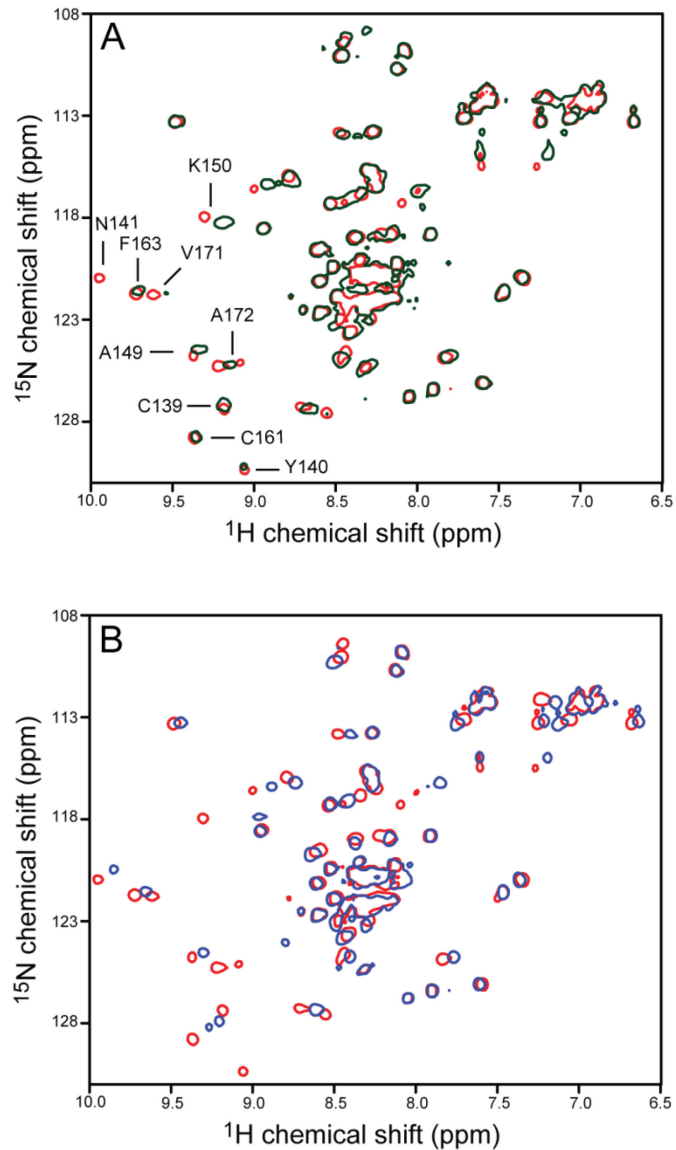
Supplementary material



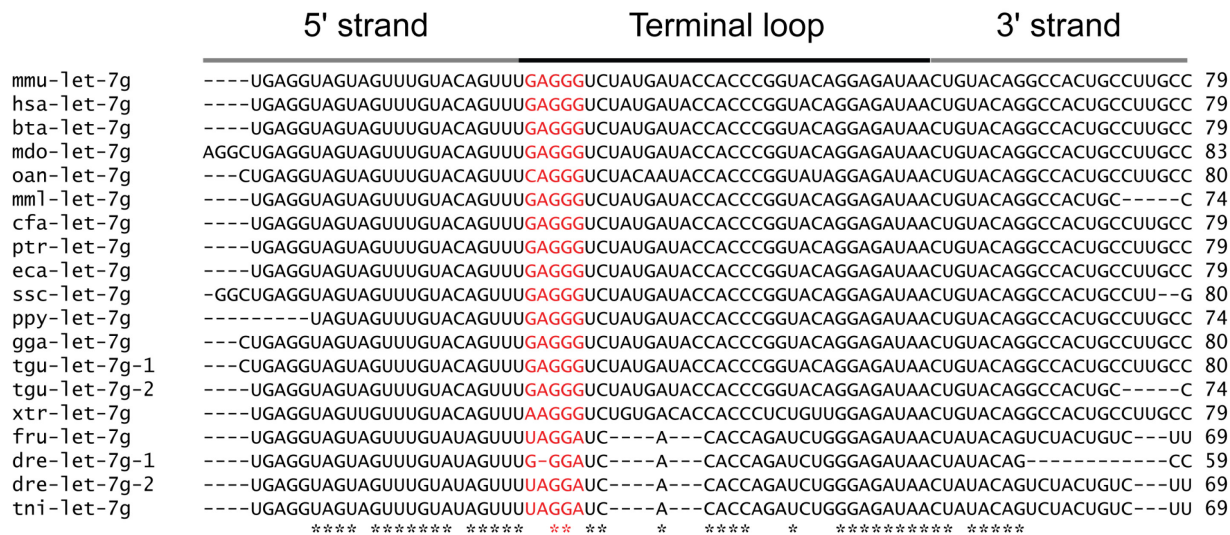
Supplementary Figure 2-1. Superposition of 2D ^1H - ^{15}N HSQC spectra of 1.3 mM ^{15}N -labeled Lin28₁₁₉₋₁₈₀ (black) and 1 mM ^{15}N -labeled Lin28₁₃₆₋₁₈₀ (red). Signals of residues 136-180 from Lin28₁₁₉₋₁₈₀ that are significantly affected by TL-let-7g binding (see Figure 4) are annotated here.



Supplementary Figure 2-2. In-line probing assay of TL-let-7g. (A) Time course of in-line probing (1 min, 1 h, 2 h, 5 h, 20 h, 30 h and 45 h). Lanes with input TL-let-7g (RNA), an alkaline hydrolysis ladder (OH-) and a T₁ hydrolysis ladder (T1) are included. Nucleotides that are most sensitive to Mg²⁺-induced cleavage are mapped on the secondary structures obtained by the (B) *mFold* (1) and (C) *MC-fold* (2) software. Red and pink circles represent nucleotides that are respectively highly sensitive and moderately sensitive to Mg²⁺-induced cleavage, as determined by the band intensities from the 20-h lane after subtraction of corresponding intensities in the input RNA lane.



Supplementary Figure 2-3. Lin28₁₁₉₋₁₈₀ binds to G34AG35A TL-let-7g and to Δ bulge TL-let-7g. Superposition of 2D ^1H - ^{15}N HSQC spectra of (A) 0.1 mM TL-let-7g: ^{15}N -labeled Lin28₁₁₉₋₁₈₀ (red) and 0.1 mM G34AG35A TL-let-7g: ^{15}N -labeled Lin28₁₁₉₋₁₈₀ (green) and (B) 0.1 mM TL-let-7g: ^{15}N -labeled Lin28₁₁₉₋₁₈₀ (red) and 0.1 mM Δ bulge TL-let-7g: ^{15}N -labeled Lin28₁₁₉₋₁₈₀ (blue). In (A) characteristic signals from the bound form that display a significant chemical shift change ($\Delta\delta > 0.4$ ppm) as a result of TL-let-7g binding are annotated (see Figure 4). Weak signals for N141 and V171 are often observed for the 0.1 mM TL-let-7g: ^{15}N -labeled Lin28₁₁₉₋₁₈₀ complex.



Supplementary Figure 2-4. Sequence alignment of all known pre-let-7g RNA : *Mus musculus* (mmu), *Homo sapiens* (hsa), *Bos taurus* (bta), *Monodelphis domestica* (mdo), *Ornithorhynchus anatinus* (oan), *Macaca mulatta* (mm1), *Canis familiaris* (cfa), *Pan troglodytes* (ptr), *Equus caballus* (eca), *Sus scrofa* (ssc), *Pongo pygmaeus* (ppy), *Gallus gallus* (gga), *Taeniopygia guttata* (tgu), *Xenopus tropicalis* (xtr), *Fugu rubripes* (fru), *Danio rerio* (dre), *Tetraodon nigroviridis* (tni). The sequence of the G-rich bulge is highlighted in red. Sequences from miRBase were aligned with T-coffee (3).

Supplementary references

1. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, 31, 3406-3415.
2. Parisien, M. and Major, F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, 452, 51-55.
3. 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.

3.0 Assemblage séquentiel de plusieurs protéines Lin28 sur TL-let-7g

Mise en contexte

Les expériences menant à ce deuxième manuscrit ont été effectuées de 2011 à 2013. L'objectif de cette deuxième étude était de réconcilier plusieurs données publiées semblant contradictoires par rapport à la liaison de Lin28 à pre-let-7. Au départ, différentes études biochimiques, structurales et fonctionnelles avaient mis en évidence l'importance de trois régions de pre-let-7g pour son interaction avec Lin28. La protéine Lin28 possédant deux domaines de liaison à l'ARN, un maximum de deux régions d'ARN devraient être requises pour la liaison de la protéine lors de la formation d'un complexe de ratio stœchiométrique 1:1.

En nous basant sur la multiplicité des bandes présentes sur gel natif lors de la liaison d'une forte concentration de Lin28 à pre-let-7g, nous avons émis l'hypothèse que plusieurs protéines Lin28 puissent lier simultanément la même boucle terminale de pre-let-7g. En utilisant une combinaison d'essais de retard sur gel pour déterminer soit l'affinité ou la stœchiométrie de la liaison, nous avons identifié trois régions importantes de l'ARN pour la liaison de Lin28. Par la suite, nous avons caractérisé l'assemblage séquentiel de complexes 1:1, 1:2 et 1:3 entre TL-let-7g et Lin28. Des essais de fluorescence ont permis d'attribuer un rôle à chacun des domaines de liaison à l'ARN de Lin28 en plus de clarifier l'ordre de liaison de la protéine sur l'ARN. L'ensemble de ces données nous ont permis de proposer un modèle d'assemblage de Lin28 sur TL-let-7g. Finalement, des essais *in vitro* avec Dicer ont permis d'adresser l'implication du phénomène de multimérisation sur le clivage de pre-let-7g par cette endonucléase.

Pour cette étude, la planification et la conception des différentes expériences ont été effectuées par AD et PL. Tout le clonage et la purification par affinité des ARN nécessaires au projet ont été réalisés par AD assisté à l'occasion de JB, d'Alexis Rompré-Brodeur et de Geneviève Di Tomasso. Le clonage et la purification des protéines ont été faits par AD (et JB pour les essais de Dicer seulement). Les essais de retard sur gel, tout comme ceux de fluorescence, ont été effectués et analysés par AD. Pour les essais de Dicer, la mise au point des expériences a été effectuée par JB, AD et PL. Ces expériences ont été réalisées par JB, analysées par JB, AD et PL, alors que les figures ont été préparées par JB. Finalement, les autres figures et la rédaction de l'article ont été réalisées par AD, puis revues par PL.

Stepwise assembly of multiple Lin28 proteins on the terminal loop of let-7 miRNA precursors

Alexandre Desjardins, Jonathan Bouvette and Pascale Legault

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

Accepté par la revue Nucleic Acids Research le 11 décembre 2013.

Abstract

Lin28 inhibits the biogenesis of let-7 miRNAs through direct interactions with let-7 precursors. Previous studies have described seemingly inconsistent Lin28 binding sites on pre-let-7 RNAs. Here, we reconcile these data by examining the binding mechanism of Lin28 to the terminal loop of pre-let-7g (TL-let-7g) using biochemical and biophysical methods. First, we investigate Lin28 binding to TL-let-7g variants and short RNA fragments and identify three independent binding sites for Lin28 on TL-let-7g. We then determine that Lin28 assembles in a stepwise manner on TL-let-7g to form a stable 1:3 complex. We show that the cold-shock domain (CSD) of Lin28 is responsible for remodeling the terminal loop of TL-let-7g, whereas the NCp7-like domain facilitates the initial binding of Lin28 to TL-let-7g. This stable binding of multiple Lin28 molecules to the terminal loop of pre-let-7g extends to other precursors of the let-7 family, but not to other pre-miRNAs tested. We propose a model for stepwise assembly of the 1:1, 1:2 and 1:3 pre-let-7g/Lin28 complexes. Stepwise multimerization of Lin28 on pre-let-7 is required for maximum inhibition of Dicer cleavage for a least one member of the let-7 family and may be important for orchestrating the activity of the several factors that regulate let-7 biogenesis.

Introduction

MicroRNAs (miRNAs) are short single-stranded RNAs that control fundamental biological processes in plants and animals by acting as post-transcriptional regulators of mRNA expression [for detailed reviews see (1-4)]. The miRNAs from the let-7 family were among the first to be identified and have been extensively studied given their central roles in development, cell differentiation and tumor suppression (5-13). The let-7 miRNA loci are often present in multiple copies in a single genome, with the mature let-7 miRNAs being highly conserved across species (6,7,13,14). Furthermore, the terminal loops of their precursors contain highly conserved nucleotides that mediate interaction with factors that control miRNA biogenesis, such as Lin28 (15-19), hnRNP A1 (20,21) and KSRP (21-23), thereby adding a layer of complexity to miRNA-mediated gene regulation.

The Lin28 protein is a key post-transcriptional inhibitor of miRNA biogenesis that acts selectively on let-7 miRNAs (16,24), and there exists two isoforms in mammals [Lin28A (also termed Lin28) and Lin28B (25)]. Lin28 inhibits Drosha cleavage of primary let-7 transcripts (pri-let-7) in the nucleus (16,24,26,27) and interferes with Dicer cleavage of precursor let-7 (pre-let-7) in the cytoplasm (18,28,29). Moreover, Lin28 can counteract the stimulation of let-7 biogenesis brought about by KSRP, a factor that directly interacts with the terminal loops of several pri/pre-let-7 miRNAs (22,23). In the absence of Lin28, several Tutases (TUT2, TUT4 and TUT7) mono-uridylylate a specific subset of pre-miRNAs, and this enhances the cleavage activity of Dicer (30). In contrast, Lin28 induces the oligo-uridylation of pre-let-7 by terminal uridylyl transferases (TUT4/Zcchc11 or TUT7/Zcchc6 in mouse and human), which inhibits cleavage by Dicer and promotes pre-let-7 decay (17,28,31-35). Thus, Lin28 negatively regulates let-7 biogenesis by inhibiting the Drosha- and Dicer-mediated cleavage of immature forms of let-7, by counteracting the action of factors that promote such processing and by enhancing mechanisms that specifically promote pre-let-7 decay.

Lin28 also regulates translation of several mRNAs [for a recent review see (36)] and functions as one of four factors that are sufficient to reprogram human somatic cells into induced pluripotent stem cells (37). Not surprisingly, recent studies have also associated Lin28

with development traits (38-40), development defects (41,42), tissue repair (43), increased cancer susceptibility (43-47) and advanced human malignancies (47-49).

RNA recognition by Lin28 is being intensively investigated given that it contains two highly-conserved RNA-binding domains [Fig. 1A; (41,50)]; an N-terminal cold shock domain (CSD) and a C-terminal NCp7-like domain composed of one KR-rich motif and two CCHC-type zinc-binding domains [ZBDs; (19)]. Recent crystal structures indicate that the Lin28 CSD binds with low sequence specificity to single-stranded RNAs derived from the terminal loop of pre-let-7 (TL-let-7) and fitting preferably either the 5'-NGNGAYNNN-3' [Y= pyrimidine and N = any base; (51)] or the 5'-GUNNUNN-3' (52) consensus. Crystal (51) and solution (53) structures also indicate that the NCp7-like domain specifically binds a G-rich region (5'-GGAG-3') found at the 3'-end of TL-let-7. These structural data are consistent with the importance of the 5'-GGAG-3' sequence for Lin28 binding and for the Lin28-dependent uridylation by the TUTase Zcchc11 in pre-let-7a-1 (17,52). In addition, biochemical studies have defined a G-rich bulge at the 5'-end of TL-let-7g (5'-UGAGGG-3') as a primary binding site for Lin28 (18,19). In agreement with these results, binding studies using small single-stranded oligoribonucleotides and genome-wide studies of Lin28-associated mRNAs identified several G-rich and U-rich sequences as Lin28 targets (51,52,54-57), with a noted preference for G-rich sequences by the NCp7-like domain and for U-rich sequences by the CSD (51,57). Globally, these studies reveal the complexity of defining a simple consensus sequence for Lin28.

Although previous biochemical and structural studies have described three potential recognition sites for Lin28 on the terminal loop of pre-let-7 RNAs, most studies report a 1:1 stoichiometry for the Lin28/TL-let-7 complex (15,17-19,51-53). Based on the observation of native gel supershifts in our previous Lin28-binding studies with TL-let-7g, we proposed that multiple molecules of Lin28 likely interact with pre-let-7g to fulfill its inhibitory function on let-7 biogenesis (19). More recently, such supershifts were also observed as a result of Lin28 binding to pre-let-7a-1 (55). In this present work, we characterize the mechanistic details for multimeric binding of Lin28 to pre-let-7 targets. We define three distinct Lin28 binding sites in the terminal loop of pre-let-7g (TL-let-7g) and describe the stepwise assembly of three molecules of Lin28 to a single molecule of TL-let-7g. Furthermore, we define the respective

roles of the CSD and NCp7-like domain of Lin28 in the formation of this multimeric complex and demonstrate its specificity towards terminal loops of the let-7 family of pre-miRNAs. Based on these results, we propose a model for stepwise assembly of Lin28 to the terminal loop of pre-let-7g and investigate how the processing enzyme Dicer intervenes in this process.

Materials and methods

RNA and protein preparation

All unlabeled RNAs longer than 14 nucleotides and all proteins used in this study are derived from murine sequences and were prepared as described previously (19), except for pre-let-7 RNAs used in the Dicer processing assay. Shorter RNAs and RNAs with fluorescent labels [5'-Cy3, 5'-Cy5 or internal 2-aminopurine (2-AP)] were obtained from Integrated DNA Technology (IDT). For radiolabeling of RNAs, [5'-³²P]-labeling and subsequent purification were performed as described previously (57). For the Dicer processing assay, the mono-uridylated human pre-let-7g (pre-let-7g-U) was purified from a CRISPR-pre-let-7g-U-ARiBo precursor (58), whereas the mono-uridylated human pre-let-7d (pre-let-7d-U) and pre-let-7a-1 (pre-let-7a-1-U) were purified from cis-cleavage of HH-pre-let-7-HDV precursors (59). For radiolabeling of RNAs, [5'-³²P]-labeling and subsequent purification were performed as described previously (60). The unlabeled pre-let-7g-U, pre-let-7d-U and pre-let-7a-1-U RNAs used in the Dicer processing assay were phosphorylated and purified similarly using non-radioactive ATP.

Determination of dissociation constants (K_d) by electrophoretic mobility shift assay

For binding studies by electrophoretic mobility shift assay (EMSA), protein and ³²P-labeled RNA samples were prepared in EMSA buffer (50 mM Tris pH 7.6, 50 mM NaCl, 10% glycerol, 0.05% NP-40 alternative and 2 mM DTT) and the binding reaction was performed as described previously (19). Binding reactions were loaded directly on a standard 8% native gel [37.5:1 polyacrylamide/bisacrylamide with Tris-Glycine buffer (25 mM Tris-Base and 200 mM glycine)] run at 200 V for 1 h with active water cooling at 4°C. The gels were then dried and exposed overnight on a storage phosphor screen (Bio-Rad). The ³²P-labeled RNA was visualized with a Molecular Imager FX densitometer or a Personal Molecular Imager system (both from Bio-Rad), and the band intensities were quantified using ImageLab (version 4.1 from Bio-Rad was used throughout this study). The fraction of bound RNA was plotted

against protein concentration and the data were fitted to the one-site binding equation or the Hill equation by non-linear regression analysis with OriginPro 8 (OriginLab).

Stoichiometric binding assay by native gel electrophoresis

For the stoichiometric binding assay, 5- μ M RNA samples (or 10X the final RNA concentration) were prepared by combining unlabeled RNA with 100 pM of 32 P-labeled RNA, and these samples were heated and snap-cooled (heated at 95°C and snap-cooled on ice for 5 min) to promote hairpin formation. The protein samples were first diluted in EMSA buffer to the appropriate concentrations and the binding reactions (20 μ L) were initiated by adding 500 nM of RNA (or lower concentrations for results in Supplementary Figure 3-2A). For each stoichiometric binding assay, binding reactions were incubated at 4°C for 30 min and loaded directly on an 8% standard native gel run at 200 V for 2 h with active water cooling at 4°C. The gels were then dried, exposed and visualized as described for K_d determination. The fractions of RNA present in the bands of the 1:1, 1:2 and 1:3 complexes were quantified with ImageLab, and reported scores in Table 3-II were obtained from at least two independent experiments.

2-Aminopurine fluorescence assay

Several TL-let-7g RNAs carrying a single adenine to 2-aminopurine (2-AP) substitution were used for the 2-AP fluorescence assay. Each 2-AP-modified RNA was first heated and snap-cooled to promote hairpin formation. The emission spectra of the 500-nM solution of 2-AP-modified RNA (or unmodified RNA used as a control) were recorded from 335/20 nm (335 nm with an emission slit of 20 nm) to 425/20 nm after excitation at 300/5 nm first in absence of protein and then 1 min after each protein addition from a concentrated stock. All experiments were performed with a Varian Cary Eclipse fluorimeter at 4°C in 50 mM Tris pH 7.6, 50 mM NaCl and 10% glycerol. Each difference emission spectrum was obtained by subtracting the emission spectrum of the sample containing the 2-AP-modified RNA with that containing unmodified RNA collected under the same protein concentration

and buffer conditions. For each titration, the fluorescence emission intensity at 370 nm was normalized with respect to the highest fluorescence emission intensity observed at 370 nm for that specific titration experiment. The normalized fluorescence intensity at 370 nm was plotted as a function of protein concentration, and the data were fitted to the dose-response equation $[y=A_1+(A_2-A_1)/(1+10^{((\log x_0-x)*p))}]$, where A_1 is the bottom asymptote, A_2 the top asymptote, x_0 the EC_{50} and p the Hill slope] by non-linear regression analysis with OriginPro 8.

Strand displacement assay monitored by fluorescence resonance energy transfer

Three different fluorophore-labeled (5'-Cy3 or 5'-Cy5) RNAs were used for the strand displacement assay by fluorescence resonance energy transfer (FRET). The forward strand (Cy5-FWD: 5'-Cy5-CGU ACA GAU UGA GGG UGA CAU CG-3') was annealed to a perfectly complementary reverse strand (Cy3-REV_{comp}: 5'-Cy3-CGA UGU CAC CCU CAA UCU GUA CG-3') to form the complementary duplex (duplex_{comp}) and to a partially complementary reverse strand (Cy3-REV_{bulge}: 5'-Cy3-CGA UGU CAU AUC UGU ACG-3') to form a duplex with a G-rich bulge (duplex_{bulge}). Solutions containing 25 nM of both forward and reverse strands were heated and slow-cooled (heated 2 min at 95°C and slow-cooled at room temperature for 20 min) to promote duplex formation. The strand displacement was monitored by FRET upon addition of a concentrated stock of protein to a solution of Cy3/Cy5-labeled RNA duplex in EMSA buffer. Following protein addition, the sample was equilibrated at 30°C for 1 min and fluorescence emission data were recorded at 30°C on a Varian Cary Eclipse fluorimeter following an excitation of Cy3 at 535/20 nm. The Fret Index (F_{Cy5}/F_{Cy3}) was calculated from the emission of Cy5 at 670/10 nm (F_{Cy5}) and Cy3 at 590/10 nm (F_{Cy3}), and this Fret Index was normalized with respect to the Fret Index obtained in the absence of protein. The $\Delta FRET_{max}$ value corresponds to the maximum difference in normalized Fret Index. The normalized Fret Index was also plotted against total protein concentration and the data were fitted to the dose-response equation by non-linear regression analysis with OriginPro.

Strand displacement assay monitored by native gels

This assay was performed with RNA duplexes formed by combining fluorophore-labeled RNAs (5'-Cy3 or 5'- Cy5) following by heating and slow cooling. As controls, individual fluorophore-labeled RNAs were also investigated, and those were prepared by heating and snap cooling. The protein binding reactions were performed at 25 nM or 250 nM RNA concentrations and separated on gels as described for the stoichiometric binding assay. The fluorophore-labeled RNAs were detected with the ChemiDoc MP system (Bio-Rad) set up for multiplex detection of Cy3 and Cy5 fluorophores.

Dicer processing assay

Dicer cleavage reactions were performed as described for the stoichiometric binding assay but using a total volume of 20 μ L containing 50 mM Tris pH 7.6, 250 mM NaCl, 2.5 mM MgCl₂, 10% glycerol, 0.05% NP-40, 2 mM DTT, 500 nM unlabeled pre-let-miRNA, 40 pM ³²P-labeled pre-miRNA, variable concentrations of Lin28 and 0.25 U of recombinant Dicer (Genlantis). These reactions were incubated 10 min on ice and then 1 h at 37°C. Afterwards, 0.05 U of proteinase K (Roche) was added, samples were incubated 15 min at 37°C and the reactions were stopped by adding 40 μ L of gel loading buffer (87% formamide, 25 mM EDTA, 0.02 % xylene cyanol, 0.02 % bromophenol blue). Samples (15 μ L) were loaded on a 10% polyacrylamide (19:1 polyacrylamide/bisacrylamide)/7 M urea gel, run at 500 V for 45 min and then dried, exposed and visualized as described for K_d determination. Band intensities for the full-length and cleaved pre-miRNA were quantified using ImageLab and used to derive the percentages of pre-miRNA cleavage. The relative Dicer processing efficiency is the percentage of pre-miRNA cleavage in the presence of Lin28 over that in the absence of Lin28.

Error Analysis

For K_d determination, 2-AP fluorescence and strand displacement assays, at least three independent binding experiments were performed. Reported values (K_d , EC_{50} , Fret Index,

$\Delta\text{FRET}_{\text{max}}$) and their errors are respectively the average values and the standard deviations from the multiple experiments.

Results

Major variants of the pre-let-7g terminal loop maintain high-affinity binding to Lin28

We previously determined by electrophoretic mobility shift assay (EMSA) that Lin28 binds the terminal loop of pre-let-7g (TL-let-7g) with a very high affinity [$K_d = 0.13$ nM; (19)]. The binding data were best fitted using the Hill equation, with a Hill coefficient (n) of 2.9. Given that supershifts were also observed in our EMSA at high concentrations of Lin28, we hypothesized that multiple molecules of Lin28 could bind a single molecule of TL-let-7g (19). To better understand the determinants of TL-let-7g binding to Lin28, we performed our EMSA with three variants of TL-let-7g (Figure 3-1B) in which key secondary structure elements were modified, namely the 5' G-rich bulge (Δ bulge), the internal loop (Δ loop) and the hairpin loop (GNRA loop). Remarkably, these variations do not greatly affect the binding affinity of Lin28 to TL-let-7g (Table 3-I), since the averaged K_d values increase by no more than seven fold relative to the wild-type sequence. The largest effect is observed with the Δ bulge variant (K_d of 0.9 nM) previously shown to abolish the TL-let-7g binding of a variant of Lin28 containing only the NCp7-like domain [Lin28₁₁₉₋₁₈₀; (19)]. The lack of a similar effect with the full-length protein concur with the hypothesis that Lin28 has the ability to bind multiple sites on TL-let-7g and that the variants being studied may destroy and/or conceal one or more of the Lin28 binding sites, but still allow high-affinity Lin28 binding to other sites.

The terminal loop of pre-let-7g comprises three distinct binding sites for Lin28

To precisely identify the different binding sites of Lin28 on TL-let-7g, we performed additional binding assays by EMSA using 14-nt RNA fragments derived from the TL-let-7g. These short RNAs were designed to cover the entire terminal loop of pre-let-7g with partial overlaps and are named according to the matching segment of residues within TL-let-7g (Table 3-I and Figure 3-1B). For example, TL5-18 containing the sequence of the 5' G-rich bulge of TL-let-7g, residues 5 to 18, was investigated for Lin28 binding by EMSA (Supplementary Figure 3-1A). The plot of the bound RNA fraction versus Lin28 protein

concentration fits the one-site binding equation (Supplementary Figure 3-1B) to yield a K_d of 1.7 nM (Table 3-I). Binding of Lin28 to a shorter fragment derived from TL5-18 (TL6-13) was not detectable at protein concentrations up to 5 μ M by EMSA (data not shown), suggesting that a minimal RNA fragment length, between 9-14 nt, is necessary for high-affinity binding. Amongst the five short RNAs with native sequences tested (TL5-18, TL13-26, TL19-32, TL28-41, TL33-36), only one (TL19-32) did not bind Lin28 by EMSA (Table 3-I). In addition to TL5-18, subnanomolar binding affinity is observed for TL28-41 ($K_d = 0.29$ nM) and TL33-46 ($K_d = 0.5$ nM), whereas nanomolar binding affinity is observed for TL13-26 (25-75 nM). The binding data with these four RNAs could be fitted to the one site binding equation and only a single shifted band is observed by EMSA, indicating that these short RNAs contain a single binding site for Lin28.

Interestingly, these four short RNAs that bind Lin28 with high affinity contain previously-identified binding sites for Lin28 (in bold in Table 3-I). First, the TL5-18 RNA contains the 5' G-rich bulge recently identified as a primary binding site for Lin28 and its NCp7-like domain in pre-let-7g (18,19). Also, both TL28-41 and TL33-46 contain the 3'-GGAG sequence defined as a key determinant of Lin28 binding and function targeted specifically by the NCp7-like domain (16,17,51-53). To confirm that the two G-rich sequences in these short RNAs were important for the observed high-affinity binding, we tested the binding of short RNA variants with nucleotide changes of key G residues (17-19,51-53). As expected, replacement of key guanines in TL5-18 and TL28-41, either abolished (TL5-18V) or greatly reduced (TL28-41V) Lin28 binding (Table 3-I). Despite the importance of these G residues, no binding could be observed in our assay between the NCp7-like domain (Lin28₁₁₉₋₁₈₀) and the G-rich fragments (TL5-18, TL28-41 or TL33-46; data not shown), indicating that the CSD is required for binding these short RNAs. Finally, TL13-26 contains the previously-identified AUGAUAC sequence recognized by the CSD of Lin28 (51,52). Again, nucleotide changes within TL13-26 at key positions for CSD recognition (51) prevent binding to Lin28 (Table 3-I; TL13-26V). In summary, by using short and most likely single-stranded RNAs, three distinct Lin28 binding sites were identified on TL-let-7g. Since most of these sites do not adopt a single-stranded conformation in the context of TL-let-7g (18,19), it is not clear yet if

they all are accessible to allow binding of multiple molecules of Lin28 to a single molecule of TL-let-7g.

Stepwise assembly of Lin28 on the terminal loop of pre-let-7g yields a stable 1:3 complex

To determine if multiple molecules of Lin28 bind one molecule of TL-let-7g, we optimized a stoichiometric binding assay. In this assay, an increasing amount of protein (0.1X to 10X the RNA concentration) is added to a non-negligible quantity (500 nM) of RNA. If multiple molecules of Lin28 bind TL-let-7g in a non-concerted fashion, supershifts will be observed on a native gel. In fact, the stoichiometric binding assay of Lin28 to TL-let-7g (Figure 3-2, top left panel) results in the sequential appearance of a band shift, a supershift, and a super-supershift that are almost fully populated at RNA:protein ratios of 1:1.5, 1:2.5 and 1:3.5, respectively. The mobility of the free RNA band and that of the first shift are invariable at RNA concentrations ranging from 1 pM to 0.5 μ M (data not shown), which strongly indicates that TL-let-7g forms a monomeric hairpin in its free and bound forms. Furthermore, 1:1 binding was previously observed for similar complexes by sedimentation equilibration ultracentrifugation (51), size-exclusion chromatography and mass spectrometry (61) at equimolar concentrations of RNA and protein. Thus, the observed shifts most likely correspond to RNA:protein stoichiometries of 1:1, 1:2 and 1:3. Additional molecules of Lin28 may associate transiently to the 1:3 complex, as indicated by smearing of the 1:3 complex band at the highest RNA:protein ratios (1:5 and 1:10). Remarkably, the largest stable multimeric complex (designated the 1:3 complex) is assembled in a stepwise manner; the free RNA must be completely shifted in a 1:1 complex, prior to formation of the 1:2 and 1:3 complexes.

This stoichiometric binding assay was also performed at lower concentration of TL-let-7g (25, 50 and 250 nM; Supplementary Figure 3-2A). The 1:1, 1:2 and 1:3 complexes form in all cases, but higher protein:RNA ratios are needed as the RNA concentration is reduced. Formation of stable multimeric complexes was also tested with the parental pre-let-7g at 500 nM RNA (Supplementary Figure 3-2B). Compared to results with TL-let-7g, somewhat higher

protein:RNA ratios are needed for formation of the 1:1 and 1:2 and 1:3 complexes, and the latter is more diffuse, possibly due to the stabilizing effect of the longer stem of pre-let-7g or to its contribution to non-specific Lin28 binding.

Both RNA-binding domains of Lin28 contribute to its stepwise assembly on TL-let-7g

To examine the individual roles of the CSD and NCp7-like domain of Lin28 in the assembly of the 1:3 complex, stoichiometric binding assays were repeated with Lin28 variants containing only a single functional RNA domain (Figure 3-1A, Figure 3-2 and Table 3-II). The Lin28₁₁₉₋₁₈₀ variant, which only contains the NCp7-like domain known to target the 5' G-rich bulge (19), forms a 1:1 complex with TL-let-7g (Figure 3-2, top right panel) with an affinity similar to the full-length protein, but forms 1:2 and 1:3 complexes with TL-let-7g very inefficiently and in a non-orderly fashion. In contrast, the Lin28 C139A/C161A variant, which contains a functional CSD but a dysfunctional NCp7-like domain, can assemble to form 1:1, 1:2 and 1:3 complexes (Figure 3-2, top middle panel), but at higher protein concentrations than for the wild-type Lin28, such that it prevents formation of the 1:3 complex at an RNA:protein ratio of 1:3.5 (Table 3-II). Given that the binding affinity for TL-let-7g of the NCp-7 like domain ($K_d = 1.3$ nM) is higher than for the CSD [$K_d = 126$ nM; (19)], it is not surprising to find in our stoichiometric assay that Lin28 C139A/C161A is slightly more deficient at forming the 1:1 complex than the NCp7-like domain (see Table 3-II). Thus, the NCp7-like domain is important for formation of a high-affinity 1:1 complex and it facilitates assembly of 1:2 and 1:3 complexes, but the intrinsic ability of Lin28 to efficiently multimerize on TL-let-7g is imparted by the CSD.

Interestingly, stoichiometric binding of one or two equivalents of Lin28 C139A/C161A to TL-let-7g is not affected by the binding of one equivalent of Lin28₁₁₉₋₁₈₀ (Supplementary Figure 3-3), but rather lead to formation of 1:1:1 and 1:1:2 RNA:Lin28₁₁₉₋₁₈₀:Lin28 C139A/C161A complexes as indicated by the gel mobility of the complexes. Thus, the Lin28₁₁₉₋₁₈₀ and Lin28 C139A/C161A variants target different residues in their respective 1:1

complex, in agreement with the different sequence preference for the NCp7-like domain and the CSD (51,57).

We also investigated stoichiometric binding to two TL-let-7g variants [Figure 3-1B; (19)]. The TL-let-7g Δ loop variant was previously shown to have a minor effect on binding of Lin28₁₁₉₋₁₈₀, in agreement with the fact that Lin28₁₁₉₋₁₈₀ mainly recognizes the 5' G-rich bulge (19). Similarly, this variant reduces to a limited extent the ability of Lin28₁₁₉₋₁₈₀ to form a 1:1 complex (Figure 3-2 and Table 3-II). In addition, it reduces the ability of both Lin28 and the Lin28 C139A/C161A variant to form 1:1, 1:2 and 1:3 complexes (Figure 3-2 and Table 3-II). Given that both the 5'-AUGAUAC-3' and 5'-GGAG-3' binding sites are destroyed in the Δ loop variant, the limited formation of the 1:2 and 1:3 complexes with these proteins likely results from the low specificity of the CSD (52). Deletion of the 5' G-rich bulge (Δ bulge; Figure 3-1B), previously identified as the main binding site for Lin28₁₁₉₋₁₈₀ (19), has a larger effect on stoichiometric binding. As expected, the Δ bulge variant strongly reduces the ability of Lin28₁₁₉₋₁₈₀ to form a 1:1 complex (Figure 3-2, bottom panels and Table 3-II). In addition, the Δ bulge variant significantly reduces the ability of Lin28 and Lin28 C139A/C161A to form 1:1, 1:2 and 1:3 complexes. Lin28 forms 1:1 and 1:2 complexes with the Δ bulge variant, but not the 1:3 complex at an RNA:protein ratio of 1:3.5 (Table 3-II), in agreement with the loss of a high-affinity binding site on the RNA. Comparatively, Lin28 C139A/C161A is more deficient than Lin28 at forming 1:1 and 1:2 complexes with the Δ bulge variant and even loses its ability for stepwise assembly. These results support the idea that the NCp7-like domain and the 5' G-rich bulge both play an important role in initiating the ordered assembly of the multimeric complexes, as previously suggested (19).

Lin28 specifically forms stable 1:3 complexes with terminal loops of other miRNA precursors from the let-7 family

Stoichiometric binding was also investigated between Lin28 and the terminal loops of other miRNA precursors to determine if, like TL-let-7g, these terminal loops have the ability for stepwise assembly of a stable 1:3 complex. Interestingly, both TL-let-7a-1 and TL-let-7d can form 1:1, 1:2 and 1:3 complexes, although the stepwise assembly of the 1:2 and 1:3

complexes requires higher Lin28 concentration than for TL-let-7g (Figure 3-3 and Table 3-II). Both TL-let-7a-1 and TL-let-7d contain 5' and 3' G-rich sequences (in bold in Figure 3-3AB), and TL-let-7d also contains an established CSD binding site [the 5'-AGAGAUUUU-3' sequence in bold in Figure 3-3B; (51,52)]. For TL-let-7d, the highly-accessible 5'-GGAG-3' binding site and the partially-accessible 5'-AGAGAUUUU-3' site likely contribute to formation of a high-affinity 1:1 complex, as observed in the X-ray crystal structure [Figure 3-3D; (51,52)]. Binding to Lin28 variants is similar for TL-let-7a-1 and TL-let-7d compared to TL-let-7g, with the Lin28₁₁₉₋₁₈₀ variant binding with high affinity to at least one site and the Lin28 C139A/C161A variant allowing stepwise assembly of multimeric complexes (Figure 3-3), but at higher protein concentration than Lin28.

To further examine the specificity for stepwise assembly of multimeric complexes between Lin28 and TL-let-7 RNAs, the binding of Lin28 was also investigated with the terminal loops of miR-99b (TL-miR-99b, Figure 3-3C) and miR-21a (TL-miR-21a, Supplementary Figure 3-4A). It was previously demonstrated that the biogenesis of miR-99b and miR-21a is unaffected by the expression or the silencing of Lin28 (15,16,24). In agreement with these results, we measured K_d values of 15 nM for TL-miR-99b and 13 nM for TL-miR-21a for formation of a 1:1 complex with Lin28 (data not shown), which represent a two-order of magnitude weaker binding than for TL-let-7g [$K_d = 0.13$ nM; (19)]. Furthermore, in our stoichiometric binding assay Lin28 has a reduced ability to form 1:1 and 1:2 complexes with TL-miR-99b compared to TL-let-7g and does not form a 1:3 complex under the conditions being investigated (Figure 3-3D). Similar results were obtained with TL-miR-21a (Supplementary Figure 3-4B). Furthermore, stoichiometric binding of the Lin28₁₁₉₋₁₈₀ and Lin28 C139A/C161A variants is severely reduced with TL-miR-99b compared with TL-let-7g (Figure 3-3D). Taken together, these results indicate that stepwise assembly is markedly impaired with TL-miR-99b and TL-miR-21a relative to TL-let-7 RNAs.

Dynamic assembly of Lin28 on the terminal loop of pre-let-7g

To better understand the dynamic assembly of Lin28 on TL-let-7g RNA, we carried out a fluorescence assay using a series of TL-let-7g RNAs containing single adenine to

2-aminopurine (2-AP) modifications (2-AP5, 2-AP9, 2-AP21, 2-AP31 and 2-AP36; Figure 3-4B). The fluorescence emission of 2-AP is highly sensitive to the immediate environment of the base analogue, particularly stacking interactions that can substantially quench the high quantum yield of free 2-AP (62-66). Therefore, 2-AP is an ideal reporter for changes in RNA structure and dynamic resulting from protein binding. Here, the fluorescence emission spectra of the modified TL-let-7g were first collected in the absence of proteins using the same detection parameters (Supplementary Figure 3-5A). The low fluorescence of 2-AP5 is consistent with this residue being part of a stable stem, whereas the higher fluorescence of the other 2-AP-modified TL-let-7g RNAs indicates that these modified residues (A9, A21, A31 and A36) are part of more flexible regions in the RNA. Following the addition of Lin28, 2-AP21 undergoes a significant decrease in fluorescence intensity with a measured EC_{50} of 0.50 μ M [Figure 3-4 (blue diamonds) and Supplementary Figure 3-5D]. This result is consistent with stabilization of 2-AP stacking, most likely as a result of direct binding with Lin28 since A21 is located in a previously-identified CSD binding site (5'-AUGAUAC-3'). According to the crystal structure of a complex between derivatives of Lin28 and TL-let-7g, A21 should stack between Trp46 and C22 when bound to the CSD [pdb code: 3TS2; (51)]. The other 2-AP-modified TL-let-7g RNAs undergo fluorescence increases upon addition of Lin28 [Figure 3-4 and Supplementary Figure 3-5], either through a direct interaction with Lin28 or through destabilization of the RNA structure. Given that the EC_{50} values of 0.5-0.6 μ M for 2-AP9, 2-AP21 and 2-AP36 nearly match the point of half binding for the first shift observed with the stoichiometric binding assay (Figure 3-2), these results indicate that binding of the first protein affects the local environment of multiple nucleotides of TL-let-7g. Furthermore, the EC_{50} of 1.3 μ M for 2-AP31 [Figure 3-4 (red triangles) and Supplementary Figure 3-5E] is consistent with binding of a second molecule of Lin28 on TL-let-7g. Only a small increase in fluorescence intensity was observed for 2-AP5 at high Lin28 concentrations, indicating that the A5-U41 base pair of TL-let-7g remains stable until the formation of the 1:3 complex [Figure 3-4 (cyan dots) and Supplementary Figure 3-5B]. In summary, formation of the 1:1 complex significantly remodels the internal loop by affecting residues within the three Lin28 binding sites, whereas formation of the 1:2 and 1:3 complexes appear to have more localized effects on individual Lin28-binding sites.

The cold shock domain is responsible for the RNA melting activity of Lin28

To investigate if Lin28 carries an RNA melting activity responsible for destabilizing the TL-let-7g upon formation of the 1:1 complex, we adapted an established fluorescence-based strand displacement assay [Figure 3-5A; (67)]. This assay relies on determining the FRET index (F_{Cy5}/F_{Cy3}) following Cy3 excitation for an RNA duplex composed of 5'-Cy5-labeled and 5'-Cy3-labeled strands. The fluorescence emission of Cy5 is strongly dependent on the annealing of the two strands, such that the decrease in fluorescence intensity upon addition of protein infers RNA melting activity. Here, two different RNA duplexes were used; one contains a G-rich bulge mimicking the 5'-bulge of TL-let-7g ($\text{duplex}_{\text{bulge}}$) and the other is the equivalent duplex without the bulge such that the two strands are perfectly complementary ($\text{duplex}_{\text{comp}}$).

The strand displacement assay performed with 25 nM $\text{duplex}_{\text{bulge}}$ reveals the RNA melting activity of both Lin28 and the Lin28 C139A/C161A variants, with respective $\Delta\text{FRET}_{\text{max}}$ of 0.84 and 0.77 and EC_{50} of 58 nM and 354 nM (Figure 3-5B-C). The fluorescence emission profiles clearly demonstrate that the decrease in FRET index upon addition of Lin28 is caused by an emission increase from Cy3 combined with an emission decrease from Cy5, which is consistent with the separation of the two fluorophores (Supplementary Figure 3-6AC). Furthermore, the stoichiometric binding assay performed with $\text{duplex}_{\text{bulge}}$ at 25 nM and 250 nM provides further evidence for dissociation of the $\text{duplex}_{\text{bulge}}$ in two separate strands as a result of addition of Lin28 (Supplementary Figure 3-7 and 8). For the Lin28₁₁₉₋₁₈₀ variant, the strand displacement assay yields a smaller $\Delta\text{FRET}_{\text{max}}$ value of 0.24 with an EC_{50} of 99 nM, which is indicative of a specific binding event at the bulge that moves the fluorophores further apart relative to each other without displacing the strands, although strand displacement may occur to a limited extent (Figure 3-5 and Supplementary Figure 3-6E, S7E and S8E). Thus, the RNA melting activity of Lin28 is present at higher protein concentration for the Lin28 C139A/C161A variant, but essentially absent for the Lin28₁₁₉₋₁₈₀ variant under all tested conditions. In contrast, Lin28 and its derivatives do not melt the perfectly complementary RNA duplex ($\text{duplex}_{\text{comp}}$), since no decrease in FRET index was

observed (Figure 3-5 and Supplementary Figure 3-6) indicating that the G-rich bulge is essential for this RNA melting activity. In summary, the strand displacement assay demonstrates that the RNA melting activity of Lin28 belongs predominantly to its CSD, but is enhanced by the NCp7-like domain of Lin28 due to its high-affinity binding to the G-rich bulge.

Formation of the 1:1 complex is not always sufficient for maximum inhibition of Dicer cleavage in vitro

Lin28 is known to inhibit the cleavage activity of Dicer on let-7 precursors, but it is not clear where this inhibition takes place in the stepwise assembly of Lin28 on pre-let-7g. To clarify this issue, we tested the effect of Lin28 concentration on Dicer cleavage of pre-let-7g-U, an optimal Dicer substrate containing a 5'-phosphate (68) and mono-uridylylated at its 3'-end [(Figure 3-6A; (30)]. In parallel, we performed a stoichiometric binding assay with pre-let-7g-U under the same high salt conditions as those used for Dicer cleavage (Figure 3-6B). A 60% reduction of Dicer cleavage (from ~80% to ~30% cleavage) is observed at Lin28 concentrations that allow formation of the 1:1 complex (Figure 3-6C-D). However, further increase in Lin28 concentration to allow formation of the 1:2 and 1:3 complexes do not further reduce Dicer cleavage. These results indicate that binding of one molecule of Lin28 to pre-let-7g-U is sufficient to significantly inhibit Dicer cleavage, and that this level of inhibition is maintained upon binding of additional molecules of Lin28.

Similar Dicer processing assays were performed with pre-let-7d-U and pre-let-7a-1-U, two additional Dicer substrates containing a 5'-phosphate (68) and mono-uridylylated at their 3'-end [Supplementary Figure 3-9A and Supplementary Figure 3-10A; (30)]. Similarly to pre-let-7g-U, a 70% reduction of Dicer cleavage (from ~91% to ~24% cleavage) is observed for pre-let-7d-U at a Lin28 concentration that allows for formation of the 1:1 complex, and further increases in Lin28 concentration that allow for formation of the 1:2 and 1:3 complexes do not further reduce Dicer cleavage ([Supplementary Figure 3-9C-D). In contrast, Dicer cleavage of pre-let-7a-1-U is reduced by 10% at a Lin28 concentration that allows for formation of the 1:1 complex, and is further reduced by 35% and 45% with Lin28 concentrations that allow for

formation of the 1:2 and 1:3 complexes, respectively. Thus, although binding of one molecule of Lin28 to pre-let-7g-U and pre-let-7d-U is sufficient for maximum Dicer cleavage, multimerization of Lin28 is required for maximum Dicer inhibition with pre-let-7a-1-U as the substrate.

Discussion

In this work, we elucidate the molecular mechanism by which Lin28 interacts with the terminal loop of pre-let-7g. This binding mechanism, which involves stepwise assembly of three molecules of Lin28 on the pre-miRNA terminal loop, is detailed below in regard to an assembly model that recapitulates results presented here and helps clarify what was thought to be conflicting data associated with previous studies. In addition, we examine the significance of this stepwise assembly for in vivo regulation of pre-let-7 biogenesis.

Binding of Lin28 to short RNAs

Using short 14-nt RNAs derived from TL-let-7g, we identified three distinct Lin28 binding sites on TL-let-7g. Fragments containing a conserved G-rich sequence bind with very high affinity to Lin28, those containing the 5'-GGAG-3' motif binding with somewhat higher affinity (K_d of 0.3-0.5 nM) than those containing the 5'-GAGGG-3' motif (K_d of 1.7 nM). These binding data are in agreement with previous studies in which short conserved G-rich elements located at both the 5'-end [5'-UGAGGG-3'; (18,19)] and the 3'-end [5'-GGAG-3'; (17,52)] of the terminal loop of pre-let-7g were individually defined as the main determinants of Lin28 binding. Short G-rich elements are also predominantly found in target sequences of Lin28-associated mRNAs as part of genome-wide studies (54-57,69). In our study with 14-nt RNA fragments, mutations of key G residues abolished high-affinity binding to Lin28, as expected for recognition by the NCp7-like domain. High-affinity binding was lost when we replaced the 14-nt TL5-18 fragment by an 8-nt 5'-UGAGGGUC-3'. Surprisingly, high-affinity binding to 14-nt G-rich fragments was also lost when we used a Lin28 variant containing just the NCp7-like domain, even though the NCp7-like domain recognizes the G-rich bulge with high affinity as part of TL-let-7g [$K_d = 1.3$ nM;(19)]. Binding of the NCp7-like domain to TL-let-7g may be favored by the presence of nearby helical elements or the unique structure of the bulge. In agreement with our results, low affinity binding ($K_d = 0.4$ μ M) was observed between an isolated NCp7-like domain and a G-rich RNA heptamer by isothermal titration calorimetry (53). Clearly, high-affinity binding of Lin28 to these 14-nt fragments is not

limited to the NCp7-like domain, but likely involves the CSD, which is known to bind short RNAs (6-9 nt) with low-sequence specificity (52). Thus, both the NCp7-like domain and the CSD contribute to recognition of these G-rich sites by Lin28, and we defined a minimum length between 9-14 nt for high-affinity binding to G-rich fragments derived from pre-let-7g.

The TL13-26 fragment containing the 5'-AUGAUAC-3' region from TL-let-7g binds Lin28 with high affinity (K_d of 25-75 nM). The 5'-AUGAUAC-3' sequence directly interacts with the CSD in a crystal structure of a complex formed by Lin28 and pre-let-7g derivatives (51), and its importance for CSD binding is supported here by the loss of binding observed with the TL13-26 variant containing the 5'-AUCAAAC-3' mutation. These results are in agreement with the ability of the CSD of Lin28B from *Xenopus tropicalis* (*Xtr*) to bind RNA fragments of 6-9 nts derived from the *Xtr*-pre-let-7 terminal loop with low nanomolar affinity (52). Given this observation and the fact that TL13-26 does not contain a GNG or GNNG sequence (19,53), binding to this fragment may only involve the CSD of Lin28.

Model for stepwise assembly of Lin28 on pre-let-7g

Based on available data, we propose a model for stepwise assembly of Lin28 on the terminal loop of pre-let-7g (Figure 3-7). In this model, the first molecule of Lin28 initiates complex formation through a high-affinity interaction between its NCp7-like domain and the 5'-GAGGG-3' site at the 5'-bulge, which is followed by binding of its CSD to the nearby 5'-AUGAUAC-3' site. Both the 5'-bulge and the NCp7-like domain have previously been identified as important determinants of the pre-let-7g/Lin28 interaction (18,19). Our stoichiometric binding assay further reveals that an intact 5'-bulge on pre-let-7g and the NCp7-like domain of Lin28 are essential for formation of a stable 1:1 complex and that disruption of both leads to random assembly of multimeric complexes. This supports the idea that they initiate the ordered assembly of the 1:1, 1:2 and 1:3 complexes, as previously suggested (19). The concept that the CSD contacts the 5'-AUGAUAC-3' site (residues 16-21) in the 1:1 complex is supported by the high-affinity of Lin28 towards the TL13-26 fragment containing this sequence as well as fluorescence quenching of 2-AP21 associated with CSD binding upon formation of a 1:1 complex (49). Furthermore, the proposed model for the 1:1 complex

matches remarkably well with RNase protection studies, where the 5'-GAGGG-3' and the 5'-AUGAUAC-3' sites are the most protected regions of pre-let-7g upon binding of Lin28 (18). Interactions at these two RNA sites in the 1:1 complex occurs with the CSD and NCp7-like domain arranged in a 3' to 5' orientation that may not be preferred for Lin28 (51,57), but that may be dictated by the availability and high affinity of these sites on pre-let-7g. Binding at these sites destabilize Watson-Crick base pairs within the terminal loop, as previously observed (18), since both RNA domains are known to interact with single-stranded RNAs. Results from our 2-AP assay are also in agreement with RNA destabilization upon formation of the 1:1 complex, whereas our FRET strand-displacement assay supports an RNA melting activity for Lin28, as previously suggested (52).

This terminal loop destabilization caused by the binding of the first Lin28 protein exposes a previously concealed high-affinity binding site on the 3'-strand, making it available for Lin28 binding (Figure 3-7). We propose that the NCp7-like domain of the second protein binds the conserved 5'-GGAG-3' sequence, whereas its CSD binds an adjacent 5'-region, possibly the 5'-CGGUAC-3' sequence, which shares a 5'-YGRUAC-3' motif with the well-characterized 5'-AUGAUAC-3' CSD binding sites of TL-let-7g (52). Binding of a second molecule of Lin28 at this proposed site is consistent with our 2-AP fluorescence assay, in which a Lin28-dependent change in 2-AP31 fluorescence correlates with formation of the 1:2 complex. We propose that the third molecule of Lin28 binds the high-affinity 5'-AUGAUAC-3' site through its CSD. This would require the CSD of the first molecule of Lin28 to relocate, which is conceivable given its low sequence specificity and the flexibility of the linker region between the CSD and the NCp7-like domain (51,52). Furthermore, such repositioning of the first molecule of Lin28 is consistent with the increase in 2-AP5 fluorescence associated with formation of the 1:3 complex.

The proposed assembly mechanism may differ for precursors of other let-7 family members. For example, formation of the 1:1 complex with pre-let-7d likely involves the exposed 5'-GGAG-3' binding site at the 3'-end of TL-let-7d, as previously reported (51,52), rather than the concealed 5'-AGGG-3' binding site at the 5'-end of TL-let-7d. Moreover, multimeric assembly could also occur through alternative and likely less important pathways,

as evidence from the reduced ability of the C139A/C161A variant to form multimeric complexes with TL-let-7g.

The stepwise assembly proposed here underlines the ability of Lin28 to efficiently multimerize on pre-let-7 to form a 1:3 complex, previously noted with pre-let-7g (19) and pre-let-7a-1 (55). One could argue that the observed stoichiometry contradicts previous reports, where 1:1 binding was observed between mouse Lin28 and TL-let-7d by sedimentation equilibration ultracentrifugation experiments (51) and between human Lin28B and pre-let-7g using both size-exclusion chromatography and mass spectrometry (61). However, these previous studies were not performed under conditions where the protein is in excess of the RNA (51,61) and that allow binding of multiple copies of Lin28 to pre-let-7.

The stepwise assembly recapitulates other important aspects of Lin28 function. It involves the contribution of both RNA domains of Lin28, in agreement with their importance for binding target RNAs *in vitro* and for *in vivo* function (15,17,19,24,25,28,51,53,57,69,70). The NCp7-like domain initiates complex formation with an accessible high-affinity site to allow an orderly stepwise assembly of Lin28 on its target pre-miRNA. After the initial binding, the CSD melts the RNA to allow Lin28 to efficiently multimerize on the RNA. In the resulting 1:3 complex, all three Lin28-binding sites in the terminal loop are occupied by one molecule of Lin28, with the two G-rich sites near the Dicer cleavage sites forming high-affinity interactions. Notably, binding at these G-rich sites in the 1:3 complex involves both RNA domains of Lin28, with the CSD and the NCp7-like domain arranged in a 5' to 3' orientation on the RNA, in agreement with biochemical and X-ray structural studies (51) as well as transcriptome-wide studies of LIN28B-bound mRNA targets (57). Given that precursors of the let-7 family have large terminal loops in which these two important G-rich sequences are highly conserved (20,71), formation of multimeric complexes observed here with pre-let-7a-1 (55), pre-let-7d and pre-let-7g likely extends to other family members and may have important regulatory functions.

Role of the stepwise assembly of Lin28 for in vivo regulation of pre-let-7 biogenesis

Lin28 is highly expressed in stem cells, progenitor cells as well as poorly differentiated tumors, and Lin28 levels decrease during differentiation (72). Assembly of multiple molecules of Lin28 on the terminal loop of pre-let-7 could therefore play important regulatory functions in cells where Lin28 is highly abundant. Here, we tested the possibility that multimerization of Lin28 on pre-let-7 RNAs, which is observed at high Lin28 concentrations, could modulate the activity of Dicer. We determined that binding of only one molecule of Lin28 on the terminal loop of pre-let-7g and pre-let-7d is sufficient to elicit maximum inhibition of Dicer cleavage, but that binding of multiple molecules of Lin28 on the terminal loop of pre-let-7a-1 is required to reach maximum inhibition of Dicer cleavage. Interestingly, recent studies in HEK293 cells indicate the let-7a (7a-1, 7a-2 and 7a-3) is not strongly regulated by Lin28 compared to other members of the let-7 family, including let-7d and let-7g (54,55). In these studies, Lin28 may have reached a level that strongly inhibited Dicer processing for several let-7 precursors, but that only weakly inhibited Dicer processing for let-7a precursors. In contrast, efficient Lin28-dependent inhibition of pre-let-7a-1 maturation was observed in other studies (16,17,27-29), including studies with mammalian embryonic stem cells, where higher levels of Lin28 may allow its multimerization on pre-let-7a-1 and more efficient Dicer inhibition.

In vivo, the specificity of the Lin28-dependent inhibition of let-7 biogenesis likely stems from several aspects of Lin28 function depending on the cell type, including its ability to specifically form a highly stable 1:1 complex with let-7 precursors, its ability to multimerize on these precursors, and the coordination of its RNA-binding process with additional factors that control let-7 biogenesis. For example, the processive activity of the terminal uridylyl transferase TUT4 on pre-let-7 RNAs depends on the formation of a specific Lin28/pre-let-7 complex (17,28,31). By comparison, although Dicer likely competes with Lin28 for pre-let-7g binding, TUT4 must recognize a Lin28/pre-let-7 complex (17,28,31), and it is not yet clear if it can associate with the 1:1, 1:2 and/or 1:3 complex. Additional factors, including the Drosha-DGCR8 complex, hnRNPA1 (20,21), KSRP (21-23) and the Dis3L2 3'-5' exonuclease (34) are present in vivo to assist Lin28 in modulating the production of mature let-7 as needed for development, cell differentiation and tumor suppression (5-13). Such

factors may intervene at different steps of the assembly of Lin28 on pre-let-7 (or pri-let-7) terminal loops, either by competing with Lin28 for its binding sites or by associating with specific complexes along the assembly pathway. Multimerization of Lin28 on pre-let-7 terminal loops may inhibit the activity of factors that promote let-7 biogenesis and/or assist factors that prevent let-7 biogenesis, and thereby contribute to maintaining a pluripotent cell state. Similarly, the previously observed multimerization of Lin28 on target mRNAs may antagonize miRNA-mediated repression by displacing the miRNA-induced silencing complexes (36,69), although such multimerization may also contribute to recruit factors that control the fate of mRNAs. Thus, the present study provides a framework for future investigations aimed at better understanding the various functions of Lin28, particularly those that examine the interplay with additional factors involved in let-7 biogenesis.

Acknowledgements

We thank Geneviève Di Tomasso for sample preparation and J.G. Omichinski for careful reading of the manuscript.

Funding

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

Tables

Table 3-I. Dissociation constants (K_d)^a for binding of Lin28 to RNAs derived from TL-let-7g.^a

RNA	Sequences (5' to 3') ^b	K_d (nM)
TL-let-7g	GCAGAUUGAGGGUCUAUGAUACCACCCGGUACAGGAGAU AUCUGCA	0.13 ± 0.02 n = 2.9 ± 0.7 ^c
TL-let-7g Δbulge	GCAGAU ----- GUCUAUGAUACCACCCGGUACAGGAGAU AUCUGCA	0.9 ± 0.2 n = 2 ± 1
TL-let-7g Δiloop	GCAGAUUGAGGGUCU - UG - UACCACCCGGUACA -- AGAU AUCUGCA	0.52 ± 0.09 n = 1.3 ± 0.1
TL-let-7g GNRA loop	GCAGAUUGAGGGUCUAUGAUACC <u>CA</u> <u>AGGU</u> ACAGGAGAU AUCUGCA	0.4 ± 0.2 n = 1.46 ± 0.01
TL5-18	AUUGAGGGUCUAUA	1.7 ± 0.4
TL5-18V	AUUGACGCUCUAUA	>5000
TL13-26	UCUAUGAUACCACC	25-75 ^e
TL13-26V	UCUAU <u>CA</u> AACCACC	n.b. ^d
TL19-32	AUACCACCCGGUAC	n.b. ^d
TL28-41	GGUACAGGAGAU AU	0.29 ± 0.08
TL28-41V	GGUACAC <u>CA</u> CAU AU	25-75 ^e
TL33-46	AGGAGAU AUCUGCA	0.5 ± 0.1

^aEach K_d value and its associated error are the average and standard deviation, respectively, from at least three independent experiments.

^bResidues in bold are part of previously-identified Lin28 binding sites (see text).

^cThe K_d value was obtained from a previous study using the Hill equation (19).

^dNo specific binding observed. The gel mobility shift assays display smearing, indicating non-specific binding.

^eOnly an approximate value could be obtained in these cases.

Table 3-II. Ability of Lin28 and variants to form multimeric complexes with various RNAs.^a

RNA	Lin28			Lin28 C139A/C161A			Lin28 ₁₁₉₋₁₈₀		
	1:1	1:2	1:3	1:1	1:2	1:3	1:1	1:2	1:3
TL-let-7g	++++	++++	++++	++	++	-	+++	+	-
TL-let-7g Δ loop	+++	++	++	+	-	-	++	-	-
TL-let-7g Δ bulge	++	++	-	-	+	-	-	-	-
TL-let-7a-1	++++	+++	++	+	-	-	+++	-	-
TL-let-7d	++++	++	+/>+++	+++	-	-	+++	+	-
TL-miR-99b	+++	-/>+	-	+	-	-	-	-	-

^aThe score indicated the amount of RNA present in the indicated stoichiometric ratio: +++++, >75%; +++, 50-75%; ++, 25-50%; +, 5-25%; -, 0-5%. The ability of the protein to form a 1:1, 1:2 or 1:3 complex with the RNA was respectively quantified at 0.750, 1.250 and 1.750 μ M of protein from gels shown in Figure 3-2 and Figure 3-3.

Figures

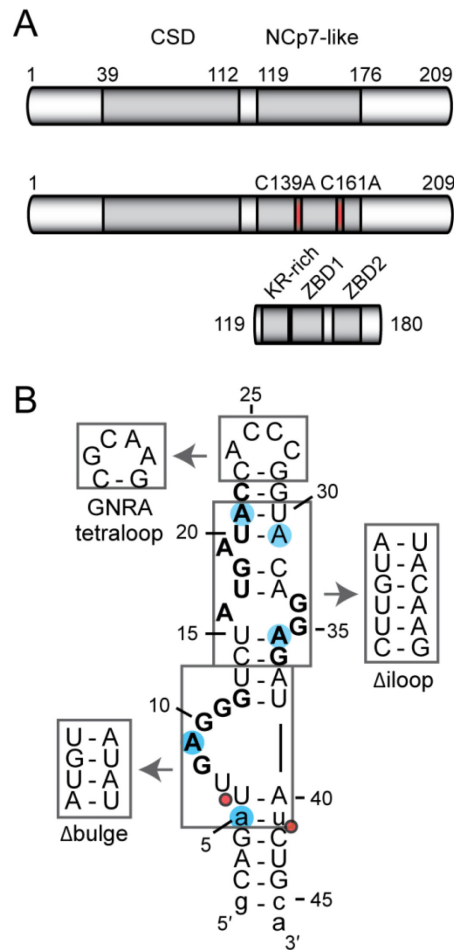


Figure 3-1. The Lin28 protein, TL-let-7g RNA and related sequences used in this study. **(A)** Schematic representation of the primary structures of Lin28 and related variants, Lin28 C139A/C161A and Lin28₁₁₉₋₁₈₀. The grey boxes delineate sequences of known RNA-binding motifs: a cold shock domain (CSD) and an NCp7-like domain with a KR-rich region (residues 125-135) N-terminal to a pair of retroviral-type CCHC zinc-binding domains [ZBD1 (residues 137-154) and ZBD2 (residues 160-176); (19)]. Site-specific substitutions of Lin28 are shown in red. **(B)** Primary and proposed secondary structures of TL-let-7g. Non-natural nucleotides are shown in lowercase, residues of previously-identified Lin28-binding sites in bold characters, substitution sites for 2-AP are as blue shadows, regions that were replaced by alternative structural elements are boxed, and Dicer cleavage sites are indicated by red dots.

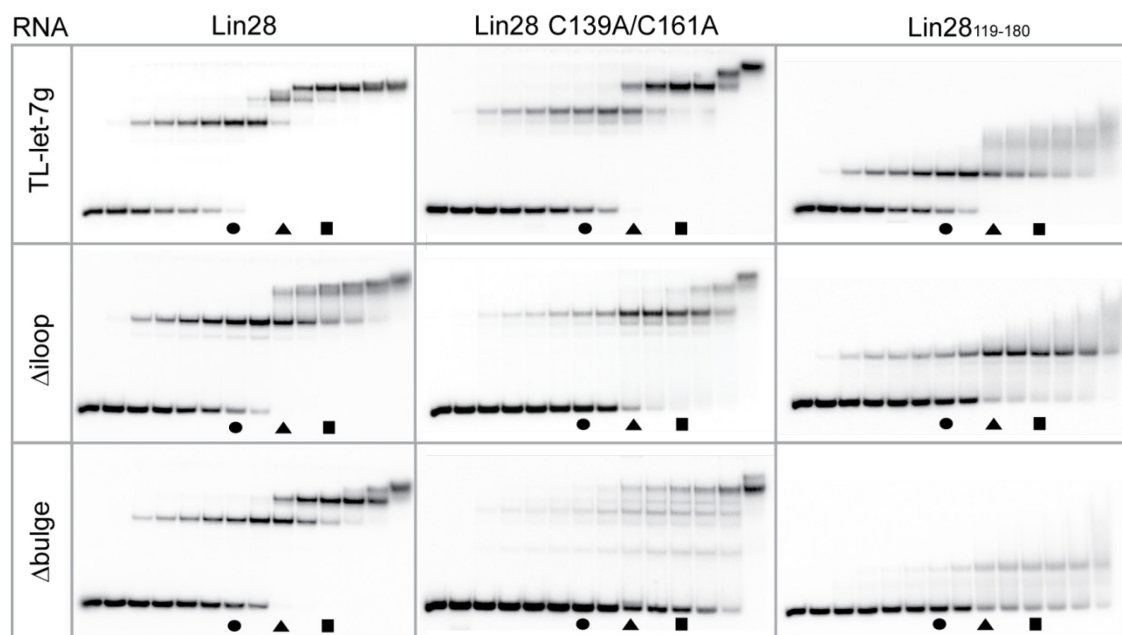


Figure 3-2. Stoichiometric binding assay by native gel electrophoresis for Lin28 binding to TL-let-7g RNA and effect of select RNA and protein variants on the stoichiometry of the complex. Each assay is performed with 0.5 μM RNA, including 10 pM 5'-[^{32}P]-labeled RNA, and increasing concentrations of protein (0.00, 0.05, 0.25, 0.375, 0.50, 0.625, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.50 and 5.00 μM). The gel lanes with RNA:protein ratios of 1:1.5, 1:2.5 and 1:3.5 are identified by a circle, a triangle and a square, respectively.

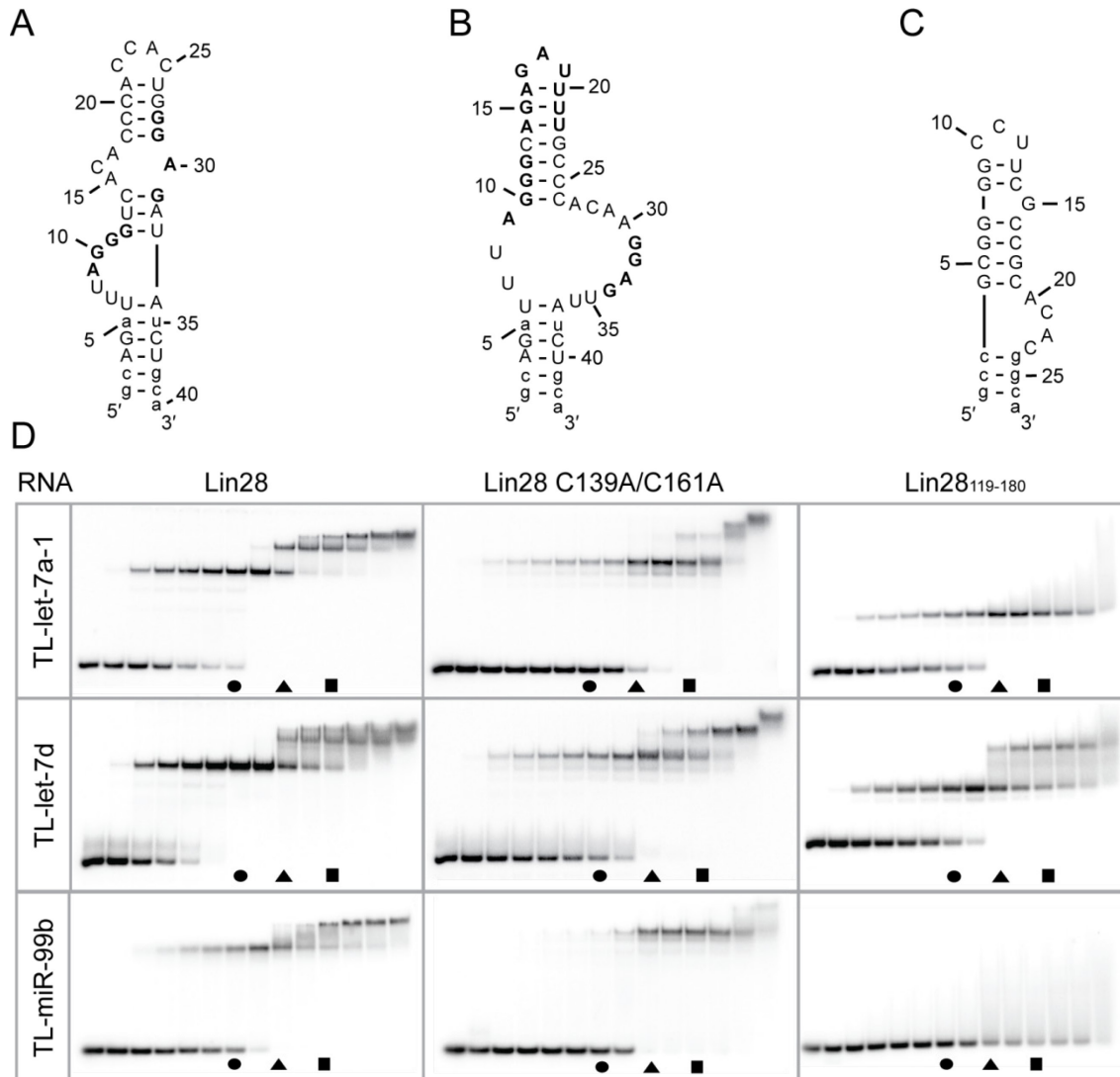


Figure 3-3. Stoichiometric binding assay by native gel electrophoresis for Lin28 binding to other pre-let-7 terminal loops. Sequences and proposed secondary structures of (A) TL-let-7a-1, (B) TL-let-7d RNA and (C) TL-miR-99b with bold residues representing previously-identified or predicted Lin28-binding sites (see text). Non-natural nucleotides are shown in lowercase. (D) Stoichiometric binding assay of TL-let-7a-1, TL-let-7d and TL-miR-99b RNA with Lin28, Lin28 C139A/C161A and Lin28₁₁₉₋₁₈₀. Each assay is performed with 0.5 μ M RNA, including 10 pM 5'-[³²P]-labeled RNA, and increasing concentrations of protein (0.00, 0.05, 0.25, 0.375, 0.50, 0.625, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.50 and 5.00 μ M). The gel lanes with RNA:protein ratios of 1:1.5, 1:2.5 and 1:3.5 are identified by a circle, a triangle and a square, respectively.

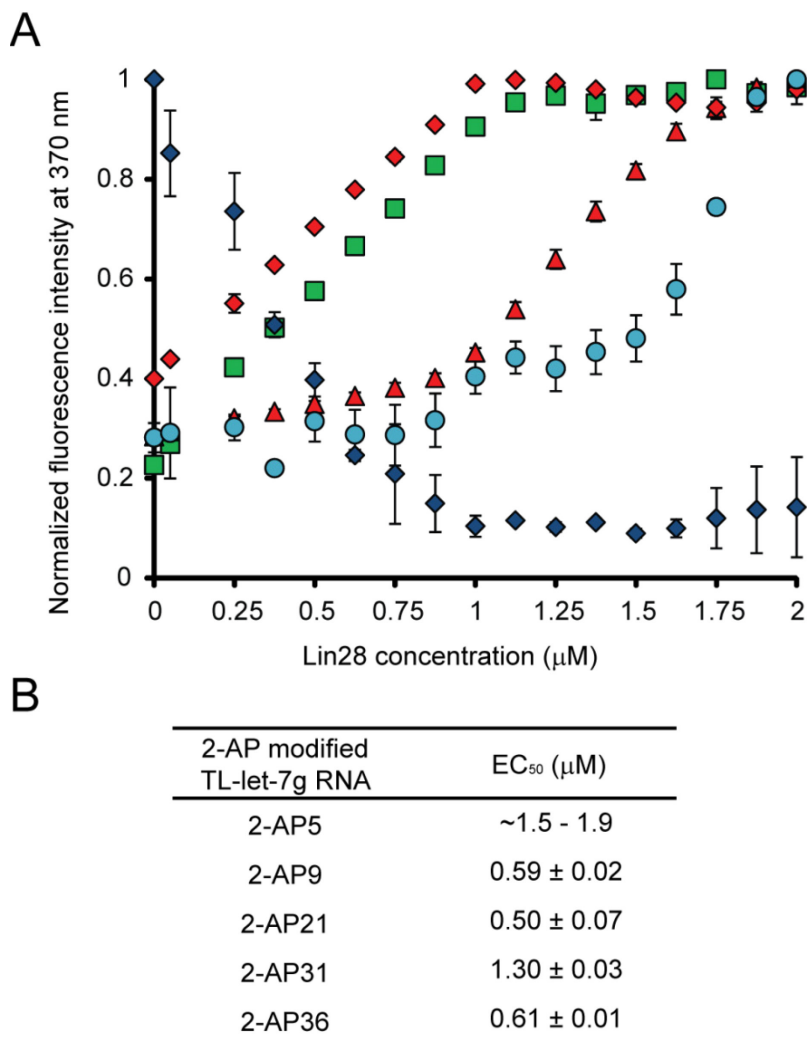


Figure 3-4. Effect of Lin28 on the fluorescence intensity of TL-let-7g containing individual 2-AP modifications. **(A)** Normalized fluorescence intensity at 370 nm of 2-AP5 (cyan dots), 2-AP9 (green squares), 2-AP21 (blue diamonds), 2-AP31 (red triangles) and 2-AP36 (red diamonds) as a function of Lin28 concentration. Each titration point and the associated error bars are respectively the average and standard deviation from multiple experiments. **(B)** Reported EC₅₀ values and their errors obtained from data in **(A)** by fitting the normalized fluorescence intensity at 370 nm with respect to Lin28 concentration using the dose-response equation.

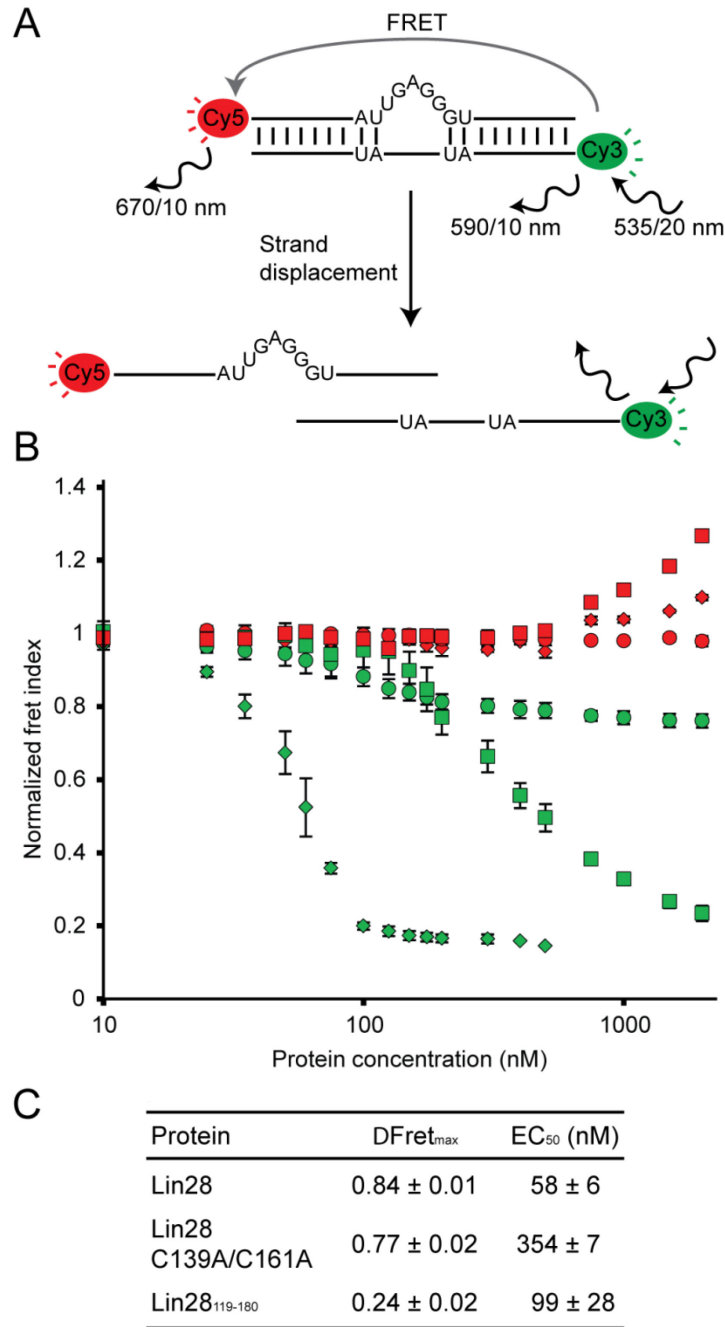


Figure 3-5. Strand displacement assay monitored by FRET to test the RNA melting activity of Lin28. **A)** Schematic of the RNA melting activity of Lin28 on the duplex_{bulge} RNA, a Cy3/Cy5-labeled duplex. Relevant excitation and emission frequencies are indicated with associated slit widths. **(B)** Normalized Fret Index ($F_{\text{Cy5}}/F_{\text{Cy3}}$) of Cy3/Cy5-labeled RNA duplexes (25 nM) as a function of total protein concentration (10, 25, 35, 50, 60, 75, 100, 125, 150, 175, 200, 300, 400, 500, 750, 1000, 1500 and 2000 nM). Proteins were sequentially

added to both the duplex_{bulge} (Lin28, green diamonds; Lin28 C139A/C161A, green squares; and Lin28₁₁₉₋₁₈₀, green circles) and the control duplex_{comp} (Lin28, red diamonds; Lin28 C139A/C161A, red squares; and Lin28₁₁₉₋₁₈₀, red circles). (C) Values of $\Delta\text{FRET}_{\text{max}}$ and EC_{50} derived from results shown in B.

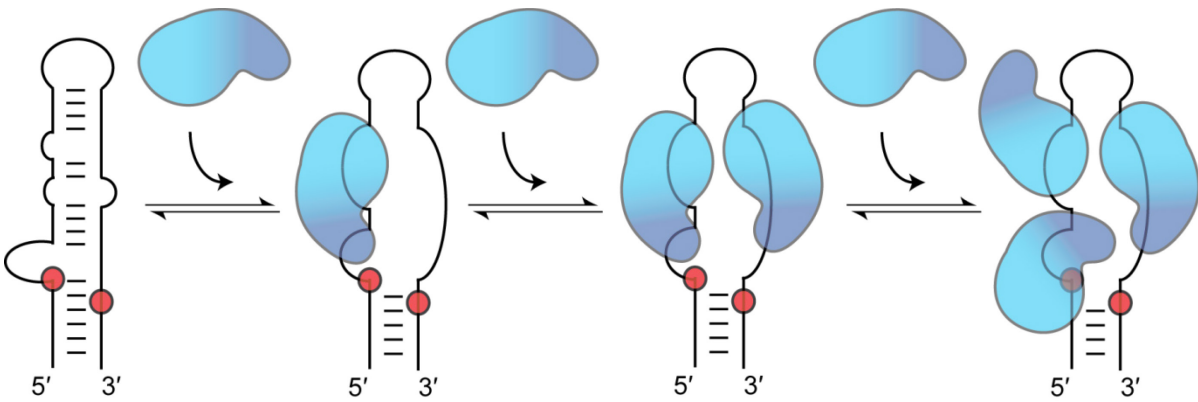


Figure 3-7. Schematic representation of the proposed model for the stepwise assembly of Lin28 on the terminal loop of pre-let-7g. In this model, only the RNA-binding domains of Lin28 are represented, with the NCp7-like domain in dark blue and the CSD in cyan. The first molecule of Lin28 initiates complex formation through interactions between its NCp7-like domain and the conserved 5'-GAGGG-3' site at the 5'-bulge and between its CSD and the nearby 5'-AUGAUAC-3' site. These interactions on the 5'-strand destabilize Watson-Crick base pairs within the terminal loop and exposes the 3'-strand, making it available for binding a second molecule of Lin28, with its NCp7-like domain targeting the conserved 5'-GGAG-3' sequence and its CSD an adjacent 5'-region. Binding of the third molecule of Lin28 to the high-affinity 5'-AUGAUAC-3' site involves only its CSD and implicates relocating the CSD of the first molecule of Lin28. The order of assembly proposed for pre-let-7g may differ for precursors of other let-7 family members.

References

1. Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215-233.
2. Kim, V.N., Han, J. and Siomi, M.C. (2009) Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol*, **10**, 126-139.
3. Newman, M.A. and Hammond, S.M. (2010) Emerging paradigms of regulated microRNA processing. *Genes Dev*, **24**, 1086-1092.
4. Winter, J. and Diederichs, S. (2011) MicroRNA biogenesis and cancer. *Methods Mol Biol*, **676**, 3-22.
5. Ambros, V. (2011) MicroRNAs and developmental timing. *Curr Opin Genet Dev*, **21**, 511-517.
6. Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R. and Ruvkun, G. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, **403**, 901-906.
7. Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P. *et al.* (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, **408**, 86-89.
8. Melton, C., Judson, R.L. and Blelloch, R. (2010) Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature*, **463**, 621-626.
9. Osada, H. and Takahashi, T. (2011) let-7 and miR-17-92: small-sized major players in lung cancer development. *Cancer Sci*, **102**, 9-17.
10. Chin, L.J., Ratner, E., Leng, S., Zhai, R., Nallur, S., Babar, I., Muller, R.U., Straka, E., Su, L., Burki, E.A. *et al.* (2008) A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk. *Cancer Res*, **68**, 8535-8540.
11. He, X.Y., Chen, J.X., Zhang, Z., Li, C.L., Peng, Q.L. and Peng, H.M. (2010) The let-7a microRNA protects from growth of lung carcinoma by suppression of k-Ras and c-Myc in nude mice. *J Cancer Res Clin Oncol*, **136**, 1023-1028.

12. Bussing, I., Slack, F.J. and Grosshans, H. (2008) let-7 microRNAs in development, stem cells and cancer. *Trends Mol Med*, **14**, 400-409.
13. Roush, S. and Slack, F.J. (2008) The let-7 family of microRNAs. *Trends Cell Biol*, **18**, 505-516.
14. Pasquinelli, A.E., McCoy, A., Jimenez, E., Salo, E., Ruvkun, G., Martindale, M.Q. and Baguna, J. (2003) Expression of the 22 nucleotide let-7 heterochronic RNA throughout the Metazoa: a role in life history evolution? *Evol Dev*, **5**, 372-378.
15. 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.
16. 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.
17. 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.
18. 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.
19. Desjardins, A., Yang, A., Bouvette, J., Omichinski, J.G. and Legault, P. (2012) Importance of the NCp7-like domain in the recognition of pre-let-7g by the pluripotency factor Lin28. *Nucleic Acids Res*, **40**, 1767-1777.
20. Michlewski, G., Guil, S., Semple, C.A. and Caceres, J.F. (2008) Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol Cell*, **32**, 383-393.
21. 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.
22. 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.

23. Nicastro, G., Garcia-Mayoral, M.F., Hollingworth, D., Kelly, G., Martin, S.R., Briata, P., Gherzi, R. and Ramos, A. (2012) Noncanonical G recognition mediates KSRP regulation of let-7 biogenesis. *Nat Struct Mol Biol*, **19**, 1282-1286.
24. Viswanathan, S.R., Daley, G.Q. and Gregory, R.I. (2008) Selective blockade of microRNA processing by Lin28. *Science*, **320**, 97-100.
25. 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.
26. 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.
27. Piskounova, E., Polytarchou, C., Thornton, J.E., LaPierre, R.J., Pothoulakis, C., Hagan, J.P., Iliopoulos, D. and Gregory, R.I. (2011) Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. *Cell*, **147**, 1066-1079.
28. 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.
29. 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.
30. Heo, I., Ha, M., Lim, J., Yoon, M.J., Park, J.E., Kwon, S.C., Chang, H. and Kim, V.N. (2012) Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. *Cell*, **151**, 521-532.
31. 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.
32. 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.

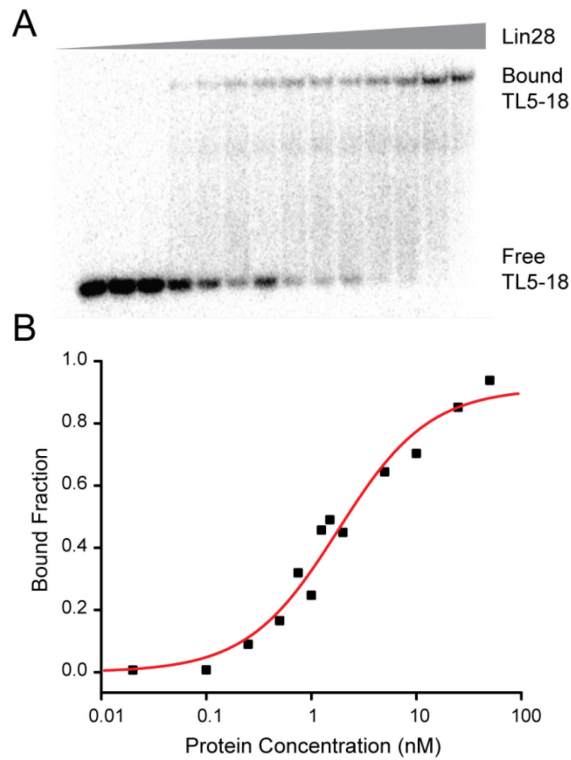
33. Thornton, J.E., Chang, H.M., Piskounova, E. and Gregory, R.I. (2012) Lin28-mediated control of let-7 microRNA expression by alternative TUTases Zcchc11 (TUT4) and Zcchc6 (TUT7). *RNA*, **18**, 1875-1885.
34. Chang, H.M., Triboulet, R., Thornton, J.E. and Gregory, R.I. (2013) A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. *Nature*, **497**, 244-248.
35. Ustianenko, D., Hrossova, D., Potesil, D., Chalupnikova, K., Hrazdilova, K., Pachernik, J., Cetkowska, K., Uldrijan, S., Zdrahal, Z. and Vanacova, S. (2013) Mammalian DIS3L2 exoribonuclease targets the uridylated precursors of let-7 miRNAs. *RNA*, **19**, 1632-1638.
36. Kallen, A.N., Ma, J. and Huang, Y. (2012) Does Lin28 Antagonize miRNA-Mediated Repression by Displacing miRISC from Target mRNAs? *Front Genet*, **3**, 240.
37. 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.
38. Lettre, 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.* (2008) Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet*, **40**, 584-591.
39. Hartge, P. (2009) Genetics of reproductive lifespan. *Nat Genet*, **41**, 637-638.
40. 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.
41. 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.
42. Faas, L., Warrander, F.C., Maguire, R., Ramsbottom, S.A., Quinn, D., Genever, P. and Isaacs, H.V. (2013) Lin28 proteins are required for germ layer specification in *Xenopus*. *Development*, **140**, 976-986.

43. Shyh-Chang, N., Zhu, H., Yvanka de Soysa, T., Shinoda, G., Seligson, M.T., Tsanov, K.M., Nguyen, L., Asara, J.M., Cantley, L.C. and Daley, G.Q. (2013) Lin28 enhances tissue repair by reprogramming cellular metabolism. *Cell*, **155**, 778-792.
44. 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.
45. Chen, A.X., Yu, K.D., Fan, L., Li, J.Y., Yang, C., Huang, A.J. and Shao, Z.M. (2011) Germline genetic variants disturbing the Let-7/LIN28 double-negative feedback loop alter breast cancer susceptibility. *PLoS Genet*, **7**, e1002259.
46. 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.
47. Thornton, J.E. and Gregory, R.I. (2012) How does Lin28 let-7 control development and disease? *Trends Cell Biol*, **22**, 474-482.
48. 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.
49. King, C.E., Cuatrecasas, M., Castells, A., Sepulveda, A.R., Lee, J.S. and Rustgi, A.K. (2011) LIN28B promotes colon cancer progression and metastasis. *Cancer Res*, **71**, 4260-4268.
50. Moss, E.G. and Tang, L. (2003) Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. *Dev Biol*, **258**, 432-442.
51. Nam, Y., Chen, C., Gregory, R.I., Chou, J.J. and Sliz, P. (2011) Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell*, **147**, 1080-1091.
52. Mayr, F., Schutz, A., Doge, N. and Heinemann, U. (2012) The Lin28 cold-shock domain remodels pre-let-7 microRNA. *Nucleic Acids Res*, **40**, 7492-7506.
53. Loughlin, F.E., Gebert, L.F., Towbin, H., Brunschweiler, A., Hall, J. and Allain, F.H. (2012) Structural basis of pre-let-7 miRNA recognition by the zinc knuckles of pluripotency factor Lin28. *Nat Struct Mol Biol*, **19**, 84-89.

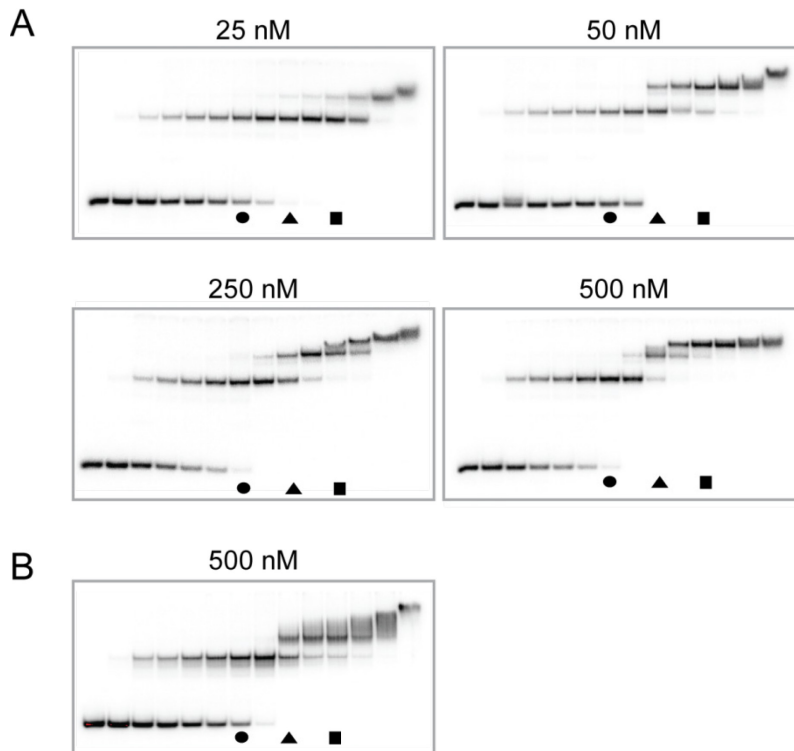
54. Wilbert, M.L., Huelga, S.C., Kapeli, K., Stark, T.J., Liang, T.Y., Chen, S.X., Yan, B.Y., Nathanson, J.L., Hutt, K.R., Lovci, M.T. *et al.* (2012) LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. *Mol Cell*, **48**, 195-206.
55. Hafner, M., Max, K.E., Bandaru, P., Morozov, P., Gerstberger, S., Brown, M., Molina, H. and Tuschl, T. (2013) Identification of mRNAs bound and regulated by human LIN28 proteins and molecular requirements for RNA recognition. *RNA*, **19**, 613-626.
56. Cho, J., Chang, H., Kwon, S.C., Kim, B., Kim, Y., Choe, J., Ha, M., Kim, Y.K. and Kim, V.N. (2012) LIN28A is a suppressor of ER-associated translation in embryonic stem cells. *Cell*, **151**, 765-777.
57. Graf, R., Munschauer, M., Mastrobuoni, G., Mayr, F., Heinemann, U., Kempa, S., Rajewsky, N. and Landthaler, M. (2013) Identification of LIN28B-bound mRNAs reveals features of target recognition and regulation. *RNA Biol*, **10**.
58. Salvail-Lacoste, A., Di Tomasso, G., Piette, B.L. and Legault, P. (2013) Affinity purification of T7 RNA transcripts with homogeneous ends using ARiBo and CRISPR tags. *RNA*, **19**, 1003-1014.
59. Walker, S.C., Avis, J.M. and Conn, G.L. (2003) General plasmids for producing RNA in vitro transcripts with homogeneous ends. *Nucleic Acids Research*, **31**, e82.
60. Bouchard, P., Lacroix-Labonte, 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.
61. Ali, P.S., Ghoshdastider, U., Hoffmann, J., Brutschy, B. and Filipek, S. (2012) Recognition of the let-7g miRNA precursor by human Lin28B. *FEBS Lett*, **586**, 3986-3990.
62. Ward, D.C., Reich, E. and Stryer, L. (1969) Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives. *J Biol Chem*, **244**, 1228-1237.
63. Menger, M., Tuschl, T., Eckstein, F. and Porschke, D. (1996) Mg(2+)-dependent conformational changes in the hammerhead ribozyme. *Biochemistry*, **35**, 14710-14716.

64. Rist, M.J. and Marino, J.P. (2002) Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry*, **41**, 14762-14770.
65. Jean, J.M. and Hall, K.B. (2001) 2-Aminopurine fluorescence quenching and lifetimes: role of base stacking. *Proc Natl Acad Sci U S A*, **98**, 37-41.
66. Heppell, B., Mulhbachter, J., Penedo, J.C. and Lafontaine, D.A. (2009) Application of fluorescent measurements for characterization of riboswitch-ligand interactions. *Methods Mol Biol*, **540**, 25-37.
67. Rajkowitsch, L. and Schroeder, R. (2007) Coupling RNA annealing and strand displacement: a FRET-based microplate reader assay for RNA chaperone activity. *Biotechniques*, **43**, 304-310.
68. Park, J.E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J. and Kim, V.N. (2011) Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature*, **475**, 201-205.
69. Lei, X.X., Xu, J., Ma, W., Qiao, C., Newman, M.A., Hammond, S.M. and Huang, Y. (2012) Determinants of mRNA recognition and translation regulation by Lin28. *Nucleic Acids Res*, **40**, 3574-3584.
70. 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.
71. 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.
72. Viswanathan, S.R. and Daley, G.Q. (2010) Lin28: A microRNA regulator with a macro role. *Cell*, **140**, 445-449.

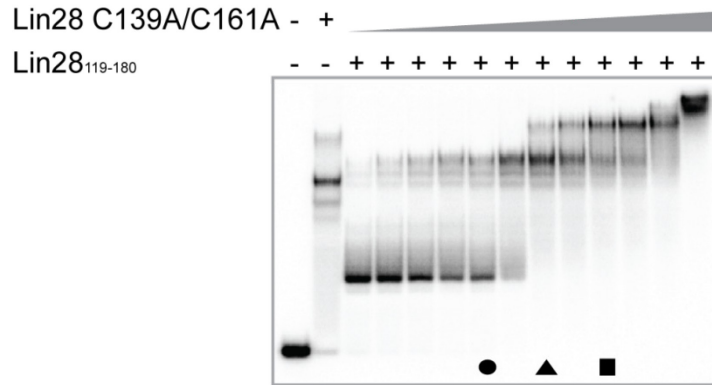
Supplementary material



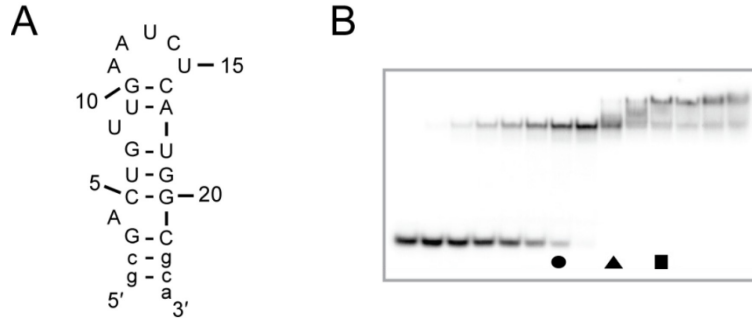
Supplementary Figure 3-1. Binding studies of TL-let-7g to Lin28 by EMSA for determination of the dissociation constant. **(A)** Typical EMSA performed with 1 pM 5'-[³²P]-labeled TL5-18 RNA and increasing concentrations of Lin28 (0.0, 0.02, 0.10, 0.25, 0.50, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 5.0, 10, 25 and 50 nM). **(B)** Fraction of bound RNA as a function of total protein concentration, fitted to the one site binding equation (red line; with $K_d = 1.8$ nM and $R^2 = 0.97$).



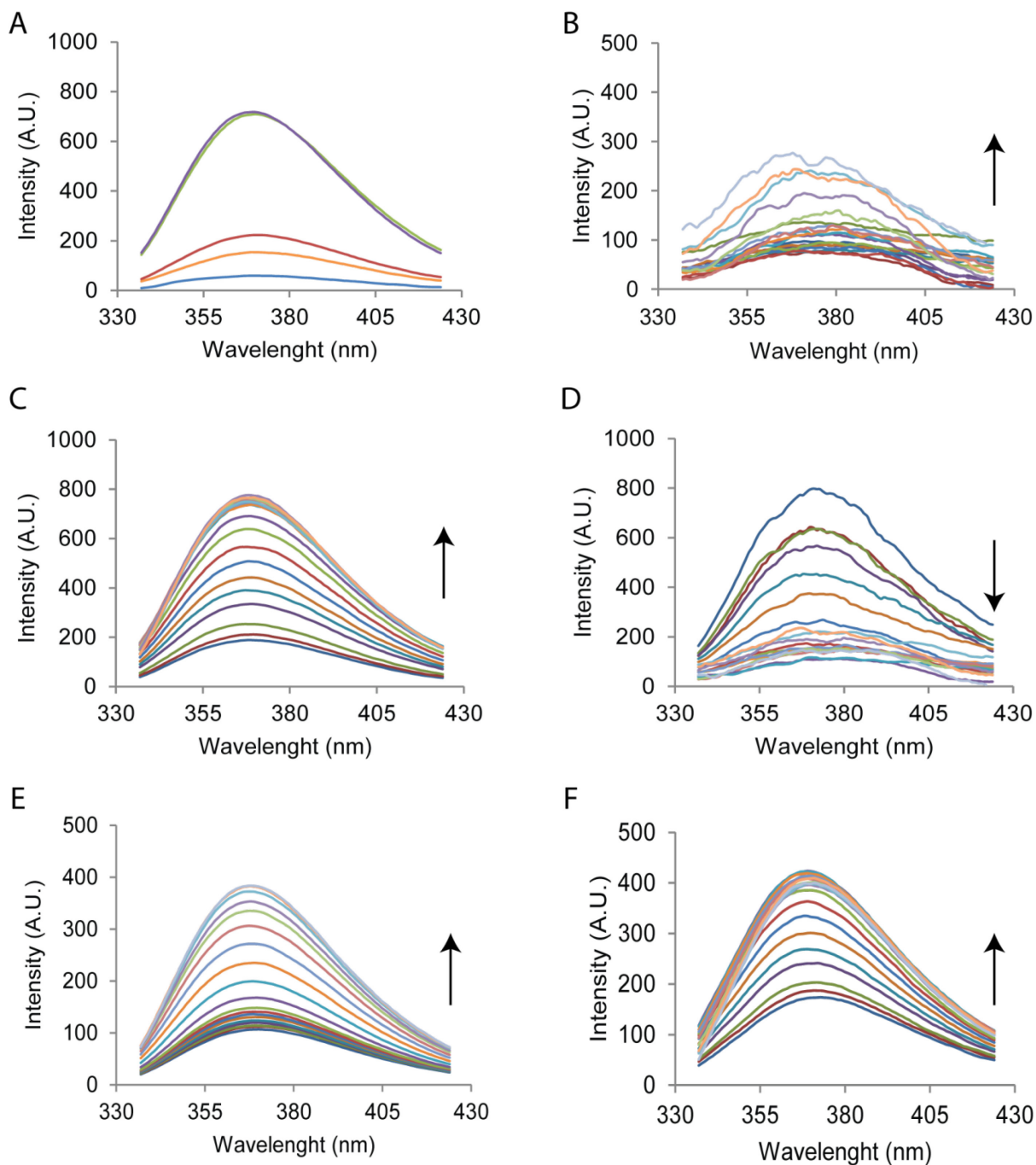
Supplementary Figure 3-2. Stoichiometric binding assay by native gel electrophoresis for Lin28 binding to different concentrations of TL-let-7g **(A)** and to 500 nM of pre-let-7g [for the exact sequence see (1)] **(B)**. Each assay is performed with the indicated RNA concentration, including 10 pM 5'-[³²P]-labeled RNA, and increasing concentrations of protein (from 0.1X to 10X the RNA concentration). The gel lanes with RNA:protein ratios of 1:1.5, 1:2.5 and 1:3.5 are identified by a circle, a triangle and a square, respectively.



Supplementary Figure 3-3. Competitive stoichiometric binding assay by native gel electrophoresis between Lin28 C139A/C161A and Lin28₁₁₉₋₁₈₀ for binding to TL-let-7g. The assay is performed with 0.5 μ M RNA, including 10 pM 5'-[³²P]-labeled RNA, increasing concentrations of Lin28 C139A/C161A (0.00, 1.00, 0.05, 0.25, 0.375, 0.50, 0.625, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.50 and 5.00 μ M) and constant concentration of Lin28₁₁₉₋₁₈₀ (1 μ M). The gel lanes with RNA:Lin28 C139A/C161A ratios of 1:1.5, 1:2.5 and 1:3.5 are identified by a circle, a triangle and a square, respectively.

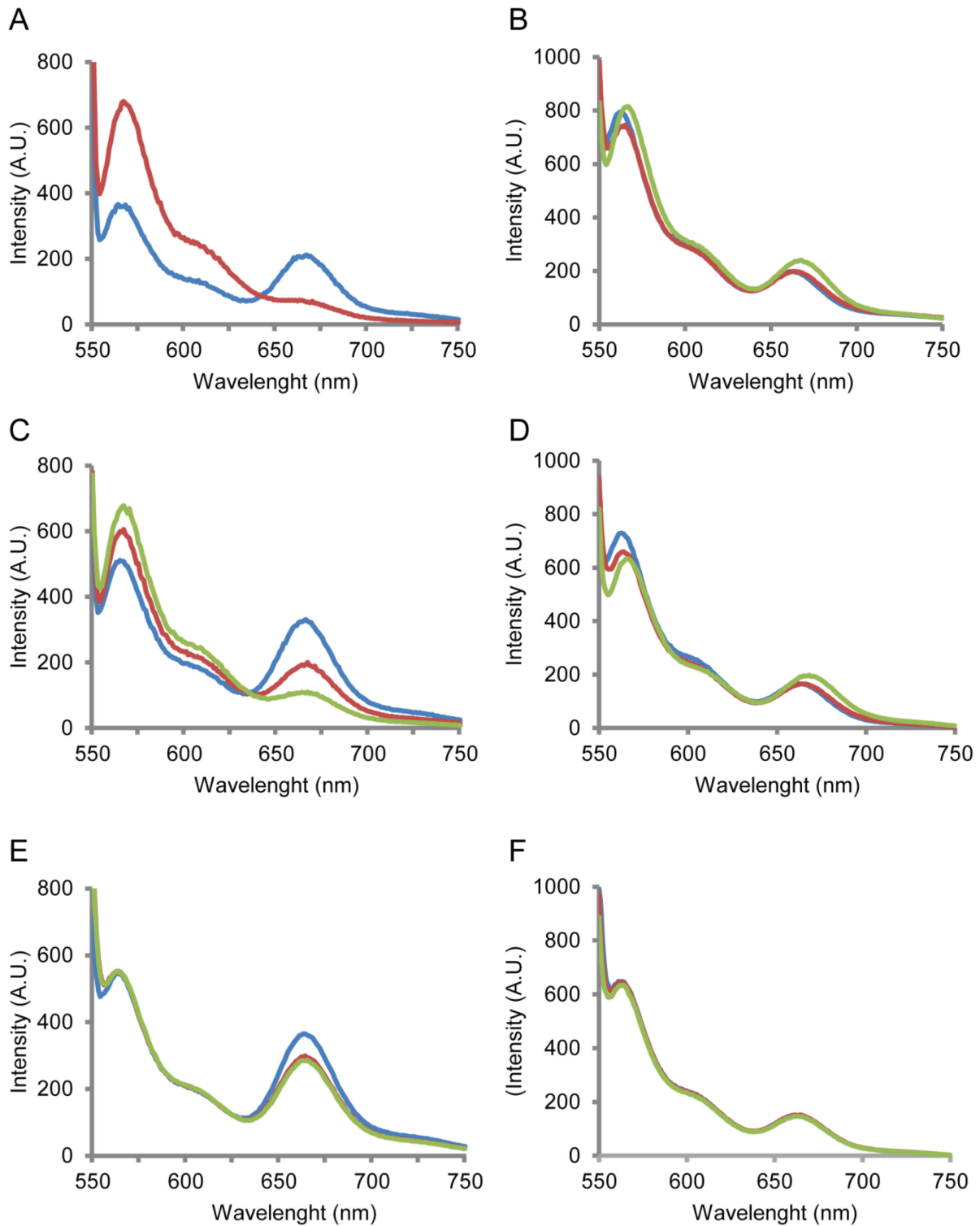


Supplementary Figure 3-4. Stoichiometric binding assay by native gel electrophoresis for Lin28 binding to TL-miR-21a. **(A)** Sequence and proposed secondary structure of TL-miR-21a. **(B)** Stoichiometric binding assay of TL-miR-21a with Lin28. The assay is performed with 0.5 μM RNA, including 10 pM 5'-[^{32}P]-labeled RNA, and increasing concentrations of protein (0.00, 0.05, 0.25, 0.375, 0.50, 0.625, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.50 and 5.00 μM). The gel lanes with RNA:protein ratios of 1:1.5, 1:2.5 and 1:3.5 are identified by a circle, a triangle and a square, respectively.



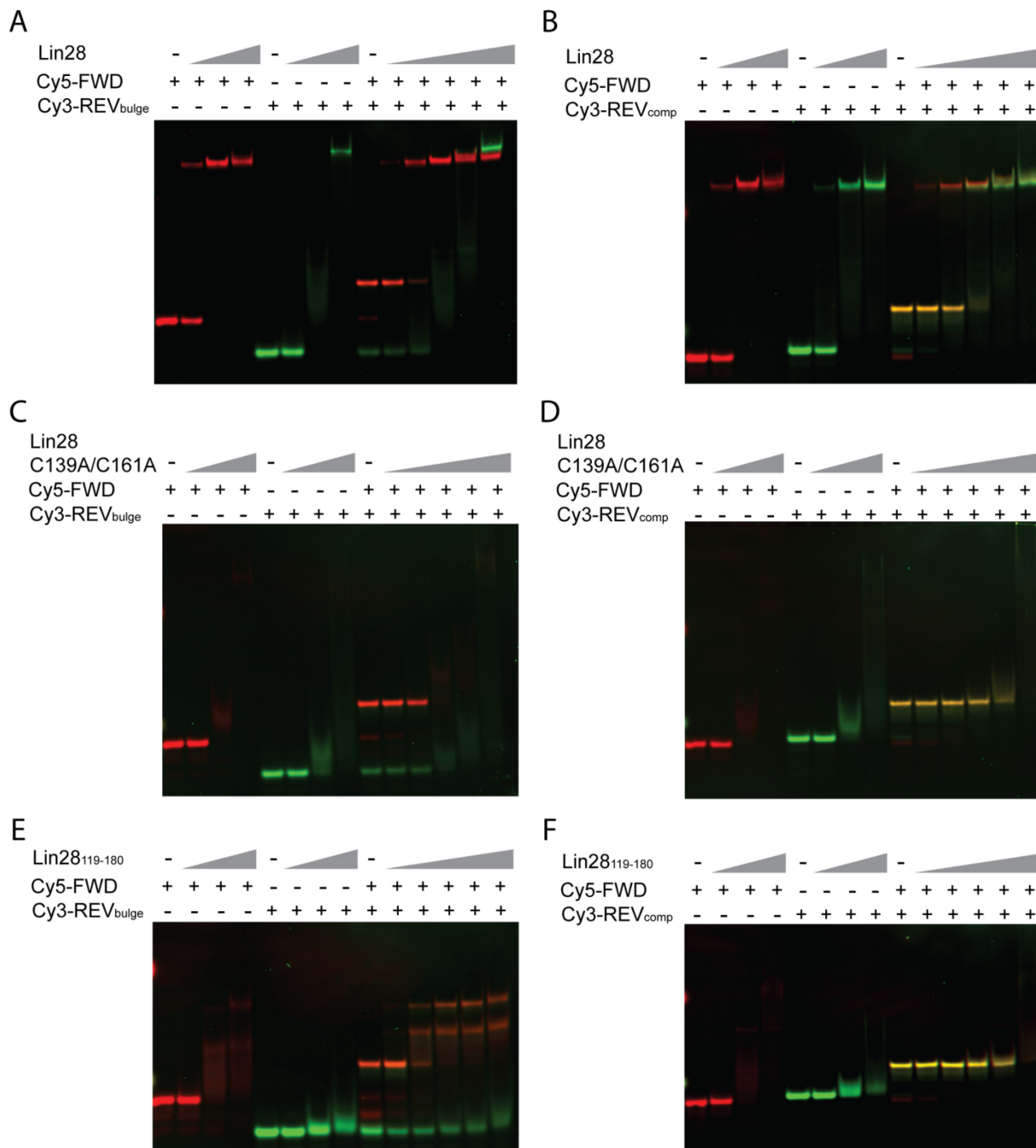
Supplementary Figure 3-5. Effect of addition of Lin28 on the emission spectra of 2-AP-modified TL-let-7g RNAs. **(A)** Emission spectra recorded in the absence of protein for TL-let-7g RNAs (500 nM) containing a single 2-AP at position A5 (blue), A9 (red), A21 (green), A31 (purple) or A36 (orange). Spectra were all recorded with the same detector gain settings for comparison purposes. **(B)-(F)** Difference emission spectra following the addition of Lin28

(0 μM , dark blue; 0.050 μM , dark red; 0.125 μM , dark green; 0.250 μM , dark purple; 0.375 μM , dark cyan; 0.500 μM , dark orange; 0.625 μM , blue; 0.750 μM , red; 0.875 μM , green; 1 μM , purple; 1.125 μM , cyan; 1.250 μM , orange; 1.375, light blue; 1.500 μM , light red; 1.625 μM , light green; 1.750 μM , light purple; 1.875 μM , light cyan; 2 μM , light orange; and 2.5 μM , grey) to different TL-let-7g RNAs containing a single 2-AP at positions **(B)** A5, **(C)** A9, **(D)** A21, **(E)** A31 or **(F)** A36. The arrows indicate the increase (arrow up) or decrease (arrow down) of fluorescence emission upon addition of protein. The detector gain setting was optimized for each titration experiment, thus the intensities detected for different RNA-protein combinations cannot be directly compared.



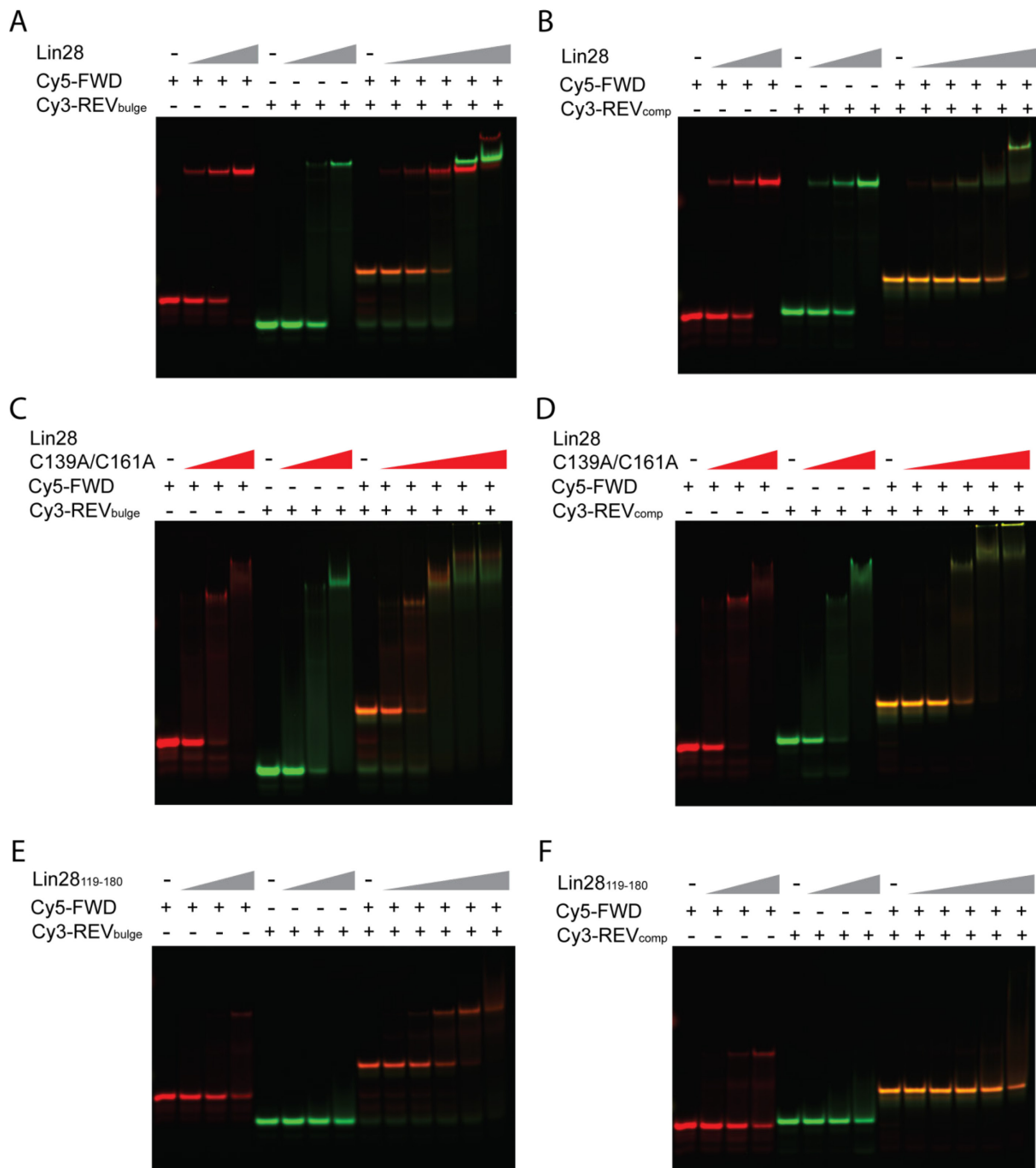
Supplementary Figure 3-6. Effect of addition of Lin28 on the emission spectra of Cy3/Cy5-labeled RNA duplexes. The RNA duplex containing a G-rich bulge (duplex_{bulge}) was titrated

with (A) Lin28, (C) Lin28 C139A/C161A and (E) Lin28₁₁₉₋₁₈₀. The control complementary RNA duplex (duplex_{comp}) was also titrated with (B) Lin28, (D) Lin28 C139A/C161A and (F) Lin28₁₁₉₋₁₈₀. The emission spectra were recorded from 550/10 nm to 750/10 nm after excitation at 535/20 nm on samples containing 25 nM Cy3/Cy5-labeled RNA duplexes and either 0 nM (blue), 500 nM (red) or 2 μ M protein (green). In (A) the 2- μ M titration point was not recorded, whereas all three titration points are present in the other panels.



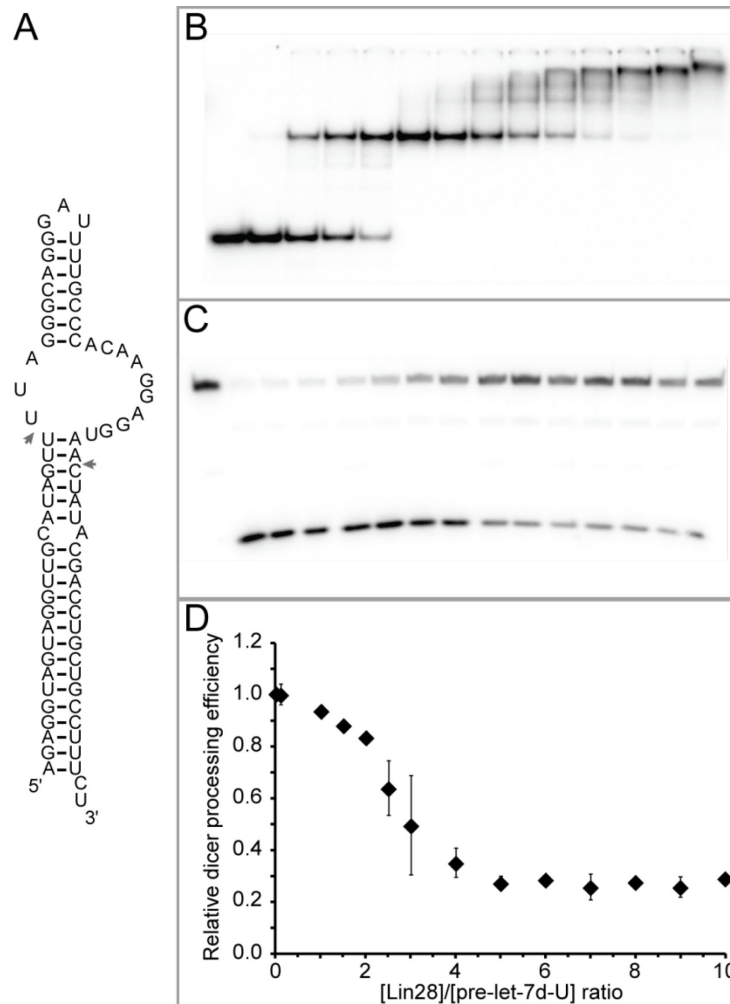
Supplementary Figure 3-7. Strand displacement assay monitored by native gel electrophoresis with 25 nM of Cy3/Cy5-labeled duplex RNAs. (A) Additions of Lin28 (0, 25, 250 and 2000 nM) to isolated RNA strands, either 25 nM of Cy5-FWD or 25 nM Cy3-REV_{bulge} followed by the additions of Lin28 (0, 25, 100, 250, 500 and 2000 nM) to 25 nM duplex_{bulge} formed by annealing Cy5-FWD to Cy3-REV_{bulge}. Titrations similar to those described in (A) were performed with (C) Lin28 C139A/C161A and (E) Lin28₁₁₉₋₁₈₀. (B)

Titration experiments were performed as in (A), except that Cy3-REV_{bulge} is replaced by Cy3-REV_{comp} and duplex_{bulge} is replaced by duplex_{comp}. Titration experiments similar to those described in (B) were performed with (D) Lin28 C139A/C161A and (F) Lin28₁₁₉₋₁₈₀. The Cy3 and Cy5 signals are represented by green and red colors, respectively, and the superposition of both signals produces a spectrum of colors from yellow to orange-red. In (B), binding of Lin28 to the duplex cannot be distinguished from binding to the individual strands. In several cases, binding of Lin28 creates such a diffuse band on the gel that the RNA seems to disappear from the gel.



Supplementary Figure 3-8. Strand displacement assay monitored by native gel electrophoresis with 250 nM of Cy3/Cy5-labeled RNA. (A) Additions of Lin28 (0, 125, 250 and 500 nM) to isolated RNA strands, either 250 nM of Cy5-FWD or 250 nM Cy3-REV_{bulge} followed by the additions of Lin28 (0, 125, 250, 500, 1000 and 2000 nM) to 250 nM Cy3/Cy5-labeled duplex_{bulge} formed by annealing Cy5-FWD to Cy3-REV_{bulge}. Titrations similar to those

described in (A) were performed with (C) Lin28 C139A/C161A and (E) Lin28₁₁₉₋₁₈₀, except that different protein concentrations (red gradients) were used in (C). (B) Titrations were performed as in (A), except that Cy3-REV_{bulge} is replaced by Cy3-REV_{comp} and duplex_{bulge} is replaced by duplex_{comp}. Titrations similar to those described in (B) were performed with (D) Lin28 C139A/C161A and (F) Lin28₁₁₉₋₁₈₀, except that different protein concentrations (red gradients) were used in (D). In (C) and (D) 0, 0.5, 1 and 2 μ M of Lin28 were sequentially added to the isolated RNA strands and 0, 0.5, 1, 2, 5 and 10 μ M of protein were sequentially added to the duplex RNA (red gradients). The Cy3 and Cy5 signals are represented by green and red colors, respectively, and the superposition of both signals produces a spectrum of colors from yellow to orange-red. In (B), binding of Lin28 to the duplex cannot be distinguished from binding to the individual strands. In several cases, binding of Lin28 creates such a diffuse band on the gel that the RNA seems to disappear from the gel.



Supplementary Figure 3-9. Dicer processing assay of pre-let-7d-U. **(A)** Primary and proposed secondary structures of the pre-let-7d-U RNA. Dicer cleavage sites are identified with arrows. **(B)** Stoichiometric binding assay detected by native gel electrophoresis for Lin28 binding to pre-let-7d-U. Each assay is performed with 0.5 μ M 5'-phosphorylated RNA, including 40 pM 5'-[32 P]-labeled RNA and increasing concentration of protein (0.00, 0.05, 0.50, 0.75, 1.25, 1.50, 2.00, 2.50, 3.00, 3.50, 4.00, 4.50 and 5.00 μ M). **(C)** Dicer processing assay detected by denaturing gel electrophoresis of pre-let-7d-U under varying Lin28 concentrations. The first well contains 0.5 μ M RNA without Lin28 and Dicer. The subsequent wells are for the Dicer assay performed under the same conditions as for the stoichiometric binding assay (in **B**), but with an additional 0.25 U of Dicer enzyme. **(D)** Relative Dicer processing efficiency plotted against Lin28/pre-let-7d-U concentration ratios (n=3).

Supplementary references

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

4.0 Discussion et conclusion

4.1 Importance physiologique du modèle proposé

L'accomplissement final des deux études présentées est la mise au point d'un nouveau modèle décrivant l'assemblage de plusieurs protéines Lin28 sur le même précurseur du miARN let-7g. Dans le premier article (304), nous avons décrit la liaison initiale de la première protéine Lin28 sur la boucle terminale de pre-let-7g. L'un des résultats majeurs obtenus est la haute affinité mesurée entre Lin28 et pre-let-7g. Ce K_d de 0.15 nM était bien différent de ce qui avait été publié précédemment (249,324). Par essais de retard sur gel, nous avons déterminé que c'est la liaison du domaine *NCp7-like* au renflement riche en guanines présent en 5' de TL-let-7g qui rend cette haute affinité possible. De plus, nous avons démontré que ce domaine C-terminal de Lin28 possède une affinité supérieure envers les séquences 5'-GxG-3' ou 5'-GxxG-3'. Dès cette première publication, nous avons observé plusieurs bandes sur gel natif lors de la liaison de forte concentration de Lin28 à pre-let-7g. Puisque l'apparente multimérisation était absente lors de l'utilisation du domaine *NCp7-like*, nous avons avancé l'hypothèse que le domaine *cold shock* était responsable de ce phénomène.

Le deuxième manuscrit (début p. 107) est la suite directe du premier. Sachant que la première protéine liait la région riche en guanines en 5' et ayant observé l'apparition de *supershifts* à haute concentration de protéine, nous avons utilisé différentes techniques pour comprendre où et comment les protéines supplémentaires liaient TL-let-7g. Une combinaison de méthodes biochimiques et biophysiques nous permettent maintenant de proposer un nouveau modèle (Figure 3-7, p. 145) où la liaison du domaine *NCp7-like* de la première protéine au renflement en 5' permet la liaison subséquente du CSD à la séquence voisine 5'-AUGAUAC-3'. Cette première liaison de Lin28 cause une dissociation des brins de la tige d'ARN et rend accessible le deuxième site de haute affinité en 3'. La séquence 5'-GGAG-3' de ce site est liée par le domaine *NCp7-like* de la deuxième protéine, alors que le CSD de cette protéine lie une région voisine en 5'. Nous croyons que la troisième protéine déplace le CSD de la première et lie la séquence 5'-AUGAUAC-3' avec son propre domaine *cold shock*. Le CSD de la première protéine Lin28 se déplace et lie sans doute la région située en 5' du renflement. Ce modèle implique que les CSD soient capables de lier une séquence d'ARN située très près du site de liaison occupé par le domaine *NCp7-like*. Cette liaison rapprochée

est possible et permet sans doute, de stabiliser le complexe tel qu'il a été démontré lors de nos expériences avec les courts ARN de 14 nt. En effet, si les courts ARN contenant les sites de liaison du domaine *NCp7-like* peuvent être liés par la protéine entière, aucun complexe de haute affinité n'a été observé lors de l'utilisation du domaine C-terminal isolé. De ces expériences, nous avons conclu que non seulement le CSD et le domaine *NCp7-like* peuvent lier des séquences d'ARN voisines, mais cette liaison supplémentaire du CSD permet une stabilisation du complexe.

Finalement, puisque cet assemblage séquentiel est également observable chez TL-let-7a-1 et TL-let-7d, nous croyons que ce type de liaison par plusieurs protéines Lin28 est présent chez tous les autres membres de la famille let-7. Cette multimérisation semble spécifique à cette famille de miARN, puisque nous n'avons pas observé ce type d'assemblage avec TL-miR-21a ou TL-miR-99b. Par conséquent, l'assemblage décrit dans notre modèle est peut-être ce qui permet à Lin28 d'inhiber spécifiquement les membres de la famille let-7 (voir sections 4.1.2 et 4.1.3).

4.1.1 Le modèle proposé et la littérature actuelle

Dans notre nouveau modèle, les deux domaines de liaison à l'ARN de Lin28 possèdent des rôles essentiels. Le domaine *NCp7-like* permet une liaison de haute affinité ordonnée de la protéine, alors que le CSD est requis pour la déstabilisation de l'ARN qui rend accessible les sites de liaison supplémentaires. L'implication de chacun des domaines est toute à fait compatible avec les études précédentes et permet de réconcilier plusieurs données provenant de différents groupes qui semblaient parfois contradictoires.

Le domaine *cold shock* de Lin28 est impliqué dans la croissance et le développement cellulaire. En effet, les mutations des acides aminés conservés G43S et M82Q présents dans les motifs RNP causent des défauts de développement chez *C. elegans* (3). Aussi, Lin28BS, un isoforme de Lin28B auquel il manque les 70 premiers acides aminés, n'a aucun effet sur la croissance cellulaire (291). Finalement, l'introduction de mutations ponctuelles à l'intérieur du domaine *cold shock* réduit l'inhibition de la maturation de let-7g par Lin28 (249,325). Le

modèle proposé est tout à fait en accord avec ces résultats qui mettent en évidence la fonction du domaine *cold shock* dans la régulation de la prolifération cellulaire. Ces fonctions sont probablement causées par l'activité chaperonne de ce domaine qui permet la déstabilisation des structures secondaires d'ARN. D'ailleurs, cette activité chaperonne observée lors de nos expériences de FRET confirme ce qui avait été récemment décrit dans une étude utilisant une sonde moléculaire [*molecular beacon*; (327)].

Dans notre modèle, le domaine *NCp7-like* est responsable du contact initial de Lin28 sur pre-let-7g ce qui permet un assemblage ordonné. Ce domaine apporte également la spécificité à la protéine en ciblant les deux régions riches en guanines. Cette spécificité concorde avec ce qui avait été préalablement décrit dans plusieurs études où la mutation d'acides aminés dans ce domaine réduisait la liaison spécifique de la protéine (200,249). Le domaine *NCp7-like* isolé a également été rapporté comme pouvant lier spécifiquement par lui-même, le précurseur du miARN let-7g (304,326). Ceci est en parfait accord avec sa fonction de favoriser l'assemblage ordonné proposée dans notre modèle. Avec ce rôle important, il n'est pas étonnant que plusieurs fonctions biologiques aient été reportées pour ce domaine. En effet, cette extrémité C-terminale de Lin28 est impliquée dans la localisation de la protéine aux *P-bodies* (292), la stimulation de la traduction (321), l'inhibition de la maturation de pre-let-7g (249,325) et dans la stimulation de l'uridylation par Tut4 (199). Nous croyons que c'est la spécificité du domaine *NCp7-like*, dont dépend l'assemblage ordonné de la protéine sur le miARN précurseur, qui explique ces différents rôles majeurs.

Notre modèle permet également de réconcilier plusieurs données qui semblaient contradictoires concernant les sites de liaison de Lin28 sur les pre-let-7. Une première région riche en guanines en 3' avait d'abord été identifiée comme importante pour la liaison de Lin28 (200), ainsi que pour l'inhibition de la maturation du pre-miARN (325). Par la suite, des études structurales utilisant des ARN minimisés ont confirmé l'importance de cette région lors de la liaison par le domaine *NCp7-like* (325,326). Puis, la région riche en guanines en 5' a été mise en évidence comme le site de liaison de Lin28 (324) ou, plus précisément, du domaine *NCp7-like* (304). Finalement, une étude structurale a mise en évidence un troisième site lié, cette fois, par le CSD de Lin28 (325). La présence de ces trois sites de liaison semblait

contradictoire avec l'idée d'un complexe 1:1 et il était difficile d'expliquer le rôle de ces différents sites. En plus d'expliquer cette apparente contradiction, notre modèle d'assemblage séquentiel est en accord avec des essais de digestion par Dicer (248) et de protection aux ribonucléases (324). Dans ces expériences, des nucléotides situés dans différentes régions de TL-let-7g jouaient un rôle dans l'interaction avec Lin28. Nous savons maintenant que ces nucléotides constituaient différents sites de liaison pour la protéine. En résumé, le modèle proposé dans notre deuxième étude concorde avec les différentes études traitant du sujet et permet de réconcilier les apparentes contradictions, concernant les sites de liaison de Lin28, qui s'y trouvaient.

4.1.2 Effet de l'assemblage séquentiel sur Dicer

Jusqu'à présent, les études permettant de clarifier les fonctions physiologiques de l'assemblage séquentiel de Lin28 sur le précurseur de let-7g sont limitées. Nous avons étudié le rôle de cet assemblage sur l'inhibition du clivage de Dicer et avons conclu que la liaison de la première protéine Lin28 était suffisante pour une inhibition efficace. La liaison subséquente de la deuxième et troisième protéine n'influence pas cette inhibition. Il est possible que cette inhibition soit causée par l'encombrement stérique provoquée par la liaison de la première protéine sur le renflement en 5' qui est situé près du site de clivage de Dicer. Néanmoins, il convient de rappeler que cet essai a été effectué avec une protéine Dicer recombinante purifiée. Dans la cellule, Dicer est en complexe avec des cofacteurs tels que TRPB, PACT et possiblement Ago2 (voir section 1.1.5). Par conséquent, le pouvoir régulateur de Lin28 est peut-être différent lorsqu'un tel complexe multimérique interagit avec pre-let-7g. Il serait intéressant de répéter l'expérience de clivage par Dicer effectuée dans la deuxième étude, en ajoutant des cofacteurs à l'endonucléase. Ceci pourrait être fait en utilisant des extraits cellulaires ou des protéines Dicer purifiées à l'aide d'anticorps plutôt qu'une protéine recombinante. Ceci permettrait de vérifier si les liaisons de la deuxième et troisième Lin28 ont un effet sur le clivage par Dicer dans des conditions plus physiologiques.

4.1.3 Possibles rôles physiologiques dans la maturation des miARN

Si la liaison de protéines Lin28 supplémentaires ne semble pas essentielle pour une inhibition *in vitro* efficace de Dicer, plusieurs autres facteurs interagissant avec pre-let-7g pourraient être affectés par cet assemblage séquentiel. L'un des candidats les plus prometteurs est la tutase Tut4. En effet, la poly-uridylation du miARN précurseur dépend de la formation du complexe Lin28/pre-let-7 (199-201). De plus, il a été observé que la liaison de Lin28 à pre-let-7g n'est pas suffisante pour l'ajout d'une queue poly-U si des mutations sont introduites dans le site riche en guanines présent en 3' (200). En considérant ces résultats, nous pouvons émettre l'hypothèse que la liaison de la deuxième protéine Lin28, qui lie ce site en 3', est nécessaire pour la poly-uridylation par Tut4. Une étude sur ce sujet pourrait mettre en lumière la reconnaissance du complexe Lin28/pre-let-7 par la tutase. En effet, il n'est toujours pas connu si la tutase reconnaît un site sur l'ARN rendu accessible par la liaison de Lin28 (203) ou si cette enzyme est spécifique à un stade de l'assemblage séquentiel de Lin28.

Une autre enzyme intervenant dans la maturation de let-7 pouvant être influencée par la liaison de Lin28 est l'endonucléase Drosha et son partenaire DGCR8. Si une nouvelle étude propose que l'inhibition de ce clivage par Lin28 s'effectue par séquestration dans le nucléole (293), plusieurs groupes ont initialement proposé une inhibition plus directe par Lin28 (245,248,339). Il serait intéressant de vérifier si la liaison de plusieurs protéines Lin28 à pri-let-7 est importante pour la séquestration du transcrite primaire lors de l'inhibition de sa maturation. Tout d'abord, la liaison de Lin28 [ou, plus précisément, Lin28B qui est localisée au noyau (293)] à différents pri-let-7 devra être caractérisée. Par la suite, l'utilisation de constructions d'ARN limitant la multimérisation pourra être utilisée dans le cadre d'étude *in vivo* étudiant l'effet de la liaison par Lin28 sur la séquestration de ce transcrite primaire.

Notre modèle permet également d'expliquer la compétition entre la protéine KSRP et Lin28 (177,178). Le site de liaison de KSRP sur pri-let-7 et pre-let-7 étant situé dans le renflement en 5', il est probable que la liaison de la première protéine Lin28 soit responsable de cette compétition. Néanmoins, l'étude de cette compétition dans des conditions contrôlées permettrait de vérifier si la présence de KSRP sur ce site en 5' bloque complètement l'assemblage séquentiel de Lin28 ou si cette dernière peut utiliser le site en 3' et former un

complexe ternaire avec KSRP et pre-let-7. Ce type d'étude permettrait également de comprendre pourquoi la liaison de KSRP au site en 5' stimule le recrutement de Drosha et Dicer, alors que la présence de Lin28 à ce même site inhibe le clivage de Dicer.

Finalement, les multiples effets potentiels qu'a l'assemblage séquentiel de Lin28 sur les précurseurs des miARN let-7 restent à caractériser. La cartographie détaillée de l'interaction entre Tut4 et le complexe multimérique Lin28/pre-let-7 pourrait, à elle seule, constituer un nouveau projet de doctorat.

4.2 Liaison coopérative de Lin28

Lors de notre première étude, de la multimérisation a été observée lors de la liaison de forte concentration de Lin28 à pre-let-7g (304). Cette multimérisation a été la source d'inspiration de notre deuxième étude menant à un modèle d'assemblage séquentiel de Lin28 présentant les caractéristiques d'une interaction coopérative. Ce modèle s'inscrit dans un contexte où une telle coopérativité est décrite chez plusieurs interactions macromoléculaires dont certaines sont présentes dans la voie de maturation des miARN.

4.2.1 La coopérativité chez les macromolécules

La coopérativité est un phénomène économique pour la cellule, puisque cela réduit les concentrations de ligands nécessaires pour obtenir l'effet régulateur désiré (340). En effet, la coopérativité chez les macromolécules décrit l'effet qu'a la liaison d'un premier ligand à un récepteur sur la liaison de ligands supplémentaires. Une coopérativité positive est observée si ce premier ligand facilite la liaison des ligands suivants. Dans le cas inverse, le terme de coopérativité négative est employé. Ce phénomène peut être utilisé pour décrire l'assemblage de ligands identiques sur une protéine composée de plusieurs sous-unités. C'est ce type de coopérativité qui est retrouvée dans l'hémoglobine où quatre molécules d'oxygène lient tour à tour les quatre sous-unités de la protéine (341,342). Aussi, les transitions dans le repliement d'une protéine, la formation d'une double hélice d'acides nucléiques ou l'assemblage de complexes formés de plusieurs macromolécules différentes peuvent également être considérés comme des manifestations de coopérativité (343).

Deux mécanismes peuvent causer le phénomène de coopérativité (343). D'abord, la liaison du premier ligand peut permettre la stabilisation du deuxième à l'aide de contacts entre ces derniers (voir Figure 4-1A). Dans cet exemple, la liaison de B est facilitée par la présence de C et un $\Delta\Delta G$ de 2 est observé. Avec ce mécanisme, la liaison d'un premier ligand augmente réellement l'affinité du second ligand envers son site d'interaction sur le récepteur. Par contre, avec le second mécanisme responsable du phénomène de coopérativité (voir Figure 4-1B), la

liaison du premier ligand est défavorisée puisque ce dernier doit induire un changement de conformation chez le récepteur. Dans ce cas, il n'y a pas de contact direct entre les deux ligands et la coopérativité est causée par le changement de conformation apportée par la première liaison. Ce second mécanisme est fréquemment retrouvé chez les acides nucléiques où la liaison d'une première protéine perturbe l'empilement des bases par une déformation de l'hélice, facilitant ainsi les liaisons subséquentes (344). Dans bien des cas, la coopérativité présente dans un complexe macromoléculaire est causée par un mélange de ces deux mécanismes (343,344).

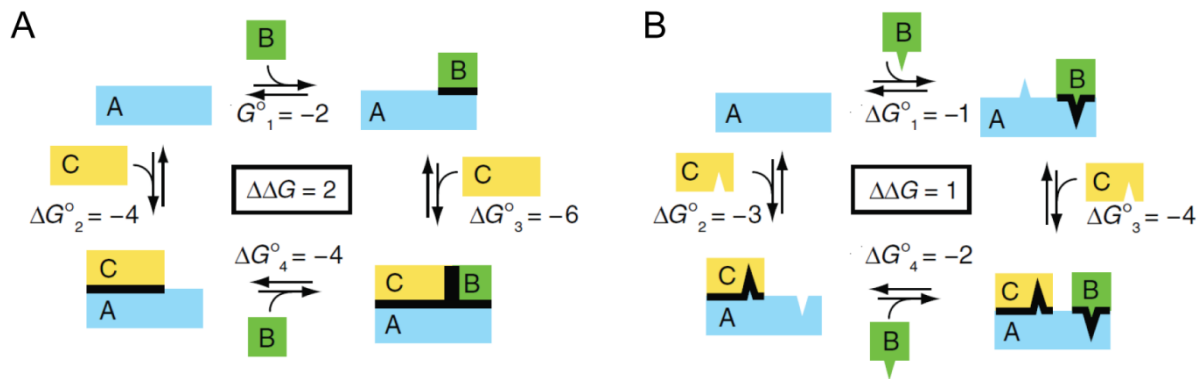


Figure 4-1. Mécanismes expliquant la coopérativité chez les macromolécules (A) La liaison du premier ligand permet de stabiliser la liaison du deuxième à l'aide de contacts entre chacun des ligands. (B) La liaison du premier ligand est défavorisée puisqu'il doit induire un changement de conformation chez le partenaire commun des ligands. Il n'y a pas de contact direct entre les ligands. Adaptée de (343).

La meilleure méthode pour caractériser la coopérativité d'un complexe macromoléculaire est la différence d'énergie libre ($\Delta\Delta G$) entre la liaison d'un ligand à son récepteur en présence et en absence du second ligand (340). Par exemple, dans la Figure 4-1A, le ligand B a une énergie libre (ΔG^0) de -2 en absence du second ligand et de -4 en présence de celui-ci. Ce $\Delta\Delta G$ de 2 indique que la liaison de B à A est favorisée en présence de C et qu'il y a coopérativité positive entre ces deux ligands. Pour mener à bien ce type de caractérisation, il est essentiel de connaître les constantes de dissociation (ou les ΔG^0) pour chacune des

interactions intermédiaires. Si ces données sont relativement aisées à obtenir avec un complexe formé de trois molécules différentes, il en est autrement pour un complexe formé d'un récepteur lié par plusieurs ligands identiques (340). En effet, avec un complexe tel que Lin28 et pre-let-7, la caractérisation de l'interaction de Lin28 avec chacun des sites est difficile, puisque l'on ne peut pas observer chaque liaison indépendamment l'une de l'autre. Par exemple, il serait difficile de mesurer l'affinité qu'a Lin28 pour le deuxième site en absence de ligand au premier site de pre-let-7g. Dans de simples cas, il est parfois possible de combiner différentes techniques pour obtenir les paramètres thermodynamiques requis. Par exemple, la constante de dissociation globale pour la liaison d'un équivalent de ligand à différents sites peut être déterminée par essais de retard sur gel. Puis, des titrations RMN peuvent ensuite indiquer le ratio d'occupation pour chaque site suite à la liaison de cet équivalent (345). Les ΔG^0 pour la liaison de chaque ligand (ΔG^0_1 et ΔG^0_2 , voir Figure 4-1) peuvent ensuite être calculés. Par contre, Lin28 n'ayant pas la même affinité pour chacun des sites, cette méthode ne peut être utilisée, puisque le premier site sera entièrement occupé avant la liaison du ligand au deuxième et troisième site. Bref, il n'est actuellement probablement pas possible d'effectuer une caractérisation détaillée de l'assemblage coopératif des trois protéines Lin28 sur un même TL-let-7.

4.2.2 Coopérativité locale à l'intérieur d'une protéine

Lorsqu'une protéine est composée de deux domaines pouvant lier séparément la même cible, il est possible d'observer de la coopérativité locale entre les domaines de la protéine pour la liaison de cette cible. Pour caractériser la coopérativité locale entre deux régions liant l'ARN du domaine RBD1 lors de la liaison de sa cible, Kranz *et al.* (346) ont mesuré l'affinité de nombreux mutants de la protéine après avoir quantifié leur stabilité par dichroïsme circulaire en présence d'hydrochlorure de guanidine (GndHCl). Cette méthode leur a permis de comparer le ΔG^0 de chacune des régions de ce domaine en présence et en absence de la seconde région lors de la liaison à l'ARN cible. En comparant les ΔG^0 pour chacune des

régions à celui de la protéine entière liant l'ARN, ils ont pu quantifier la coopérativité locale entre les deux régions de ce domaine.

Bien que Lin28 soit de plus grande taille que RBD1, ce type d'expérience pourrait être utilisé pour mieux comprendre l'apparente coopérativité observée lors de la liaison de la première protéine Lin28 à TL-let-7g (voir Figure 2-2C du chapitre 2.0, p. 90). Cette coopérativité pourrait être causée par une interaction locale entre le domaine *NCp7-like* et le domaine *cold shock* suite à leur liaison à l'ARN. Malgré que ces deux domaines isolés soient capables de lier l'ARN, il reste possible que leur liaison ne soit pas totalement indépendante de la présence du second domaine. Tout comme pour RBD1, les domaines de liaison de l'ARN de Lin28 communiquent peut-être ensemble via l'ARN à l'aide d'un réseau d'empilement de bases hydrophobes ou par une déstabilisation initiale de l'ARN. Pour mener à bien ce type d'études, il est important que les sites de liaison spécifiques à chaque domaine, lorsque présents dans la protéine pleine longueur, restent les mêmes pour les domaines individuels. Par contre, la faible spécificité du domaine *cold shock* et la présence du motif riche en lysine et arginine sont des sources de complications importantes puisque les sites d'interactions de ces derniers sont sans doute variables en absence du domaine *NCp7-like*. La mise au point d'un ARN consensus minimal réduisant les interactions non spécifiques permettrait peut-être d'éliminer ce problème (voir section 4.3.3).

4.2.3 La coopérativité et les interactions entre protéines et acides nucléiques

Lin28 n'est pas la seule protéine à lier son ARN cible avec coopérativité. Par exemple, les dimères de DGCR8 (voir introduction, p. 7) s'associent pour former des trimères de dimères de façon coopérative lors de la liaison des transcrits primaires dans le noyau. En fait, ce serait cette coopérativité qui rendrait DGCR8 spécifique envers les pri-miARN (31). Il a été démontré que la queue terminale de cette protéine est nécessaire pour obtenir une trimérisation efficace, mais le mécanisme exact responsable de la coopérativité (voir Figure 4-1) reste à définir.

Les protéines *cold shock* bactériennes auxquelles ressemble le CSD de Lin28 lient également l'ARN de façon coopérative. D'ailleurs, la protéine CspA possède aussi une activité chaperonne permettant de déstabiliser les structures secondaires voisines avec peu de spécificité (297). Encore une fois, les mécanismes moléculaires permettant d'expliquer cette coopérativité sont mal définis. Par contre, nous pouvons avancer l'hypothèse que, tout comme pour Lin28, la liaison initiale provoque un changement de conformation chez l'ARN, facilitant ainsi la liaison d'une deuxième CspA qui propage ensuite cette déstabilisation le long de la structure d'acides nucléiques.

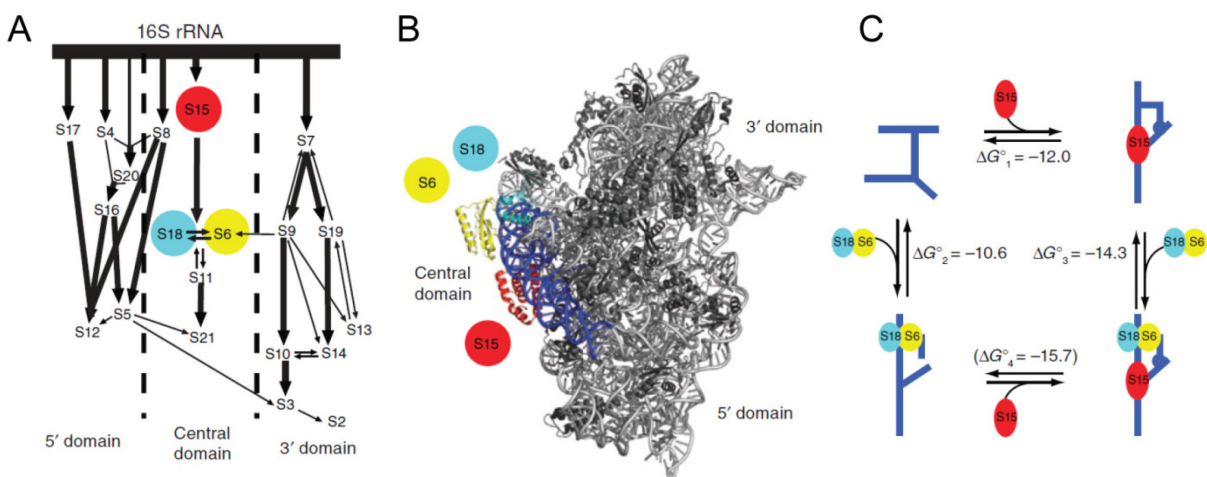


Figure 4-2. La coopérativité dans l'assemblage de l'unité 30S du ribosome. La protéine S15 lie l'ARN 16S avant le dimère S6/S18 puisque cette route est thermodynamiquement favorable. (A) Carte de l'assemblage des différents composants de l'unité 30S. (B) Structure de cette unité avec les protéines S6, S15 et S18 mises en évidence. (C) Cycle thermodynamique pour la formation du complexe ARNr 16S/S6/S15/S18. Adaptée de (343).

La caractérisation thermodynamique de la coopérativité peut également permettre de comprendre l'assemblage d'un complexe macromoléculaire. L'unité 30S du ribosome est composée de vingt protéines qui respectent un ordre d'assemblage précis lors de leur liaison à l'ARN ribosomal (ARNr) 16S. Par exemple, la protéine S15 peut lier directement l'ARNr 16S, mais les protéines S6 et S18 sont interdépendantes et ont besoin de la présence préalable de S15 (voir Figure 4-2A). La structure cristalline démontre que S6 et S18 forment un

hétérodimère ce qui explique leur interdépendance. Par contre, la protéine S15 n'a aucun contact direct avec ces protéines [voir Figure 4-2B; (343)]. Des études thermodynamiques détaillées utilisant l'ITC et des essais de retard sur gel ont déterminé, en se basant sur différentes études structurales, que la protéine S15 induit un changement de conformation chez l'ARNr 16S, ce qui facilite la liaison de l'hétérodimère S6/S18 [voir Figure 4-2C; (347,348)]. En résumé, la combinaison d'un bon modèle d'assemblage, de structures à haute résolution et d'études thermodynamiques caractérisant la coopérativité, permettent d'obtenir une excellente compréhension concernant l'assemblage d'importants complexes macromoléculaires. La combinaison de méthodes utilisées, pour comprendre l'assemblage coopératif de l'unité 30S du ribosome, pourrait servir d'inspiration pour les études futures visant à caractériser l'assemblage de Lin28, pre-let-7 et la tutase Tut4.

4.3 Perspectives

Dans cette section, différentes suites possibles à ce projet seront proposées. La multimérisation de Lin28 sur l'ARN, décrite dans cette thèse, peut servir de base à plusieurs opportunités pour ce champ d'études.

4.3.1 Interactions de Lin28 avec Tut4, Drosha et KSRP

Tel que mentionné précédemment, une suite intéressante à ce projet est la caractérisation de l'interaction entre Tut4 et le complexe Lin28/pre-let-7. Puisqu'il est connu que la liaison de Lin28 à l'ARN est essentielle pour la reconnaissance par Tut4 (199-201), il devient important de bien définir quels sont les facteurs permettant cette interaction. Ce type d'étude devra s'intéresser à l'effet de l'assemblage séquentiel de Lin28 sur la liaison et la poly-uridylation par la tutase. Le même type d'expérience pourrait par la suite être entrepris pour évaluer l'effet de l'assemblage séquentiel de Lin28 sur la liaison de Drosha et KSRP à pri/pre-let-7. À terme, ces résultats clarifieront le rôle et le mode de régulation de Lin28 dans la voie de maturation des membres de la famille let-7.

4.3.2 Études structurales

Différentes études structurales se sont déjà intéressées au complexe Lin28/pre-let-7 (325-327). Bien que ces études aient apporté des informations importantes concernant la formation de ce complexe, elles ont toutes assumé un ratio stœchiométrique de 1:1 lors de la préparation de leurs expériences. Aussi, dans tous les cas, le site en 5' n'a pas été inclus dans l'ARN minimisé utilisé. L'obtention de cristaux avec des complexes Lin28/pre-let-7 à différents ratios stœchiométriques permettraient d'obtenir des informations à haute résolution concernant l'assemblage séquentiel de Lin28. Éventuellement, ces nouvelles études structurales pourraient également inclure la protéine Tut4. Toutefois, les conditions permettant

la reconnaissance du complexe Lin28/pre-let-7 par cette tutase devront préalablement être définies.

4.3.3 Séquence consensus

L'identification d'une séquence d'ARN consensus reconnue par Lin28 a fait l'objet de plusieurs études. Des études biochimiques, fonctionnelles et structurales ont identifié les séquences 5'-GGAG-3' et 5'-UGAGGG-3' comme les sites de liaison du domaine *NCp7-like* sur certains précurseurs de let-7, alors que les séquences 5'-NGNGAYNNN-3' et 5'-GUNNUNN-3' ont été proposées comme consensus pour le CSD (voir section 1.6.2, p. 54). De plus, les études s'intéressant aux interactions entre Lin28 et les ARN messagers ont associé des séquences riches en guanine au domaine *NCp7-like* et d'autres riches en uracile au CSD (322,325). Pourtant, la séquence consensus de la protéine entière n'a pas encore été étudiée de façon détaillée. Il a récemment été démontré que les sites de liaison de chacun des domaines peuvent être situés très près l'un de l'autre, sur la même molécule d'ARN et que le CSD est habituellement situé en 5' du domaine *NCp7-like* (322,325). Ceci est en accord avec notre deuxième étude, où nous avons observé que Lin28 peut lier de courts brins d'ARN de 14 nucléotides, mais est incapable de lier des brins de huit nucléotides. Toutefois, la faible spécificité de séquence du CSD et de la région riche en arginine et lysine complique l'utilisation d'ARN simple brin pour la détermination d'un consensus. Un ARN composé de deux régions simple brin, séparées l'une de l'autre par une double hélice, pourrait permettre de faciliter l'interprétation des résultats.

Afin d'étudier simultanément la liaison de chacun des domaines de Lin28, la mise au point d'un ARN consensus minimal et structuré pourrait s'avérer intéressante. Un tel ARN devra être composé d'une extrémité simple brin pour la liaison du domaine *NCp7-like* et une tige stabilisée complétée par une boucle apicale servant de site de liaison pour le CSD (voir Figure 4-3). La structure proposée de cet ARN consensus est fortement inspiré de TL-let-7d, une tige terminale avec laquelle Lin28 forme un complexe 1:1 (voir p. 140) dont la structure à haute résolution est connue (325). Afin de déterminer un consensus pour la liaison de Lin28,

une librairie de ce type d'ARN structuré pourrait être constituée. Tous les ARN de cette librairie auraient une structure secondaire similaire, mais la longueur et la composition de la boucle apicale, ainsi que celles de l'extrémité simple brin seraient variables. De plus, cette extrémité simple brin pourrait être positionnée en 5' ou en 3', alors que la longueur de la tige stabilisée pourrait être modifiée. Pour examiner la liaison de Lin28 à un aussi grand nombre d'ARN, les essais de liaison devront se faire en commun. En ajoutant la protéine directement à la librairie totale, il est ensuite possible de recueillir les ARN pour lesquels Lin28 a la plus grande affinité. Par la suite, l'ajout de protéinase et la liaison d'adaptateurs aux ARN copurifiés rendront possible la transcription inverse, l'amplification et le séquençage de ces ARN. Cette technique devrait permettre d'identifier un ARN structuré minimal comportant des séquences consensus pour les domaines *cold shock* et *NCp7-like*. Ces connaissances permettront de mieux définir un consensus pour Lin28 ce qui mènera à la prédiction et à l'identification *in vivo* de nouvelles cibles de Lin28.

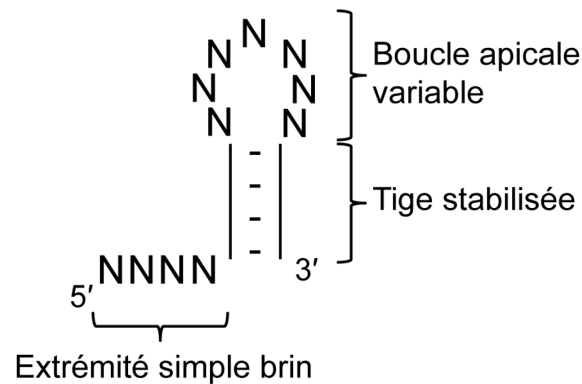


Figure 4-3. ARN consensus minimal et structuré. Le nombre de nucléotides composant la boucle apicale, la tige stabilisée ainsi que l'extrémité simple brin sont variables. L'extrémité simple brin peut être positionnée en 5' ou en 3'. La lettre N indique que l'un des quatre nucléotides sera inséré aléatoirement à cette position.

4.3.4 Les autres membres de la famille let-7

L'interaction entre la boucle terminale de let-7g et Lin28 a été caractérisée en détail dans les études présentées précédemment. Par contre, la liaison de Lin28 aux autres membres de la famille let-7 a seulement été brièvement abordée. Nous savons que la protéine peut lier de façon séquentielle TL-let-7a-1 et TL-let-7d pour former des complexes 1:3, mais les autres membres de la famille let-7 n'ont pas fait l'objet d'examen. De plus, l'affinité de Lin28 envers tous les autres membres de cette famille reste à déterminer, tout comme le site de liaison initial qui n'est peut-être pas toujours la région 5' riche en guanine. Finalement, il serait intéressant de vérifier comment le clivage des différents pre-let-7 par Dicer est inhibé par Lin28. En résumé, les méthodes optimisées dans cette thèse pourraient facilement être appliquées aux autres membres de la famille let-7.

4.3.5 Autres rôles potentiels pour l'assemblage séquentiel de Lin28

Sans être la suite directe à cette thèse, l'étude des mécanismes de reconnaissance de l'ARN par Lin28B s'avère nécessaire. Une récente publication (293) propose des fonctions différentes pour Lin28(A) et Lin28B. Cette dernière agirait dans le noyau et séquestrerait les transcrits primaires dans le nucléole. Jusqu'à présent, peu de groupes se sont intéressés au mode de reconnaissance de cet homologue de Lin28 et il serait important de vérifier si l'assemblage séquentiel sur les ARN cibles y est conservé.

La protéine Lin28 joue également un rôle dans la régulation de la traduction (voir section 1.5.2, p. 47). Dans la plupart des cas, la liaison de Lin28 aux ARNm permet de stimuler la traduction de ceux-ci par le recrutement de la *RNA Helicase A* (RHA). Cette enzyme stimulerait la traduction en déstabilisant certaines structures secondaires présentes sur l'ARNm (309,319). Puisque notre modèle propose que la présence de Lin28 soit suffisante pour déstabiliser de telles structures, une étude s'intéressant à l'effet de la liaison de Lin28 sur la structure secondaire de ces ARNm cibles est à envisager. D'ailleurs, la liaison de plusieurs protéines Lin28 sur le même élément sensible à Lin28 provenant de l'ARNm Oct4 a déjà été observée (321). Il est donc possible que Lin28 utilise le même type d'assemblage séquentiel

sur certains ARNm afin de stimuler la traduction de ces derniers avec ou sans l'assistance de RHA. Aussi, tel qu'il a déjà été proposé (349), la présence de plusieurs protéines Lin28 sur un même ARNm pourrait en stimuler l'expression par le déplacement des RISC qui y sont liés. De plus, cet assemblage séquentiel sur les ARNm pourrait servir de mécanisme régulateur. Ceci pourrait expliquer pourquoi la liaison de Lin28 à certains ARNm stimule leur traduction, alors qu'elle inhibe l'expression des ARNm dont la protéine est destinée au réticulum endoplasmique (320). En résumé, l'assemblage séquentiel de plusieurs protéines Lin28 proposé dans cette thèse s'applique peut-être à la liaison de Lin28 aux ARNm. Une nouvelle étude sur ce sujet permettrait de mieux comprendre les multiples fonctions de Lin28 dans la régulation de la traduction.

4.4 Conclusion générale

En utilisant différentes techniques biochimiques et biophysiques, les études présentées dans cette thèse ont permis de caractériser une interaction protéine-ARN de haute affinité entre Lin28 et la boucle terminale du microARN précurseur let-7g. La surprenante haute affinité initialement observée est maintenant acceptée par la communauté (293). Nous avons proposé le premier modèle expliquant l'assemblage séquentiel de Lin28 sur le précurseur du microARN let-7g. Notre modèle devra dorénavant être considéré lors de la planification d'expériences, ce qui permettra sans doute, de clarifier le rôle physiologique d'une telle multimérisation.

Les travaux présentés ici, permettent de mieux comprendre les mécanismes régissant l'interaction entre Lin28 et pre-let-7. Ces mécanismes associés à de bonnes études structurales et fonctionnelles permettront, à terme, de concevoir des inhibiteurs ciblant spécifiquement l'une des nombreuses fonctions cellulaires de Lin28.

Bibliographie

1. Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, **75**, 843-854.
2. Wightman, B., Ha, I. and Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell*, **75**, 855-862.
3. 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.
4. Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R. and Ruvkun, G. (2000) The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, **403**, 901-906.
5. Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P. *et al.* (2000) Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. *Nature*, **408**, 86-89.
6. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853-858.
7. Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, **294**, 858-862.
8. Lee, R.C. and Ambros, V. (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, **294**, 862-864.
9. Lee, Y., Jeon, K., Lee, J.T., Kim, S. and Kim, V.N. (2002) MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J*, **21**, 4663-4670.
10. Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H. and Kim, V.N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*, **23**, 4051-4060.
11. Borchert, G.M., Lanier, W. and Davidson, B.L. (2006) RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol*, **13**, 1097-1101.

12. Schanen, B.C. and Li, X. (2011) Transcriptional regulation of mammalian miRNA genes. *Genomics*, **97**, 1-6.
13. Song Gao, J., Zhang, Y., Li, M., Tucker, L.D., Machan, J.T., Quesenberry, P., Rigoutsos, I. and Ramratnam, B. (2010) Atypical transcription of microRNA gene fragments. *Nucleic Acids Res*, **38**, 2775-2787.
14. Kim, V.N., Han, J. and Siomi, M.C. (2009) Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol*, **10**, 126-139.
15. Spizzo, R., Nicoloso, M.S., Lupini, L., Lu, Y., Fogarty, J., Rossi, S., Zagatti, B., Fabbri, M., Veronese, A., Liu, X. *et al.* (2010) miR-145 participates with TP53 in a death-promoting regulatory loop and targets estrogen receptor-alpha in human breast cancer cells. *Cell Death Differ*, **17**, 246-254.
16. Wang, Z., Lin, S., Li, J.J., Xu, Z., Yao, H., Zhu, X., Xie, D., Shen, Z., Sze, J., Li, K. *et al.* (2011) MYC protein inhibits transcription of the microRNA cluster MC-let-7a-1~let-7d via noncanonical E-box. *J Biol Chem*, **286**, 39703-39714.
17. Sampson, V.B., Rong, N.H., Han, J., Yang, Q., Aris, V., Soteropoulos, P., Petrelli, N.J., Dunn, S.P. and Krueger, L.J. (2007) MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. *Cancer Res*, **67**, 9762-9770.
18. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S. *et al.* (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature*, **425**, 415-419.
19. Filippov, V., Solovyev, V., Filippova, M. and Gill, S.S. (2000) A novel type of RNase III family proteins in eukaryotes. *Gene*, **245**, 213-221.
20. Blaszczyk, J., Tropea, J.E., Bubunenko, M., Routzahn, K.M., Waugh, D.S., Court, D.L. and Ji, X. (2001) Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure*, **9**, 1225-1236.
21. Lee, Y., Han, J., Yeom, K.H., Jin, H. and Kim, V.N. (2006) Drosha in primary microRNA processing. *Cold Spring Harb Symp Quant Biol*, **71**, 51-57.
22. Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E. and Filipowicz, W. (2004) Single processing center models for human Dicer and bacterial RNase III. *Cell*, **118**, 57-68.

23. Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H. and Kim, V.N. (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev*, **18**, 3016-3027.
24. Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F. and Hannon, G.J. (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature*, **432**, 231-235.
25. Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N. and Shiekhattar, R. (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature*, **432**, 235-240.
26. Landthaler, M., Yalcin, A. and Tuschl, T. (2004) The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. *Curr Biol*, **14**, 2162-2167.
27. Yeom, K.H., Lee, Y., Han, J., Suh, M.R. and Kim, V.N. (2006) Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. *Nucleic Acids Res*, **34**, 4622-4629.
28. Faller, M., Matsunaga, M., Yin, S., Loo, J.A. and Guo, F. (2007) Heme is involved in microRNA processing. *Nat Struct Mol Biol*, **14**, 23-29.
29. Senturia, R., Faller, M., Yin, S., Loo, J.A., Cascio, D., Sawaya, M.R., Hwang, D., Clubb, R.T. and Guo, F. (2010) Structure of the dimerization domain of DiGeorge critical region 8. *Protein Sci*, **19**, 1354-1365.
30. 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.
31. Faller, M., Toso, D., Matsunaga, M., Atanasov, I., Senturia, R., Chen, Y., Zhou, Z.H. and Guo, F. (2010) DGCR8 recognizes primary transcripts of microRNAs through highly cooperative binding and formation of higher-order structures. *RNA*, **16**, 1570-1583.
32. Kim, Y.K. and Kim, V.N. (2007) Processing of intronic microRNAs. *EMBO J*, **26**, 775-783.

33. Morlando, M., Ballarino, M., Gromak, N., Pagano, F., Bozzoni, I. and Proudfoot, N.J. (2008) Primary microRNA transcripts are processed co-transcriptionally. *Nat Struct Mol Biol*, **15**, 902-909.
34. Pawlicki, J.M. and Steitz, J.A. (2008) Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production. *J Cell Biol*, **182**, 61-76.
35. Ruby, J.G., Jan, C.H. and Bartel, D.P. (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature*, **448**, 83-86.
36. Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M. and Lai, E.C. (2007) The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell*, **130**, 89-100.
37. Berezikov, E., Chung, W.J., Willis, J., Cuppen, E. and Lai, E.C. (2007) Mammalian mirtron genes. *Mol Cell*, **28**, 328-336.
38. Bohnsack, M.T., Czaplinski, K. and Gorlich, D. (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA*, **10**, 185-191.
39. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E. and Kutay, U. (2004) Nuclear export of microRNA precursors. *Science*, **303**, 95-98.
40. Calado, A., Treichel, N., Muller, E.C., Otto, A. and Kutay, U. (2002) Exportin-5-mediated nuclear export of eukaryotic elongation factor 1A and tRNA. *EMBO J*, **21**, 6216-6224.
41. Zeng, Y. and Cullen, B.R. (2004) Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res*, **32**, 4776-4785.
42. Yi, R., Doehle, B.P., Qin, Y., Macara, I.G. and Cullen, B.R. (2005) Overexpression of exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs. *RNA*, **11**, 220-226.
43. Winter, J. and Diederichs, S. (2011) MicroRNA biogenesis and cancer. *Methods Mol Biol*, **676**, 3-22.
44. Bernstein, E., Caudy, A.A., Hammond, S.M. and Hannon, G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, **409**, 363-366.

45. Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J. and Plasterk, R.H. (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev*, **15**, 2654-2659.
46. Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T. and Zamore, P.D. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science*, **293**, 834-838.
47. Knight, S.W. and Bass, B.L. (2001) A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science*, **293**, 2269-2271.
48. Provost, P., Dishart, D., Doucet, J., Frendewey, D., Samuelsson, B. and Radmark, O. (2002) Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J*, **21**, 5864-5874.
49. Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V. and Hannon, G.J. (2003) Dicer is essential for mouse development. *Nat Genet*, **35**, 215-217.
50. Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J. and Carthew, R.W. (2004) Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*, **117**, 69-81.
51. Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E. and Carrington, J.C. (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol*, **2**, E104.
52. Dlakic, M. (2006) DUF283 domain of Dicer proteins has a double-stranded RNA-binding fold. *Bioinformatics*, **22**, 2711-2714.
53. 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.
54. Lau, P.W., Guiley, K.Z., De, N., Potter, C.S., Carragher, B. and MacRae, I.J. (2012) The molecular architecture of human Dicer. *Nat Struct Mol Biol*, **19**, 436-440.

55. Cerutti, L., Mian, N. and Bateman, A. (2000) Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends Biochem Sci*, **25**, 481-482.
56. Carmell, M.A., Xuan, Z., Zhang, M.Q. and Hannon, G.J. (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev*, **16**, 2733-2742.
57. Ma, J.B., Ye, K. and Patel, D.J. (2004) Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, **429**, 318-322.
58. Lingel, A., Simon, B., Izaurralde, E. and Sattler, M. (2004) Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol*, **11**, 576-577.
59. Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L. and Zhou, M.M. (2003) Structure and conserved RNA binding of the PAZ domain. *Nature*, **426**, 468-474.
60. Park, J.E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J. and Kim, V.N. (2011) Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature*, **475**, 201-205.
61. Soifer, H.S., Sano, M., Sakurai, K., Chomchan, P., Saetrom, P., Sherman, M.A., Collingwood, M.A., Behlke, M.A. and Rossi, J.J. (2008) A role for the Dicer helicase domain in the processing of thermodynamically unstable hairpin RNAs. *Nucleic Acids Res*, **36**, 6511-6522.
62. Ma, E., MacRae, I.J., Kirsch, J.F. and Doudna, J.A. (2008) Autoinhibition of human dicer by its internal helicase domain. *J Mol Biol*, **380**, 237-243.
63. Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H. and Tomari, Y. (2011) Recognition of the pre-miRNA structure by Drosophila Dicer-1. *Nat Struct Mol Biol*, **18**, 1153-1158.
64. Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K. and Shiekhattar, R. (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, **436**, 740-744.
65. Haase, A.D., Jaskiewicz, L., Zhang, H., Laine, S., Sack, R., Gatignol, A. and Filipowicz, W. (2005) TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep*, **6**, 961-967.

66. Lee, Y., Hur, I., Park, S.Y., Kim, Y.K., Suh, M.R. and Kim, V.N. (2006) The role of PACT in the RNA silencing pathway. *EMBO J*, **25**, 522-532.
67. Maniataki, E. and Mourelatos, Z. (2005) A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev*, **19**, 2979-2990.
68. Gregory, R.I., Chendrimada, T.P., Cooch, N. and Shiekhattar, R. (2005) Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell*, **123**, 631-640.
69. Lingel, A., Simon, B., Izaurralde, E. and Sattler, M. (2003) Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. *Nature*, **426**, 465-469.
70. Song, J.J., Smith, S.K., Hannon, G.J. and Joshua-Tor, L. (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*, **305**, 1434-1437.
71. Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T. and Patel, D.J. (2005) Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature*, **434**, 666-670.
72. Parker, J.S., Roe, S.M. and Barford, D. (2005) Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature*, **434**, 663-666.
73. Yuan, Y.R., Pei, Y., Ma, J.B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.Y., Dauter, Z., Tuschl, T. and Patel, D.J. (2005) Crystal structure of *A. aeolicus* argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol Cell*, **19**, 405-419.
74. Leuschner, P.J., Ameres, S.L., Kueng, S. and Martinez, J. (2006) Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep*, **7**, 314-320.
75. Diederichs, S. and Haber, D.A. (2007) Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell*, **131**, 1097-1108.
76. Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L. and Hannon, G.J. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, **305**, 1437-1441.

77. Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G. and Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell*, **15**, 185-197.
78. Azuma-Mukai, A., Oguri, H., Mituyama, T., Qian, Z.R., Asai, K., Siomi, H. and Siomi, M.C. (2008) Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Proc Natl Acad Sci U S A*, **105**, 7964-7969.
79. Tolia, N.H. and Joshua-Tor, L. (2007) Slicer and the argonautes. *Nat Chem Biol*, **3**, 36-43.
80. Matranga, C., Tomari, Y., Shin, C., Bartel, D.P. and Zamore, P.D. (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell*, **123**, 607-620.
81. Fukuda, T., Yamagata, K., Fujiyama, S., Matsumoto, T., Koshida, I., Yoshimura, K., Mihara, M., Naitou, M., Endoh, H., Nakamura, T. *et al.* (2007) DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nat Cell Biol*, **9**, 604-611.
82. Tomari, Y., Matranga, C., Haley, B., Martinez, N. and Zamore, P.D. (2004) A protein sensor for siRNA asymmetry. *Science*, **306**, 1377-1380.
83. Preall, J.B. and Sontheimer, E.J. (2005) RNAi: RISC gets loaded. *Cell*, **123**, 543-545.
84. Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N. and Zamore, P.D. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, **115**, 199-208.
85. Khvorova, A., Reynolds, A. and Jayasena, S.D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**, 209-216.
86. Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B. and Bartel, D.P. (2003) The microRNAs of *Caenorhabditis elegans*. *Genes Dev*, **17**, 991-1008.
87. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. and Burge, C.B. (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787-798.
88. Doench, J.G. and Sharp, P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev*, **18**, 504-511.

89. Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T. and Patel, D.J. (2008) Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature*, **456**, 921-926.
90. Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15-20.
91. Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215-233.
92. Nielsen, C.B., Shomron, N., Sandberg, R., Hornstein, E., Kitzman, J. and Burge, C.B. (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA*, **13**, 1894-1910.
93. Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P. and Bartel, D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell*, **27**, 91-105.
94. Abbott, A.L., Alvarez-Saavedra, E., Miska, E.A., Lau, N.C., Bartel, D.P., Horvitz, H.R. and Ambros, V. (2005) The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in *Caenorhabditis elegans*. *Dev Cell*, **9**, 403-414.
95. Saetrom, P., Heale, B.S., Snove, O., Jr., Aagaard, L., Alluin, J. and Rossi, J.J. (2007) Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Res*, **35**, 2333-2342.
96. Tay, Y., Zhang, J., Thomson, A.M., Lim, B. and Rigoutsos, I. (2008) MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, **455**, 1124-1128.
97. Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. and Tuschl, T. (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell*, **110**, 563-574.
98. Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W. and Sontheimer, E.J. (2004) A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in *Drosophila*. *Cell*, **117**, 83-94.

99. Pratt, A.J. and MacRae, I.J. (2009) The RNA-induced silencing complex: a versatile gene-silencing machine. *J Biol Chem*, **284**, 17897-17901.
100. Pillai, R.S., Artus, C.G. and Filipowicz, W. (2004) Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *RNA*, **10**, 1518-1525.
101. Eulalio, A., Huntzinger, E. and Izaurralde, E. (2008) GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat Struct Mol Biol*, **15**, 346-353.
102. Yao, B., Li, S., Lian, S.L., Fritzler, M.J. and Chan, E.K. (2011) Mapping of Ago2-GW182 functional interactions. *Methods Mol Biol*, **725**, 45-62.
103. El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.A., Jacobsen, S.E., Cooke, R. and Lagrange, T. (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev*, **21**, 2539-2544.
104. Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R. and Filipowicz, W. (2011) miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. *Nat Struct Mol Biol*, **18**, 1218-1226.
105. Fabian, M.R., Cieplak, M.K., Frank, F., Morita, M., Green, J., Srikumar, T., Nagar, B., Yamamoto, T., Raught, B., Duchaine, T.F. *et al.* (2011) miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. *Nat Struct Mol Biol*, **18**, 1211-1217.
106. Braun, J.E., Huntzinger, E., Fauser, M. and Izaurralde, E. (2011) GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Mol Cell*, **44**, 120-133.
107. Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P. and Izaurralde, E. (2006) mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev*, **20**, 1885-1898.
108. Kulkarni, M., Ozgur, S. and Stoecklin, G. (2010) On track with P-bodies. *Biochem Soc Trans*, **38**, 242-251.

109. Fabian, M.R. and Sonenberg, N. (2012) The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat Struct Mol Biol*, **19**, 586-593.
110. Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E. and Filipowicz, W. (2005) Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science*, **309**, 1573-1576.
111. Humphreys, D.T., Westman, B.J., Martin, D.I. and Preiss, T. (2005) MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc Natl Acad Sci U S A*, **102**, 16961-16966.
112. Mathonnet, G., Fabian, M.R., Svitkin, Y.V., Parsyan, A., Huck, L., Murata, T., Biffo, S., Merrick, W.C., Darzynkiewicz, E., Pillai, R.S. *et al.* (2007) MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science*, **317**, 1764-1767.
113. Djuranovic, S., Nahvi, A. and Green, R. (2011) A parsimonious model for gene regulation by miRNAs. *Science*, **331**, 550-553.
114. Djuranovic, S., Zinchenko, M.K., Hur, J.K., Nahvi, A., Brunelle, J.L., Rogers, E.J. and Green, R. (2010) Allosteric regulation of Argonaute proteins by miRNAs. *Nat Struct Mol Biol*, **17**, 144-150.
115. Kiriakidou, M., Tan, G.S., Lamprinaki, S., De Planell-Saguer, M., Nelson, P.T. and Mourelatos, Z. (2007) An mRNA m7G cap binding-like motif within human Ago2 represses translation. *Cell*, **129**, 1141-1151.
116. Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T., Svitkin, Y.V., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J.A. *et al.* (2009) Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. *Mol Cell*, **35**, 868-880.
117. Zekri, L., Huntzinger, E., Heimstadt, S. and Izaurralde, E. (2009) The silencing domain of GW182 interacts with PABPC1 to promote translational repression and degradation of microRNA targets and is required for target release. *Mol Cell Biol*, **29**, 6220-6231.
118. Kinch, L.N. and Grishin, N.V. (2009) The human Ago2 MC region does not contain an eIF4E-like mRNA cap binding motif. *Biol Direct*, **4**, 2.

119. Fukaya, T. and Tomari, Y. (2012) MicroRNAs mediate gene silencing via multiple different pathways in drosophila. *Mol Cell*, **48**, 825-836.
120. Nottrott, S., Simard, M.J. and Richter, J.D. (2006) Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat Struct Mol Biol*, **13**, 1108-1114.
121. Kim, J., Krichevsky, A., Grad, Y., Hayes, G.D., Kosik, K.S., Church, G.M. and Ruvkun, G. (2004) Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. *Proc Natl Acad Sci U S A*, **101**, 360-365.
122. Nelson, P.T., Hatzigeorgiou, A.G. and Mourelatos, Z. (2004) miRNP:mRNA association in polyribosomes in a human neuronal cell line. *RNA*, **10**, 387-394.
123. Olsen, P.H. and Ambros, V. (1999) The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol*, **216**, 671-680.
124. Fabian, M.R., Sonenberg, N. and Filipowicz, W. (2010) Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem*, **79**, 351-379.
125. Coller, J. and Parker, R. (2004) Eukaryotic mRNA decapping. *Annu Rev Biochem*, **73**, 861-890.
126. Djuranovic, S., Nahvi, A. and Green, R. (2012) miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science*, **336**, 237-240.
127. Bazzini, A.A., Lee, M.T. and Giraldez, A.J. (2012) Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science*, **336**, 233-237.
128. Henke, J.I., Goergen, D., Zheng, J., Song, Y., Schuttler, C.G., Fehr, C., Junemann, C. and Niepmann, M. (2008) microRNA-122 stimulates translation of hepatitis C virus RNA. *EMBO J*, **27**, 3300-3310.
129. Vasudevan, S., Tong, Y. and Steitz, J.A. (2007) Switching from repression to activation: microRNAs can up-regulate translation. *Science*, **318**, 1931-1934.
130. Orom, U.A., Nielsen, F.C. and Lund, A.H. (2008) MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell*, **30**, 460-471.

131. Vasudevan, S. and Steitz, J.A. (2007) AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell*, **128**, 1105-1118.
132. Flynt, A.S., Greimann, J.C., Chung, W.J., Lima, C.D. and Lai, E.C. (2010) MicroRNA biogenesis via splicing and exosome-mediated trimming in *Drosophila*. *Mol Cell*, **38**, 900-907.
133. Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P. and Blelloch, R. (2008) Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev*, **22**, 2773-2785.
134. Bogerd, H.P., Karnowski, H.W., Cai, X., Shin, J., Pohlers, M. and Cullen, B.R. (2010) A mammalian herpesvirus uses noncanonical expression and processing mechanisms to generate viral MicroRNAs. *Mol Cell*, **37**, 135-142.
135. Yang, J.S. and Lai, E.C. (2011) Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. *Mol Cell*, **43**, 892-903.
136. Ding, S.W. and Voinnet, O. (2007) Antiviral immunity directed by small RNAs. *Cell*, **130**, 413-426.
137. Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T. *et al.* (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, **453**, 539-543.
138. Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M. *et al.* (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature*, **453**, 534-538.
139. Axtell, M.J., Westholm, J.O. and Lai, E.C. (2011) Vive la difference: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol*, **12**, 221.
140. Merchan, F., Boualem, A., Crespi, M. and Frugier, F. (2009) Plant polycistronic precursors containing non-homologous microRNAs target transcripts encoding functionally related proteins. *Genome Biol*, **10**, R136.
141. Kurihara, Y. and Watanabe, Y. (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci U S A*, **101**, 12753-12758.

142. Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H. and Poethig, R.S. (2005) Nuclear processing and export of microRNAs in Arabidopsis. *Proc Natl Acad Sci U S A*, **102**, 3691-3696.
143. Fang, Y. and Spector, D.L. (2007) Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants. *Curr Biol*, **17**, 818-823.
144. Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R. and Chen, X. (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science*, **307**, 932-935.
145. Li, J., Yang, Z., Yu, B., Liu, J. and Chen, X. (2005) Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr Biol*, **15**, 1501-1507.
146. Baumberger, N. and Baulcombe, D.C. (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci U S A*, **102**, 11928-11933.
147. Vaucheret, H. (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev*, **20**, 759-771.
148. Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B. and Bartel, D.P. (2002) Prediction of plant microRNA targets. *Cell*, **110**, 513-520.
149. Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y.Y., Sieburth, L. and Voinnet, O. (2008) Widespread translational inhibition by plant miRNAs and siRNAs. *Science*, **320**, 1185-1190.
150. Melcher, T., Maas, S., Herb, A., Sprengel, R., Seeburg, P.H. and Higuchi, M. (1996) A mammalian RNA editing enzyme. *Nature*, **379**, 460-464.
151. Bass, B.L. (2002) RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem*, **71**, 817-846.
152. Keegan, L.P., Gallo, A. and O'Connell, M.A. (2001) The many roles of an RNA editor. *Nat Rev Genet*, **2**, 869-878.
153. Melcher, T., Maas, S., Herb, A., Sprengel, R., Higuchi, M. and Seeburg, P.H. (1996) RED2, a brain-specific member of the RNA-specific adenosine deaminase family. *J Biol Chem*, **271**, 31795-31798.

154. Chen, C.X., Cho, D.S., Wang, Q., Lai, F., Carter, K.C. and Nishikura, K. (2000) A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA*, **6**, 755-767.
155. Yang, W., Chendrimada, T.P., Wang, Q., Higuchi, M., Seeburg, P.H., Shiekhattar, R. and Nishikura, K. (2006) Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol*, **13**, 13-21.
156. Ohman, M. (2007) A-to-I editing challenger or ally to the microRNA process. *Biochimie*, **89**, 1171-1176.
157. Kawahara, Y., Zinshteyn, B., Chendrimada, T.P., Shiekhattar, R. and Nishikura, K. (2007) RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. *EMBO Rep*, **8**, 763-769.
158. Scadden, A.D. (2005) The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nat Struct Mol Biol*, **12**, 489-496.
159. Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A.G. and Nishikura, K. (2007) Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science*, **315**, 1137-1140.
160. Blow, M.J., Grocock, R.J., van Dongen, S., Enright, A.J., Dicks, E., Futreal, P.A., Wooster, R. and Stratton, M.R. (2006) RNA editing of human microRNAs. *Genome Biol*, **7**, R27.
161. Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S. and Johnson, J.M. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, **433**, 769-773.
162. Lehmann, K.A. and Bass, B.L. (1999) The importance of internal loops within RNA substrates of ADAR1. *J Mol Biol*, **291**, 1-13.
163. Polson, A.G. and Bass, B.L. (1994) Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J*, **13**, 5701-5711.
164. Lehmann, K.A. and Bass, B.L. (2000) Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry*, **39**, 12875-12884.
165. Kuttan, A. and Bass, B.L. (2012) Mechanistic insights into editing-site specificity of ADARs. *Proc Natl Acad Sci U S A*, **109**, E3295-3304.

166. Heo, I., Ha, M., Lim, J., Yoon, M.J., Park, J.E., Kwon, S.C., Chang, H. and Kim, V.N. (2012) Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. *Cell*, **151**, 521-532.
167. Thornton, J.E. and Gregory, R.I. (2012) How does Lin28 let-7 control development and disease? *Trends Cell Biol*, **22**, 474-482.
168. Yeom, K.H., Heo, I., Lee, J., Hohng, S., Kim, V.N. and Joo, C. (2011) Single-molecule approach to immunoprecipitated protein complexes: insights into miRNA uridylation. *EMBO Rep*, **12**, 690-696.
169. Krainer, A.R., Conway, G.C. and Kozak, D. (1990) Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev*, **4**, 1158-1171.
170. Ge, H. and Manley, J.L. (1990) A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell*, **62**, 25-34.
171. Sun, S., Zhang, Z., Sinha, R., Karni, R. and Krainer, A.R. (2010) SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. *Nat Struct Mol Biol*, **17**, 306-312.
172. Wu, H., Sun, S., Tu, K., Gao, Y., Xie, B., Krainer, A.R. and Zhu, J. (2010) A splicing-independent function of SF2/ASF in microRNA processing. *Mol Cell*, **38**, 67-77.
173. Fang, Y., Xue, J.L., Shen, Q., Chen, J. and Tian, L. (2012) MicroRNA-7 inhibits tumor growth and metastasis by targeting the phosphoinositide 3-kinase/Akt pathway in hepatocellular carcinoma. *Hepatology*, **55**, 1852-1862.
174. Min, H., Turck, C.W., Nikolic, J.M. and Black, D.L. (1997) A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. *Genes Dev*, **11**, 1023-1036.
175. Kroll, T.T., Zhao, W.M., Jiang, C. and Huber, P.W. (2002) A homolog of FBP2/KSRP binds to localized mRNAs in *Xenopus* oocytes. *Development*, **129**, 5609-5619.
176. Gherzi, R., Chen, C.Y., Trabucchi, M., Ramos, A. and Briata, P. (2010) The role of KSRP in mRNA decay and microRNA precursor maturation. *Wiley Interdiscip Rev RNA*, **1**, 230-239.

177. Nicastro, G., Garcia-Mayoral, M.F., Hollingworth, D., Kelly, G., Martin, S.R., Briata, P., Gherzi, R. and Ramos, A. (2012) Noncanonical G recognition mediates KSRP regulation of let-7 biogenesis. *Nat Struct Mol Biol*, **19**, 1282-1286.
178. 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.
179. Mayeda, A., Munroe, S.H., Caceres, J.F. and Krainer, A.R. (1994) Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J*, **13**, 5483-5495.
180. Mayeda, A. and Krainer, A.R. (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell*, **68**, 365-375.
181. Caceres, J.F., Stamm, S., Helfman, D.M. and Krainer, A.R. (1994) Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science*, **265**, 1706-1709.
182. Krecic, A.M. and Swanson, M.S. (1999) hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol*, **11**, 363-371.
183. Pinol-Roma, S. and Dreyfuss, G. (1992) Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature*, **355**, 730-732.
184. Dreyfuss, G., Matunis, M.J., Pinol-Roma, S. and Burd, C.G. (1993) hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem*, **62**, 289-321.
185. Bonnal, S., Pileur, F., Orsini, C., Parker, F., Pujol, F., Prats, A.C. and Vagner, S. (2005) Heterogeneous nuclear ribonucleoprotein A1 is a novel internal ribosome entry site trans-acting factor that modulates alternative initiation of translation of the fibroblast growth factor 2 mRNA. *J Biol Chem*, **280**, 4144-4153.
186. Hamilton, B.J., Burns, C.M., Nichols, R.C. and Rigby, W.F. (1997) Modulation of AUUUA response element binding by heterogeneous nuclear ribonucleoprotein A1 in human T lymphocytes. The roles of cytoplasmic location, transcription, and phosphorylation. *J Biol Chem*, **272**, 28732-28741.
187. 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.

188. Michlewski, G., Guil, S., Semple, C.A. and Caceres, J.F. (2008) Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol Cell*, **32**, 383-393.
189. Guil, S. and Caceres, J.F. (2007) The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat Struct Mol Biol*, **14**, 591-596.
190. Xu, R.M., Jokhan, L., Cheng, X., Mayeda, A. and Krainer, A.R. (1997) Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure*, **5**, 559-570.
191. Shamoo, Y., Krueger, U., Rice, L.M., Williams, K.R. and Steitz, T.A. (1997) Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nat Struct Biol*, **4**, 215-222.
192. Ding, J., Hayashi, M.K., Zhang, Y., Manche, L., Krainer, A.R. and Xu, R.M. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev*, **13**, 1102-1115.
193. Herrick, G. and Alberts, B. (1976) Purification and physical characterization of nucleic acid helix-unwinding proteins from calf thymus. *J Biol Chem*, **251**, 2124-2132.
194. Pandolfo, M., Valentini, O., Biamonti, G., Morandi, C. and Riva, S. (1985) Single stranded DNA binding proteins derive from hnRNP proteins by proteolysis in mammalian cells. *Nucleic Acids Res*, **13**, 6577-6590.
195. Okunola, H.L. and Krainer, A.R. (2009) Cooperative-binding and splicing-repressive properties of hnRNP A1. *Mol Cell Biol*, **29**, 5620-5631.
196. Nielsen, A.F., Leuschner, P.J. and Martinez, J. (2007) Not miR-ly a splicing factor: hnRNP A1 succumbs to microRNA temptation. *Nat Struct Mol Biol*, **14**, 572-573.
197. Hanamura, A., Caceres, J.F., Mayeda, A., Franza, B.R., Jr. and Krainer, A.R. (1998) Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA*, **4**, 430-444.
198. Golan-Gerstl, R., Cohen, M., Shilo, A., Suh, S.S., Bakacs, A., Coppola, L. and Karni, R. (2011) Splicing factor hnRNP A2/B1 regulates tumor suppressor gene splicing and is an oncogenic driver in glioblastoma. *Cancer Res*, **71**, 4464-4472.
199. 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.

200. 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.
201. 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.
202. 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.
203. Thornton, J.E., Chang, H.M., Piskounova, E. and Gregory, R.I. (2012) Lin28-mediated control of let-7 microRNA expression by alternative TUTases Zcchc11 (TUT4) and Zcchc6 (TUT7). *RNA*, **18**, 1875-1885.
204. Chang, H.M., Triboulet, R., Thornton, J.E. and Gregory, R.I. (2013) A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. *Nature*, **497**, 244-248.
205. Bushati, N. and Cohen, S.M. (2007) microRNA functions. *Annu Rev Cell Dev Biol*, **23**, 175-205.
206. Bussing, I., Slack, F.J. and Grosshans, H. (2008) let-7 microRNAs in development, stem cells and cancer. *Trends Mol Med*, **14**, 400-409.
207. Ambros, V. (2011) MicroRNAs and developmental timing. *Curr Opin Genet Dev*, **21**, 511-517.
208. Esquela-Kerscher, A., Johnson, S.M., Bai, L., Saito, K., Partridge, J., Reinert, K.L. and Slack, F.J. (2005) Post-embryonic expression of *C. elegans* microRNAs belonging to the lin-4 and let-7 families in the hypodermis and the reproductive system. *Dev Dyn*, **234**, 868-877.
209. Zhang, H. and Fire, A.Z. (2010) Cell autonomous specification of temporal identity by *Caenorhabditis elegans* microRNA lin-4. *Dev Biol*, **344**, 603-610.
210. Hammell, C.M., Lubin, I., Boag, P.R., Blackwell, T.K. and Ambros, V. (2009) nhl-2 Modulates microRNA activity in *Caenorhabditis elegans*. *Cell*, **136**, 926-938.

211. Hammell, C.M., Karp, X. and Ambros, V. (2009) A feedback circuit involving let-7-family miRNAs and DAF-12 integrates environmental signals and developmental timing in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A*, **106**, 18668-18673.
212. Schulman, B.R., Esquela-Kerscher, A. and Slack, F.J. (2005) Reciprocal expression of lin-41 and the microRNAs let-7 and mir-125 during mouse embryogenesis. *Dev Dyn*, **234**, 1046-1054.
213. Ooi, A.G., Sahoo, D., Adorno, M., Wang, Y., Weissman, I.L. and Park, C.Y. (2010) MicroRNA-125b expands hematopoietic stem cells and enriches for the lymphoid-balanced and lymphoid-biased subsets. *Proc Natl Acad Sci U S A*, **107**, 21505-21510.
214. Willmann, M.R., Mehalick, A.J., Packer, R.L. and Jenik, P.D. (2011) MicroRNAs regulate the timing of embryo maturation in *Arabidopsis*. *Plant Physiol*, **155**, 1871-1884.
215. Lee, E.J., Baek, M., Gusev, Y., Brackett, D.J., Nuovo, G.J. and Schmittgen, T.D. (2008) Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors. *RNA*, **14**, 35-42.
216. Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834-838.
217. Kumar, M.S., Lu, J., Mercer, K.L., Golub, T.R. and Jacks, T. (2007) Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat Genet*, **39**, 673-677.
218. Muralidhar, B., Goldstein, L.D., Ng, G., Winder, D.M., Palmer, R.D., Gooding, E.L., Barbosa-Morais, N.L., Mukherjee, G., Thorne, N.P., Roberts, I. *et al.* (2007) Global microRNA profiles in cervical squamous cell carcinoma depend on Drosha expression levels. *J Pathol*, **212**, 368-377.
219. Faber, C., Horst, D., Hlubek, F. and Kirchner, T. (2011) Overexpression of Dicer predicts poor survival in colorectal cancer. *Eur J Cancer*, **47**, 1414-1419.
220. Chiosea, S., Jelezcova, E., Chandran, U., Luo, J., Mantha, G., Sobol, R.W. and Dacic, S. (2007) Overexpression of Dicer in precursor lesions of lung adenocarcinoma. *Cancer Res*, **67**, 2345-2350.

221. Karube, Y., Tanaka, H., Osada, H., Tomida, S., Tatematsu, Y., Yanagisawa, K., Yatabe, Y., Takamizawa, J., Miyoshi, S., Mitsudomi, T. *et al.* (2005) Reduced expression of Dicer associated with poor prognosis in lung cancer patients. *Cancer Sci*, **96**, 111-115.
222. Melo, S.A., Ropero, S., Moutinho, C., Aaltonen, L.A., Yamamoto, H., Calin, G.A., Rossi, S., Fernandez, A.F., Carneiro, F., Oliveira, C. *et al.* (2009) A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. *Nat Genet*, **41**, 365-370.
223. Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K. *et al.* (2002) Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*, **99**, 15524-15529.
224. Cimmino, A., Calin, G.A., Fabbri, M., Iorio, M.V., Ferracin, M., Shimizu, M., Wojcik, S.E., Aqeilan, R.I., Zupo, S., Dono, M. *et al.* (2005) miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A*, **102**, 13944-13949.
225. He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J. *et al.* (2005) A microRNA polycistron as a potential human oncogene. *Nature*, **435**, 828-833.
226. O'Donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V. and Mendell, J.T. (2005) c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*, **435**, 839-843.
227. Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M. *et al.* (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res*, **65**, 7065-7070.
228. Ciafre, S.A., Galardi, S., Mangiola, A., Ferracin, M., Liu, C.G., Sabatino, G., Negrini, M., Maira, G., Croce, C.M. and Farace, M.G. (2005) Extensive modulation of a set of microRNAs in primary glioblastoma. *Biochem Biophys Res Commun*, **334**, 1351-1358.
229. Bloomston, M., Frankel, W.L., Petrocca, F., Volinia, S., Alder, H., Hagan, J.P., Liu, C.G., Bhatt, D., Taccioli, C. and Croce, C.M. (2007) MicroRNA expression patterns to differentiate pancreatic adenocarcinoma from normal pancreas and chronic pancreatitis. *JAMA*, **297**, 1901-1908.

230. Frankel, L.B., Christoffersen, N.R., Jacobsen, A., Lindow, M., Krogh, A. and Lund, A.H. (2008) Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. *J Biol Chem*, **283**, 1026-1033.
231. Ma, L., Reinhardt, F., Pan, E., Soutschek, J., Bhat, B., Marcusson, E.G., Teruya-Feldstein, J., Bell, G.W. and Weinberg, R.A. (2010) Therapeutic silencing of miR-10b inhibits metastasis in a mouse mammary tumor model. *Nat Biotechnol*, **28**, 341-347.
232. Asangani, I.A., Rasheed, S.A., Nikolova, D.A., Leupold, J.H., Colburn, N.H., Post, S. and Allgayer, H. (2008) MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene*, **27**, 2128-2136.
233. Voorhoeve, P.M., le Sage, C., Schrier, M., Gillis, A.J., Stoop, H., Nagel, R., Liu, Y.P., van Duijse, J., Drost, J., Griekspoor, A. *et al.* (2006) A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell*, **124**, 1169-1181.
234. Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D. and Slack, F.J. (2005) RAS is regulated by the let-7 microRNA family. *Cell*, **120**, 635-647.
235. Lee, Y.S. and Dutta, A. (2007) The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev*, **21**, 1025-1030.
236. Osada, H. and Takahashi, T. (2011) let-7 and miR-17-92: small-sized major players in lung cancer development. *Cancer Sci*, **102**, 9-17.
237. Chin, L.J., Ratner, E., Leng, S., Zhai, R., Nallur, S., Babar, I., Muller, R.U., Straka, E., Su, L., Burki, E.A. *et al.* (2008) A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk. *Cancer Res*, **68**, 8535-8540.
238. He, L., He, X., Lim, L.P., de Stanchina, E., Xuan, Z., Liang, Y., Xue, W., Zender, L., Magnus, J., Ridzon, D. *et al.* (2007) A microRNA component of the p53 tumour suppressor network. *Nature*, **447**, 1130-1134.

239. Mott, J.L., Kurita, S., Cazanave, S.C., Bronk, S.F., Werneburg, N.W. and Fernandez-Zapico, M.E. (2010) Transcriptional suppression of mir-29b-1/mir-29a promoter by c-Myc, hedgehog, and NF-kappaB. *J Cell Biochem*, **110**, 1155-1164.
240. Martello, G., Rosato, A., Ferrari, F., Manfrin, A., Cordenonsi, M., Dupont, S., Enzo, E., Guzzardo, V., Rondina, M., Spruce, T. *et al.* (2010) A MicroRNA targeting dicer for metastasis control. *Cell*, **141**, 1195-1207.
241. Roush, S. and Slack, F.J. (2008) The let-7 family of microRNAs. *Trends Cell Biol*, **18**, 505-516.
242. Johnson, C.D., Esquela-Kerscher, A., Stefani, G., Byrom, M., Kelnar, K., Ovcharenko, D., Wilson, M., Wang, X., Shelton, J., Shingara, J. *et al.* (2007) The let-7 microRNA represses cell proliferation pathways in human cells. *Cancer Res*, **67**, 7713-7722.
243. Mondol, V. and Pasquinelli, A.E. (2012) Let's make it happen: the role of let-7 microRNA in development. *Curr Top Dev Biol*, **99**, 1-30.
244. 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.
245. Viswanathan, S.R., Daley, G.Q. and Gregory, R.I. (2008) Selective blockade of microRNA processing by Lin28. *Science*, **320**, 97-100.
246. Suh, M.R., Lee, Y., Kim, J.Y., Kim, S.K., Moon, S.H., Lee, J.Y., Cha, K.Y., Chung, H.M., Yoon, H.S., Moon, S.Y. *et al.* (2004) Human embryonic stem cells express a unique set of microRNAs. *Dev Biol*, **270**, 488-498.
247. 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.
248. 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.
249. 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.
250. 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.
251. Griffiths-Jones, S. (2004) The microRNA Registry. *Nucleic Acids Res*, **32**, D109-111.
252. 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.
253. Sokol, N.S., Xu, P., Jan, Y.N. and Ambros, V. (2008) Drosophila let-7 microRNA is required for remodeling of the neuromusculature during metamorphosis. *Genes Dev*, **22**, 1591-1596.
254. Caygill, E.E. and Johnston, L.A. (2008) Temporal regulation of metamorphic processes in Drosophila by the let-7 and miR-125 heterochronic microRNAs. *Curr Biol*, **18**, 943-950.
255. Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J. *et al.* (2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell*, **134**, 521-533.
256. Houbaviy, H.B., Murray, M.F. and Sharp, P.A. (2003) Embryonic stem cell-specific MicroRNAs. *Dev Cell*, **5**, 351-358.
257. Li, M.A. and He, L. (2012) microRNAs as novel regulators of stem cell pluripotency and somatic cell reprogramming. *Bioessays*, **34**, 670-680.
258. Wang, Y., Baskerville, S., Shenoy, A., Babiarz, J.E., Baehner, L. and Blelloch, R. (2008) Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat Genet*, **40**, 1478-1483.
259. Lichner, Z., Pall, E., Kerekes, A., Pallinger, E., Maraghechi, P., Bosze, Z. and Gocza, E. (2011) The miR-290-295 cluster promotes pluripotency maintenance by regulating cell cycle phase distribution in mouse embryonic stem cells. *Differentiation*, **81**, 11-24.
260. Melton, C., Judson, R.L. and Blelloch, R. (2010) Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature*, **463**, 621-626.

261. Schultz, J., Lorenz, P., Gross, G., Ibrahim, S. and Kunz, M. (2008) MicroRNA let-7b targets important cell cycle molecules in malignant melanoma cells and interferes with anchorage-independent growth. *Cell Res*, **18**, 549-557.
262. Zhong, X., Li, N., Liang, S., Huang, Q., Coukos, G. and Zhang, L. (2010) Identification of microRNAs regulating reprogramming factor LIN28 in embryonic stem cells and cancer cells. *J Biol Chem*, **285**, 41961-41971.
263. Johnson, L., Mercer, K., Greenbaum, D., Bronson, R.T., Crowley, D., Tuveson, D.A. and Jacks, T. (2001) Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. *Nature*, **410**, 1111-1116.
264. Ahrendt, S.A., Decker, P.A., Alawi, E.A., Zhu Yr, Y.R., Sanchez-Cespedes, M., Yang, S.C., Haasler, G.B., Kajdacsy-Balla, A., Demeure, M.J. and Sidransky, D. (2001) Cigarette smoking is strongly associated with mutation of the K-ras gene in patients with primary adenocarcinoma of the lung. *Cancer*, **92**, 1525-1530.
265. Inamura, K., Togashi, Y., Nomura, K., Ninomiya, H., Hiramatsu, M., Satoh, Y., Okumura, S., Nakagawa, K. and Ishikawa, Y. (2007) let-7 microRNA expression is reduced in bronchioloalveolar carcinoma, a non-invasive carcinoma, and is not correlated with prognosis. *Lung Cancer*, **58**, 392-396.
266. Chang, T.C., Yu, D., Lee, Y.S., Wentzel, E.A., Arking, D.E., West, K.M., Dang, C.V., Thomas-Tikhonenko, A. and Mendell, J.T. (2008) Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat Genet*, **40**, 43-50.
267. Boyerinas, B., Park, S.M., Shomron, N., Hedegaard, M.M., Vinther, J., Andersen, J.S., Feig, C., Xu, J., Burge, C.B. and Peter, M.E. (2008) Identification of let-7-regulated oncofetal genes. *Cancer Res*, **68**, 2587-2591.
268. Peter, M.E. (2009) Let-7 and miR-200 microRNAs: guardians against pluripotency and cancer progression. *Cell Cycle*, **8**, 843-852.
269. Mayr, C., Hemann, M.T. and Bartel, D.P. (2007) Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science*, **315**, 1576-1579.
270. Shell, S., Park, S.M., Radjabi, A.R., Schickel, R., Kistner, E.O., Jewell, D.A., Feig, C., Lengyel, E. and Peter, M.E. (2007) Let-7 expression defines two differentiation stages of cancer. *Proc Natl Acad Sci U S A*, **104**, 11400-11405.

271. Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., Huang, Y., Hu, X., Su, F., Lieberman, J. *et al.* (2007) let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell*, **131**, 1109-1123.
272. Yu, S.L., Chen, H.Y., Chang, G.C., Chen, C.Y., Chen, H.W., Singh, S., Cheng, C.L., Yu, C.J., Lee, Y.C., Chen, H.S. *et al.* (2008) MicroRNA signature predicts survival and relapse in lung cancer. *Cancer Cell*, **13**, 48-57.
273. Lobo, N.A., Shimono, Y., Qian, D. and Clarke, M.F. (2007) The biology of cancer stem cells. *Annu Rev Cell Dev Biol*, **23**, 675-699.
274. Dalerba, P., Cho, R.W. and Clarke, M.F. (2007) Cancer stem cells: models and concepts. *Annu Rev Med*, **58**, 267-284.
275. Wu, L. and Belasco, J.G. (2005) Micro-RNA regulation of the mammalian lin-28 gene during neuronal differentiation of embryonal carcinoma cells. *Mol Cell Biol*, **25**, 9198-9208.
276. Moss, E.G. and Tang, L. (2003) Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. *Dev Biol*, **258**, 432-442.
277. 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.
278. 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.
279. Lettre, G., Jackson, A.U., Gieger, C., Schumacher, F.R., Berndt, S.I., Sanna, S., Eyheramendy, S., Voight, B.F., Butler, J.L., Guducci, C. *et al.* (2008) Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet*, **40**, 584-591.
280. Hartge, P. (2009) Genetics of reproductive lifespan. *Nat Genet*, **41**, 637-638.
281. Zhu, H., Shah, S., Shyh-Chang, N., Shinoda, G., Einhorn, W.S., Viswanathan, S.R., Takeuchi, A., Grasmann, 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.

282. Faas, L., Warrander, F.C., Maguire, R., Ramsbottom, S.A., Quinn, D., Genever, P. and Isaacs, H.V. (2013) Lin28 proteins are required for germ layer specification in *Xenopus*. *Development*, **140**, 976-986.
283. 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.
284. King, C.E., Cuatrecasas, M., Castells, A., Sepulveda, A.R., Lee, J.S. and Rustgi, A.K. (2011) LIN28B promotes colon cancer progression and metastasis. *Cancer Res*, **71**, 4260-4268.
285. Chen, A.X., Yu, K.D., Fan, L., Li, J.Y., Yang, C., Huang, A.J. and Shao, Z.M. (2011) Germline genetic variants disturbing the Let-7/LIN28 double-negative feedback loop alter breast cancer susceptibility. *PLoS Genet*, **7**, e1002259.
286. Permut-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.
287. Viswanathan, S.R. and Daley, G.Q. (2010) Lin28: A microRNA regulator with a macro role. *Cell*, **140**, 445-449.
288. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, **7**, 539.
289. Kingsley, P.D. and Palis, J. (1994) GRP2 proteins contain both CCHC zinc fingers and a cold shock domain. *Plant Cell*, **6**, 1522-1523.
290. Kim, J.S., Park, S.J., Kwak, K.J., Kim, Y.O., Kim, J.Y., Song, J., Jang, B., Jung, C.H. and Kang, H. (2007) Cold shock domain proteins and glycine-rich RNA-binding proteins from *Arabidopsis thaliana* can promote the cold adaptation process in *Escherichia coli*. *Nucleic Acids Res*, **35**, 506-516.

291. 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.
292. 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.
293. Piskounova, E., Polytarchou, C., Thornton, J.E., LaPierre, R.J., Pothoulakis, C., Hagan, J.P., Iliopoulos, D. and Gregory, R.I. (2011) Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. *Cell*, **147**, 1066-1079.
294. Wilhelm, J.E., Mansfield, J., Hom-Booher, N., Wang, S., Turck, C.W., Hazelrigg, T. and Vale, R.D. (2000) Isolation of a ribonucleoprotein complex involved in mRNA localization in *Drosophila* oocytes. *J Cell Biol*, **148**, 427-440.
295. Kohno, K., Izumi, H., Uchiumi, T., Ashizuka, M. and Kuwano, M. (2003) The pleiotropic functions of the Y-box-binding protein, YB-1. *Bioessays*, **25**, 691-698.
296. Ermolenko, D.N. and Makhatadze, G.I. (2002) Bacterial cold-shock proteins. *Cell Mol Life Sci*, **59**, 1902-1913.
297. 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.
298. Landsman, D. (1992) RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain. *Nucleic Acids Res*, **20**, 2861-2864.
299. 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.
300. 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.
301. 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.

302. 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 x 3 internal loop. *Biochemistry*, **42**, 5259-5269.
303. 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.
304. Desjardins, A., Yang, A., Bouvette, J., Omichinski, J.G. and Legault, P. (2012) Importance of the NCp7-like domain in the recognition of pre-let-7g by the pluripotency factor Lin28. *Nucleic Acids Res*, **40**, 1767-1777.
305. Liu, J., Rivas, F.V., Wohlschlegel, J., Yates, J.R., 3rd, Parker, R. and Hannon, G.J. (2005) A role for the P-body component GW182 in microRNA function. *Nat Cell Biol*, **7**, 1261-1266.
306. Liu, J., Valencia-Sanchez, M.A., Hannon, G.J. and Parker, R. (2005) MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat Cell Biol*, **7**, 719-723.
307. Poleskaya, 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.
308. Poleskaya, A. and Harel-Bellan, A. (2007) [A novel role for an embryonic regulatory protein Lin-28 in adult skeletal muscle differentiation]. *Med Sci (Paris)*, **23**, 796-797.
309. Huang, Y. (2012) A mirror of two faces: Lin28 as a master regulator of both miRNA and mRNA. *Wiley Interdiscip Rev RNA*, **3**, 483-494.
310. 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.
311. 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.

312. 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.
313. Wilbert, M.L., Huelga, S.C., Kapeli, K., Stark, T.J., Liang, T.Y., Chen, S.X., Yan, B.Y., Nathanson, J.L., Hutt, K.R., Lovci, M.T. *et al.* (2012) LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. *Mol Cell*, **48**, 195-206.
314. Hafner, M., Max, K.E., Bandaru, P., Morozov, P., Gerstberger, S., Brown, M., Molina, H. and Tuschl, T. (2013) Identification of mRNAs bound and regulated by human LIN28 proteins and molecular requirements for RNA recognition. *RNA*, **19**, 613-626.
315. 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.
316. Li, N., Zhong, X., Lin, X., Guo, J., Zou, L., Tanyi, J.L., Shao, Z., Liang, S., Wang, L.P., Hwang, W.T. *et al.* (2012) Lin-28 homologue A (LIN28A) promotes cell cycle progression via regulation of cyclin-dependent kinase 2 (CDK2), cyclin D1 (CCND1), and cell division cycle 25 homolog A (CDC25A) expression in cancer. *J Biol Chem*, **287**, 17386-17397.
317. Bolinger, C., Sharma, A., Singh, D., Yu, L. and Boris-Lawrie, K. (2010) RNA helicase A modulates translation of HIV-1 and infectivity of progeny virions. *Nucleic Acids Res*, **38**, 1686-1696.
318. Hartman, T.R., Qian, S., Bolinger, C., Fernandez, S., Schoenberg, D.R. and Boris-Lawrie, K. (2006) RNA helicase A is necessary for translation of selected messenger RNAs. *Nat Struct Mol Biol*, **13**, 509-516.
319. 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.

320. Cho, J., Chang, H., Kwon, S.C., Kim, B., Kim, Y., Choe, J., Ha, M., Kim, Y.K. and Kim, V.N. (2012) LIN28A is a suppressor of ER-associated translation in embryonic stem cells. *Cell*, **151**, 765-777.
321. Lei, X.X., Xu, J., Ma, W., Qiao, C., Newman, M.A., Hammond, S.M. and Huang, Y. (2012) Determinants of mRNA recognition and translation regulation by Lin28. *Nucleic Acids Res*, **40**, 3574-3584.
322. Graf, R., Munschauer, M., Mastrobuoni, G., Mayr, F., Heinemann, U., Kempa, S., Rajewsky, N. and Landthaler, M. (2013) Identification of LIN28B-bound mRNAs reveals features of target recognition and regulation. *RNA Biol*, **10**, 1146-1159.
323. Ali, P.S., Ghoshdastider, U., Hoffmann, J., Brutschy, B. and Filippek, S. (2012) Recognition of the let-7g miRNA precursor by human Lin28B. *FEBS Lett*, **586**, 3986-3990.
324. 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.
325. Nam, Y., Chen, C., Gregory, R.I., Chou, J.J. and Sliz, P. (2011) Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell*, **147**, 1080-1091.
326. Loughlin, F.E., Gebert, L.F., Towbin, H., Brunschweiler, A., Hall, J. and Allain, F.H. (2012) Structural basis of pre-let-7 miRNA recognition by the zinc knuckles of pluripotency factor Lin28. *Nat Struct Mol Biol*, **19**, 84-89.
327. Mayr, F., Schutz, A., Doge, N. and Heinemann, U. (2012) The Lin28 cold-shock domain remodels pre-let-7 microRNA. *Nucleic Acids Res*, **40**, 7492-7506.
328. Di Tomasso, G., Dagenais, P., Desjardins, A., Rompre-Brodeur, A., Delfosse, V. and Legault, P. (2012) Affinity purification of RNA using an ARiBo tag. *Methods Mol Biol*, **941**, 137-155.
329. 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.
330. Goodrich, J.A. and Kugel, J.F. (2007) *Binding and kinetics for molecular biologists*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

331. Rachofsky, E.L., Osman, R. and Ross, J.B. (2001) Probing structure and dynamics of DNA with 2-aminopurine: effects of local environment on fluorescence. *Biochemistry*, **40**, 946-956.
332. Heppell, B., Mulhbachter, J., Penedo, J.C. and Lafontaine, D.A. (2009) Application of fluorescent measurements for characterization of riboswitch-ligand interactions. *Methods Mol Biol*, **540**, 25-37.
333. Rist, M.J. and Marino, J.P. (2002) Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry*, **41**, 14762-14770.
334. Windbichler, N., Werner, M. and Schroeder, R. (2003) Kissing complex-mediated dimerisation of HIV-1 RNA: coupling extended duplex formation to ribozyme cleavage. *Nucleic Acids Res*, **31**, 6419-6427.
335. Phadtare, S., Inouye, M. and Severinov, K. (2004) The mechanism of nucleic acid melting by a CspA family protein. *J Mol Biol*, **337**, 147-155.
336. Rajkowitsch, L. and Schroeder, R. (2007) Dissecting RNA chaperone activity. *RNA*, **13**, 2053-2060.
337. Rajkowitsch, L. and Schroeder, R. (2007) Coupling RNA annealing and strand displacement: a FRET-based microplate reader assay for RNA chaperone activity. *Biotechniques*, **43**, 304-310.
338. Dragan, A.I. and Privalov, P.L. (2008) Use of Fluorescence Resonance Energy Transfer (FRET) in Studying Protein-Induced DNA Bending. *Method Enzymol*, **450**, 185-199.
339. 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.
340. Forsen, S. and Linse, S. (1995) Cooperativity: over the Hill. *Trends Biochem Sci*, **20**, 495-497.
341. Hill, A.V. (1910) The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *The Journal of Physiology*, **40**, iv-vii.

342. Perutz, M.F., Wilkinson, A.J., Paoli, M. and Dodson, G.G. (1998) The stereochemical mechanism of the cooperative effects in hemoglobin revisited. *Annu Rev Biophys Biomol Struct*, **27**, 1-34.
343. Williamson, J.R. (2008) Cooperativity in macromolecular assembly. *Nat Chem Biol*, **4**, 458-465.
344. Sackett, D.L. and Saroff, H.A. (1996) The multiple origins of cooperativity in binding to multi-site lattices. *FEBS Lett*, **397**, 1-6.
345. Linse, S. and Chazin, W.J. (1995) Quantitative measurements of the cooperativity in an EF-hand protein with sequential calcium binding. *Protein Sci*, **4**, 1038-1044.
346. Kranz, J.K. and Hall, K.B. (1998) RNA binding mediates the local cooperativity between the beta-sheet and the C-terminal tail of the human U1A RBD1 protein. *J Mol Biol*, **275**, 465-481.
347. Recht, M.I. and Williamson, J.R. (2004) RNA tertiary structure and cooperative assembly of a large ribonucleoprotein complex. *J Mol Biol*, **344**, 395-407.
348. Recht, M.I. and Williamson, J.R. (2001) Central domain assembly: thermodynamics and kinetics of S6 and S18 binding to an S15-RNA complex. *J Mol Biol*, **313**, 35-48.
349. Kallen, A.N., Ma, J. and Huang, Y. (2012) Does Lin28 antagonize miRNA-mediated repression by displacing miRISC from target mRNAs? *Front Genet*, **3**, 240.