

Université de Montréal

**Adaptations de la méthode de purification d'ARN par affinité avec
l'étiquette ARiBo**

Par

Alix-Salvail-Lacoste

Département de biochimie et médecine moléculaire
Faculté de médecine

Mémoire présenté à la faculté de médecine
en vue de l'obtention du grade de M.Sc.
en biochimie
option dynamique cellulaire des complexes macromoléculaires

31 août, 2014

© Alix Salvail-Lacoste

Université de Montréal
Faculté des études supérieures et postdoctorales

Ce mémoire intitulé :

**Adaptations de la méthode de purification d'ARN par affinité avec
l'étiquette ARiBo**

Présenté par :

Alix Salvail-Lacoste

A été évalué par un jury composé des personnes suivantes :

Luc Desgroseillers, président-rapporteur

Pascale Legault, directeur de recherche

James G. Omichinski, co-directeur de recherche

Eric Lécuyer, membre du jury

Résumé

Dans les dernières années, une explosion de la recherche sur les ARN a eu lieu à cause de nombreuses découvertes démontrant l'importance de l'ARN dans plusieurs processus biologiques. Ainsi, de grandes quantités d'ARN sont devenues indispensables au bon déroulement de plusieurs études, notamment pour la biologie structurale et la caractérisation fonctionnelle. Cependant, il existe encore peu de méthodes de purification simples, efficaces, fiables et produisant un ARN sous forme native. Dans les dernières années, le laboratoire Legault a mis au point une méthode de purification par affinité utilisant une étiquette ARiBo pour la purification d'ARN transcrits *in vitro* par la polymérase à ARN du phage T7. Cette méthode de purification d'ARN a été spécifiquement développée pour maximiser la pureté et le rendement. De plus, elle est très rapide et fonctionne avec plusieurs types d'ARN. Cependant, comme plusieurs autres méthodes de purification, cette méthode produit des ARN avec des extrémités 5' hétérogènes. Dans ce mémoire, des solutions sont proposées pour remédier au problème d'hétérogénéité en 5' des ARN transcrits avec la polymérase à ARN du phage T7 et purifiés par la méthode ARiBo. La première solution consiste à choisir la séquence en 5' parmi celles des 32 séquences testées qui ne présentent pas d'hétérogénéité en 5'. La seconde solution est d'utiliser une étiquette clivable en 5' de l'ARN d'intérêt, tel que le ribozyme hammerhead, déjà utilisée pour ce genre d'application, ou le système CRISPR/Cse3 que nous proposons dans l'article présenté dans ce mémoire. De plus, nous avons adapté la méthode ARiBo pour rendre possible la purification d'un long ARN de 614 nt, le polycistron miR-106b-25. Nous avons également démontré la possibilité d'utiliser la méthode ARiBo pour l'isolation de protéines qui se lient à un ARN donné, le précurseur de miRNA pre-miR-153-2. En conclusion, ce mémoire démontre la possibilité d'adapter la méthode ARiBo à plusieurs applications.

Mots-clés : Purification d'ARN par affinité, étiquette ARiBo, polymérase à ARN du phage T7, hétérogénéité en 5', CRISPR/Cse3, ribozyme hammerhead, complexe ribonucléoprotéique.

Abstract

In recent years, the field of RNA research has exploded due to several discoveries demonstrating the importance of RNA in many biological processes. Along with the increased interest in this field, large amounts of RNA have become essential to the success of several studies, in particular for structural biology and functional characterization. However, there are still very few native purification methods that are simple, efficient and reliable. In the past few years, the Legault laboratory has established an affinity purification method using an ARiBo tag to purify RNAs produced by *in vitro* transcription with the T7 RNA polymerase. This RNA purification method was specifically developed to maximise purity and yield. In addition, this method is fast and works with several types of RNAs. However, like several other purification methods, this method produces RNAs with 5' heterogeneity. This Master's thesis propose solutions to overcome the problem of 5' heterogeneity for RNAs transcribed with the T7 RNA polymerase and purified with the ARiBo method. The first solution proposed is to choose a 5' sequence among those of the 32 sequences tested that do not present 5'-heterogeneity. The other possibility is the use of a cleavable tag at the 5'-end of the RNA of interest, such as the hammerhead ribozyme, already used for this purpose or the CRISPR/Cse3 system, which is presented here. Furthermore, we have adapted the ARiBo method to purify an RNA of 614 nt, the miRNAs cluster miR-106b-25. We also demonstrate the possibility to use the ARiBo method to isolate proteins that bind a given RNA, the miRNA precursor pre-miR-153-2. In conclusion, this Master's thesis demonstrates the possibility of adapting the ARiBo method for several applications.

Keywords: RNA affinity purification, ARiBo tag, T7 RNA polymerase, 5' heterogeneity, CRISPR/Cse3, hammerhead ribozyme, ribonucleoprotein complex.

Table des matières

Résumé	iii
Abstract.....	iv
Table des matières	v
Liste des tableaux.....	viii
Liste des figures.....	ix
Liste des abréviations	xii
Remerciements	xiv
Chapitre 1 – Introduction	1
1.1 – Importance de l’ARN dans les processus biologiques.....	2
1.2 – Préparation d’ARN <i>in vitro</i>	2
1.2.1 – Synthèse enzymatique d’ARN.....	3
1.2.2 – Utilisation de ribozymes pour la préparation d’ARN.....	6
1.3 – Purification d’ARN.....	8
1.3.1 – Méthode de purification par gels dénaturants de polyacrylamide	8
1.3.2 – Méthodes de purification d’ARN par chromatographie	10
1.3.3 – Purification d’ARN par affinité.....	10
1.4 – Purification d’ARN par affinité utilisant l’étiquette ARiBo.....	13
1.4.1 – Description de la méthode	13
1.4.2 – Avantages et inconvénients de la méthode ARiBo	17

1.5 – Description de l’hétérogénéité de séquences en 5’ et 3’	18
1.6 – Perspectives pour réduire le problème d’hétérogénéité en 5’ avec la méthode ARiBo.....	19
<i>1.6.1 – Utilisation du ribozyme hammerhead pour l’élimination de l’hétérogénéité en 5’.....</i>	<i>19</i>
<i>1.6.2 – Utilisation potentielle du CRISPR comme étiquette à l’extrémité 5’ d’ARN.....</i>	<i>20</i>
1.6 – Hypothèses et Objectifs	22
Chapitre 2 – Élimination d’hétérogénéité d’ARN transcrits avec la polymérase à ARN du phage T7 et purifiés par affinité en utilisant les étiquettes ARiBo et CRISPR	25
2.1 – Mise en contexte de l’article	26
2.2 – Article.....	28
<i>Abstract</i>	<i>29</i>
<i>Introduction.....</i>	<i>29</i>
<i>Results</i>	<i>33</i>
<i>Discussion</i>	<i>40</i>
<i>Materials and methods.....</i>	<i>43</i>
<i>Acknowledgments.....</i>	<i>49</i>
<i>Tables.....</i>	<i>50</i>
<i>Figures</i>	<i>52</i>
<i>References</i>	<i>57</i>
<i>Supplemental material</i>	<i>63</i>
<i>References</i>	<i>69</i>

Chapitre 3 – Discussion	70
3.1 – Retour sur les résultats	71
3.1.1 – Sélection des séquences en 5'	71
3.1.2 – Utilisation du système CRIPSR/Cse3 pour assurer l’homogénéité en 5’	74
3.1.3 – Utilisation du ribozyme HH pour éliminer l’hétérogénéité en 5’	75
3.2 – Suite du projet	77
3.2.1 – Purification d’un long ARN, le polycistron miR-106b-25	78
3.2.2 – Isolation de complexes ribonucléoprotéiques (RNP).....	86
Chapitre 4 – Conclusion	92
Références	95

Liste des tableaux

Table I. 5'-sequence heterogeneity ^a of affinity-purified SLI RNAs synthesized as SLI-ARiBo precursors from the T7 class III promoter	50
Table II. 5'-sequence heterogeneity ^a of affinity-purified SLI RNAs synthesized as SLI-ARiBo precursors from the T7 class II ϕ 2.5 promoter	51

Liste des figures

Chapitre 1

Figure 1.1 : Schéma de la transcription par la polymérase à ARN du phage T7.	4
Figure 1.2 : Présentation générale de ribozymes nucléolytiques.	7
Figure 1.3 : Purification d'ARN par gel dénaturant de polyacrylamide.	9
Figure 1.4 : Représentation schématique des étapes de la purification d'ARN par affinité utilisant l'étiquette ARiBo.	14
Figure 1.5 : Structure secondaire et réaction de clivage de l'étiquette ARiBo.	16
Figure 1.6 : Protéine de fusion $\lambda N^+ - L^+ - GST$	17
Figure 1.7 : Structure secondaire du ribozyme hammerhead.	20
Figure 1.8 : Mode de fonctionnement du CRISPR.	21

Chapitre 2

Figure 2.1 : Evidence of 5'-heterogeneity revealed from affinity purification of an SLI RNA transcribed from an SLI-ARiBo precursor.	52
Figure 2.2 : Effect of the 5' sequence on the heterogeneity of SLI RNAs transcribed as SLI-ARiBo precursors from the consensus T7 class III promoter.	53
Figure 2.3 : Effect of the 5' sequence on the heterogeneity of SLI RNAs transcribed as SLI-ARiBo precursors from the T7 class II $\phi 2.5$ promoter. ...	54
Figure 2.4 : Effect of CRISPR-RNA junction sequence on Cse3 cleavage.	55
Figure 2.5 : Affinity purification of RNA with homogeneous ends using CRISPR and ARiBo tags.	56
Supplementary Figure 2.1 : The ARiBo4 tag.	63

Supplementary Figure 2.2 : Evidence of 5'-heterogeneity from VS ribozyme cleavage.....	64
Supplementary Figure 2.3 : Effect of CRISPR-RNA junction sequence on Cse3 cleavage at 37°C.....	65
Supplementary Figure 2.4 : Reversion of the CRISPR 5'-tail back to the wild-type sequence favors Cse3 cleavage of a CRISPR-SLI-ARiBo double-fusion RNA with a U-A base pair closing the SLI stem.....	66
Supplementary Figure 2.5 : Reversion of the CRISPR 5'-tail back to the wild-type sequence favors Cse3 cleavage of a CRISPR-SLI-ARiBo double-fusion RNA with a U linker between the CRISPR and the SLI(2) RNA.	67
Supplementary Figure 2.6 : Small-scale affinity batch purifications.....	69

Chapitre 3

Figure 3.1 : Formule du calcul de l'hétérogénéité de séquence en 5' d'un ARN transcrit par la polymérase à ARN du phage T7.....	72
Figure 3.2 : Structure secondaire du ribozyme VS de <i>Neurospora</i> et de son substrat SLI.	73
Figure 3.3 : Représentation schématique des étapes de la méthode de purification ARiBo jumelée à l'utilisation d'une étiquette hammerhead pour éliminer l'hétérogénéité en 5'.....	77
Figure 3.4 : Représentation schématique simplifiée des grandes étapes de la biogénèse d'un polycistron de miARN.....	79
Figure 3.5 : Séquence du polycistron de miARN miR106b-25.	81
Figure 3.6 : Prédiction Sfold de miR-106b-25-ARiBo1 et miR-106b-25-ARiBo4.	82
Figure 3.7 : Différentes étiquettes ARiBo utilisées pour la purification du polycistron miR-106b-25.	84

Figure 3.8 : Gel de polyacrylamide de la purification du polycistron miR-106b-25.	85
Figure 3.9 : Représentation schématique de différentes méthodes d'immobilisation de l'ARN pour l'isolation de RNP.	87
Figure 3.10 : Représentation schématique des étapes d'isolation de complexes ribonucléoprotéiques dérivées de la méthode de purification par affinité utilisant l'étiquette ARiBo.	88
Figure 3.11 : Isolation de complexes ribonucléoprotéiques impliquant pre-miR- 153-2.	90

Liste des abréviations

^{13}C	Carbone 13
^{15}N	Azote 15
ADN	Acide désoxyribonucléique (DNA : <i>Desoxyribonucleic acid</i>)
ARiBo	<i>Activatable ribozyme</i> avec une séquence d'ARN λ <i>BoxB</i>
ARN	Acide ribonucléique (RNA : <i>Ribonucleic acid</i>)
ARNm	ARN messenger
ARNr	ARN ribosomal
ARNt	ARN de transfert
ATP	Adenosine triphosphate
CAS	<i>CRISPR-associated</i>
CPG	<i>Controlled pore glass</i>
CRISPR	<i>Clustered regularly interspace short palindromic repeats</i>
CTP	Cytidine triphosphate
DGCR8	<i>DiGeorge syndrome critical region gene 8</i>
DMT	Dimethoxytrityl
GlcN6P	Glucosamine-6-phosphate
GSH	Glutathion réduite
GST	Glutathion S-transférase
GTP	Guanosine triphosphate
HDV	<i>Hepatitis delta virus</i>
HH	<i>Hammerhead</i>
HPLC	<i>High performance liquid chromatography</i>
miARN	Micro ARN
nt	Nucléotide
NTP	Nucléotide triphosphate
oligo(dT)	Oligomère de deoxy-thymine
PAGE	<i>Polyacrylamide gel electrophoresis</i>
PACT	<i>Protein activator of PKR</i>
PCR	<i>Polymerase chain reaction</i>

piARN	Piwi ARN
PIWI	<i>P-element induced wimpy testis</i>
pre-miARN	Précurseur de miRNA
pri-miARN	Transcrit primaire de miARN
R-PCR	<i>Recursive polymerase chain reaction</i>
RMN	Résonance magnétique nucléaire
RNP	Ribonucléoprotéine
RNase	Ribonucléase
SA	Streptavidine
SDS	<i>Sodium dodecyl sulfate</i>
SEC	<i>Size exclusion chromatography</i>
siARN	<i>Silencing ARN</i>
TRBP	<i>Trans-activation RNA-binding protein</i>
tRSA	Étiquette ARNt jumelé à un aptamère streptavidine
UTP	Uridine triphosphate
UV	Ultraviolet
VS	<i>Varkud satellite</i>

Remerciements

Je tiens tout d'abord à remercier Pascale Legault pour m'avoir accueilli dans son laboratoire d'abord comme stagiaire et ensuite comme étudiante. L'encadrement et le soutien dont j'ai bénéficié tout au long de mon parcours m'ont grandement aidé à avancer et ont su me garder motivée même lorsque rien ne fonctionnait. Merci de m'avoir donné cette opportunité, mais surtout merci de m'avoir fait confiance et de m'avoir donné un projet bien à moi dès le début de mon stage.

Merci à toute l'équipe des laboratoires Legault et Omichinski pour votre aide et votre soutien moral, je n'aurais pas pu me retrouver au sein d'une équipe aussi dynamique et unie. Merci à Nico, Éric, Pat, Jo, Alex, Lauriane, Pierre, Geneviève, Matt, Phil, Julie pour les discussions, les 5 à 7 qui s'étirent parfois jusqu'au lendemain, le canot-camping, les parties de volleyball et j'en passe. Un merci tout particulier à Geneviève pour avoir eu la patience de tout me montrer dans le laboratoire et de m'aider dans tout ce que je fais. Merci également à Nico et Lauriane pour avoir pris la peine de lire certaines sections de ce mémoire et à Matt et Pat pour m'avoir partager vos expériences récentes de rédaction. Merci à tous les gens du bureau pour répondre à toutes mes questions, je sais que j'en ai souvent beaucoup, et pour toutes nos discussions d'ordre personnelles ou scientifiques, vous m'aidez beaucoup à avancer. Finalement, merci à tous les autres membres du laboratoire, passés ou présents, que ce soit pour avoir raquer mes tips, avoir des longues discussions interminables autour d'une table de billard ou simplement dit bonjour le matin, c'est à ce niveau que commence le travail d'équipe et c'est souvent ce qui me motive à me lever le matin.

Mes remerciements ne seraient pas complets sans mentionner les gens du département. Je ne veux nommer personne ici de peur d'en oublier, mais que ce soit les employés ou des étudiants provenant d'autres laboratoires, vous m'avez beaucoup aidé à plusieurs niveaux et je vous en remercie. La vie départementale est quelque chose qui me tient énormément à cœur et je suis contente de voir que je ne suis pas la seule.

Merci à ma famille qui s'est toujours intéressée à ce que je fais, même si pour la plupart vous n'y comprenez rien. C'est vous qui à chaque occasion m'apportez tous les articles à caractère un peu scientifique que vous trouvez, même si ce n'est pas du tout dans mon domaine ça me fait toujours plaisir. Merci à mes parents de me supporter et de m'encourager depuis le tout début. Quand je n'étais pas certaine d'une décision que j'avais prise, vous l'étiez pour moi, vous m'avez poussée à continuer et à toujours donner le meilleur de moi-même et je ne vous en remercierai jamais assez.

Finalement, merci à Seb, mon amour, même si tout ce que tu as accepté de lire dans ce mémoire est l'abstract, saches que ton soutien moral, ton aide et ta patience me sont chers. Ta passion pour ton travail et ta soif de réussir sont inspirantes. Merci d'être là pour moi quand j'ai besoin de toi.

Merci à vous tous, je suis contente de pouvoir continuer cette aventure aussi bien entourée, je suis extrêmement choyée de tous vous avoir dans ma vie.

Chapitre 1 – Introduction

1.1 – Importance de l'ARN dans les processus biologiques

Dans les dernières années, de nombreuses recherches ont démontré l'importance de l'ARN dans les processus biologiques. Les premiers ARN identifiés furent l'ARN messager (ARNm), l'ARN de transfert (ARNt) et l'ARN ribosomal (ARNr)¹. Par la suite, les ARN non codants sont devenus des sujets de recherche d'une importance capitale dans plusieurs domaines tels le cancer, le développement, les maladies auto-immunes ou les maladies neurodégénératives². Les ARN non codants regroupent plusieurs catégories d'ARN dont les longs ARN non-codants responsables de plusieurs phénomènes tels que la régulation transcriptionnelle, l'épissage alternatif et bien d'autres². Cette catégorie d'ARN regroupe également les petits ARN non-codant tels que les ARN interférant (siARN), les microARN (miARN) et les ARN interagissant avec la protéine PIWI (piARN). Les siARN et les miARN agissent sur la régulation génique au niveau de l'ARNm alors que les piARN, beaucoup moins bien caractérisés, vont supprimer l'activité des transposons. Parmi les autres classes d'ARN non codants, on retrouve les riborégulateurs qui sont des ARN régulateurs de gènes qui s'activent ou s'inactivent selon la présence d'un ligand spécifique³, et les ribozymes qui sont des ARN possédant une fonction catalytique⁴. Ces ARN non codants jouent un rôle important dans la régulation de l'expression génique à presque tous les niveaux, depuis la régulation épigénétique jusqu'à la régulation transcriptionnelle ou post-transcriptionnelle. Considérant l'importance de l'ARN en biologie et le nombre grandissant d'études sur le sujet, la purification de grandes quantités d'ARN est devenue nécessaire à plusieurs sphères de recherche.

1.2 – Préparation d'ARN *in vitro*

Différentes méthodes existent afin de préparer un échantillon d'ARN. La synthèse d'acides nucléiques *in vitro* est primordiale afin d'obtenir des quantités suffisantes d'ARN pour plusieurs types d'études. Par exemple, cette méthode permet le marquage isotopique d'ARN grâce à des NTP marqués (¹³C/¹⁵N), pour faire des expériences de résonance magnétique nucléaire (RMN). Il existe deux

méthodes pour la synthèse d'ARN *in vitro*. La première consiste à synthétiser chimiquement l'ensemble de la molécule à l'aide d'un synthétiseur automatique⁵. Plusieurs compagnies offrent ce genre de service, qui peut rapidement devenir très coûteux. De plus, pour les séquences excédant environ 60 nucléotides, la méthode devient moins efficace en raison de faibles rendements obtenus. Une méthode alternative à la synthèse chimique d'ARN est la synthèse enzymatique qui permet l'obtention de plus longues séquences d'ARN, généralement à moindre coût et avec un meilleur rendement.

1.2.1 – Synthèse enzymatique d'ARN

La synthèse enzymatique d'ARN consiste à produire la séquence d'ARN désirée à partir de nucléotides triphosphates (des NTP, soit ATP, CTP, GTP et UTP). Il existe différentes polymérases utilisées à cette fin, la plus utilisée étant la polymérase à ARN du phage T7⁶⁻⁸. La transcription *in vitro* par cette polymérase est particulièrement répandue de par sa facilité d'utilisation. Elle permet de transcrire des ARN d'une longueur pratiquement illimitée et en grande quantité (de l'ordre du milligramme) pour une très grande variété d'applications *in vitro* et *in vivo*.

Type de matrice utilisée

Pour la synthèse enzymatique d'ARN avec la polymérase du phage T7, une matrice d'ADN est utilisée. Cette matrice doit contenir une séquence complémentaire à la séquence d'ARN désirée ainsi qu'une région double brin de 18 nucléotides qui constitue le promoteur du phage T7⁹ (Figure 1.1). La matrice peut prendre différentes formes, soit celle d'un ADN synthétique (double-brin ou partiellement simple-brin), d'un produit de PCR ou d'un plasmide. Différentes stratégies peuvent être utilisées pour modifier le plasmide qui permettra la synthèse de l'ARN désiré. Pour de courtes séquences, la mutagénèse dirigée¹⁰ est suggérée alors que pour les séquences plus longues (à partir d'environ 75 nt), le PCR récursif (R-PCR) est l'option idéale¹¹.

Le plasmide doit être linéarisé avant le début de la transcription en utilisant une enzyme de restriction dont le site de reconnaissance est situé à l'extrémité 3'

de la séquence à transcrire. En effet, un plasmide linéaire permet l'arrêt de la transcription au site désiré, alors qu'avec un plasmide circulaire la transcription se poursuivrait tout au long du plasmide.

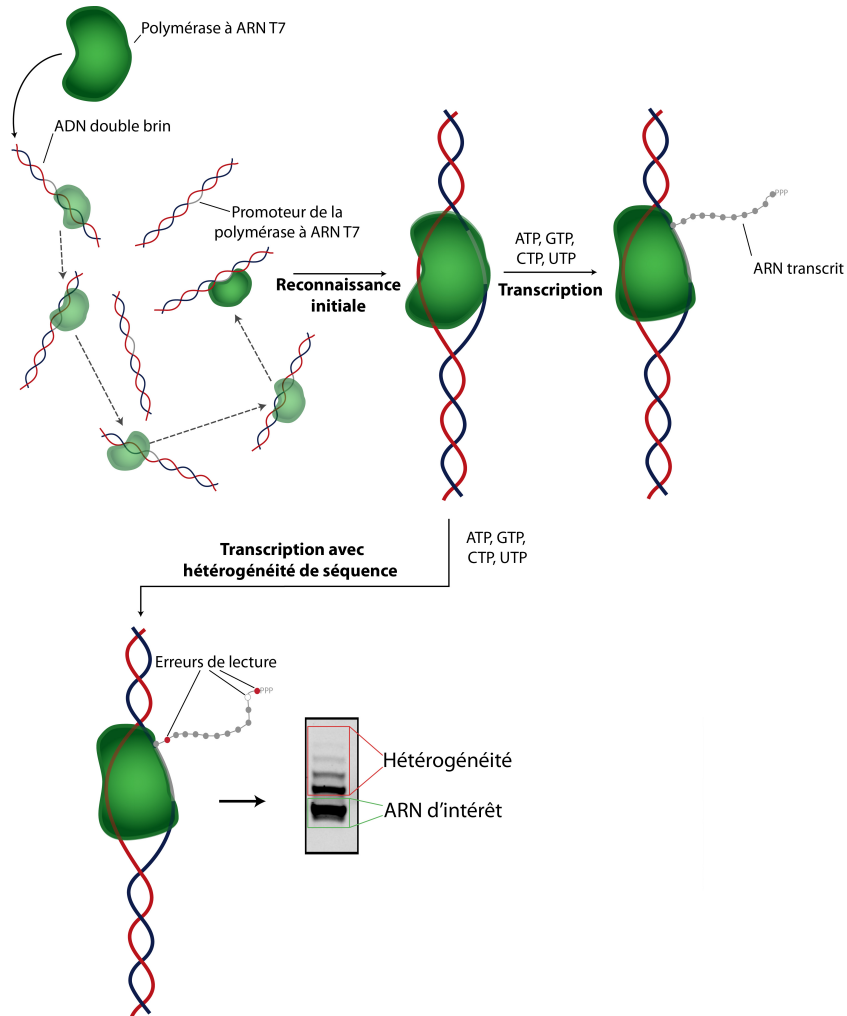


Figure 1.1 : Schéma de la transcription par la polymérase à ARN du phage T7. La transcription de la séquence désirée est illustrée, ainsi que les cas où de l'hétérogénéité de séquence pourrait survenir aux extrémités 5' et 3'. Un gel de polyacrylamide dénaturant montre la présence d'hétérogénéité résultant de la synthèse d'un ARN débutant par GGG à son extrémité 5' (Adapté de Chamberlin, 1974)¹².

Type de promoteur

Une des limites de l'utilisation de la polymérase à ARN du phage T7 se situe au niveau du choix de la séquence transcrite. En effet, cette polymérase

reconnait et transcrit uniquement des séquences débutant par une purine (soit une adénine ou une guanine)⁸. Le promoteur est une séquence d'ADN double brin de 18 nucléotides qui est reconnue par la polymérase pour permettre l'initiation de la transcription. Deux classes de promoteurs sont généralement utilisées dépendamment de la nature du nucléotide par lequel débute la séquence. Dans la majorité des cas, un promoteur de classe III est utilisé pour produire un ARN débutant par un G^{13,14}. Pour les séquences débutant par un A, un promoteur de classe II ϕ 2.5 peut aussi être utilisé, mais a été moins bien caractérisé¹⁵. Une étude suggère que les deux types de promoteur donnent des rendements de transcription semblables, mais que le promoteur de classe II ϕ 2.5 produirait moins d'hétérogénéité en 5'¹⁵.

Erreurs de transcription dues à la polymérase à ARN du phage T7

L'hétérogénéité de séquence des ARN transcrits est due à des erreurs de lecture de la polymérase. La polymérase à ARN du phage T7 peut produire des avortons, c'est-à-dire qu'elle peut terminer prématurément la transcription de l'ARN. Ce problème survient surtout au niveau de l'initiation de la transcription. Par contre, ces avortons allant généralement de 2 nt à 12 nt⁷ peuvent facilement être séparés de l'ARN pleine longueur par une étape de purification subséquente, en raison de la grande différence de taille entre ces petits ARN avortons et l'ARN pleine longueur.

L'hétérogénéité en 5' et en 3' est causée par des erreurs de lecture de la polymérase à ARN du phage T7 lors de l'initiation (en 5') et lors de la terminaison (en 3')¹⁶. Lors de l'initiation, de l'hétérogénéité en 5' est produite en fonction de la nature de la séquence¹⁷. Par exemple, les séquences ayant plusieurs G en 5' (GGG) présentent beaucoup plus d'hétérogénéité qu'une séquence débutant par GCG qui ne présente que très peu d'hétérogénéité¹⁸ (Figure 1.1). Lors de la terminaison, la polymérase ajoute souvent un nucléotide additionnel en 3' (A, C, G ou U) non codé par la matrice. L'hétérogénéité de séquence en 5' et 3' est problématique car les ARN contaminants diffèrent généralement de l'ARN désiré par seulement un, deux ou trois nucléotides (N+1, N+2, N+3, etc.), ce qui rend la purification de

l'ARN désiré très difficile par les méthodes classiques.

Mutant P266L

Afin de réduire la quantité d'avortons, la polymérase à ARN du phage T7 a été mutée en changeant la proline 266 par une leucine. Cette polymérase mutante P266L permet de réduire les arrêts prématurés au niveau de l'initiation de la transcription^{7,19}, et ainsi produit moins d'avortons que la protéine sauvage. Dans certains cas, cette polymérase mutante produit de meilleurs rendements de transcription²⁰. Il est également possible qu'elle aide à diminuer l'hétérogénéité en 5' des ARN transcrits, mais cette hypothèse n'a pas été testée de façon systématique.

1.2.2 – Utilisation de ribozymes pour la préparation d'ARN

Les ribozymes sont des ARN qui catalysent une réaction chimique en absence de protéines et représentent un outil important dans la préparation d'ARN. Une classe particulièrement intéressante est celle des ribozymes nucléolytiques qui incluent le hammerhead (HH)²¹, le *hepatitis delta virus* (HDV)²², le *Varkud Satellite* (VS)²³ et le *glmS*²⁴ (Figure 1.2A). Les trois premiers ribozymes font parti des ARN satellites, viral ou virusoïde alors que le *glmS* est un ARN commutateur (*riboswitch*) retrouvé chez les bactéries Gram positive^{25,26}. Malgré leur structures secondaires différentes, tous ces ribozymes s'autoclivent *in vivo* suivant le même type de réaction, mais sont chacun spécifiques à une séquence qui leur est propre (Figure 1.2B).

Cette réaction chimique est une trans-estérification où le 2'-OH d'un ribose spécifique attaque le 3'-phosphate adjacent (Figure 1.2B)^{26,27}. Les produits de clivage ont des extrémités 2', 3'-cyclique phosphate et 5'-OH. Les ribozymes nucléolytiques catalysent des réactions intramoléculaires et plusieurs ont été modifiés pour catalyser des réactions en *trans in vitro*, c'est-à-dire que la réaction chimique décrite plus haut servira au clivage d'un substrat externe. En effet, il est possible de changer la séquence de ces ribozymes, tout en conservant les résidus essentiels au clivage, afin de cliver un substrat donné. Ces modifications des ribozymes nucléolytiques permettent leur utilisation *in vitro* pour différentes

applications nécessitant le clivage d'un ARN²⁵. Par exemple, ils peuvent être facilement utilisés dans la préparation d'ARN pour éliminer l'hétérogénéité aux extrémités 5' et 3'²⁸.

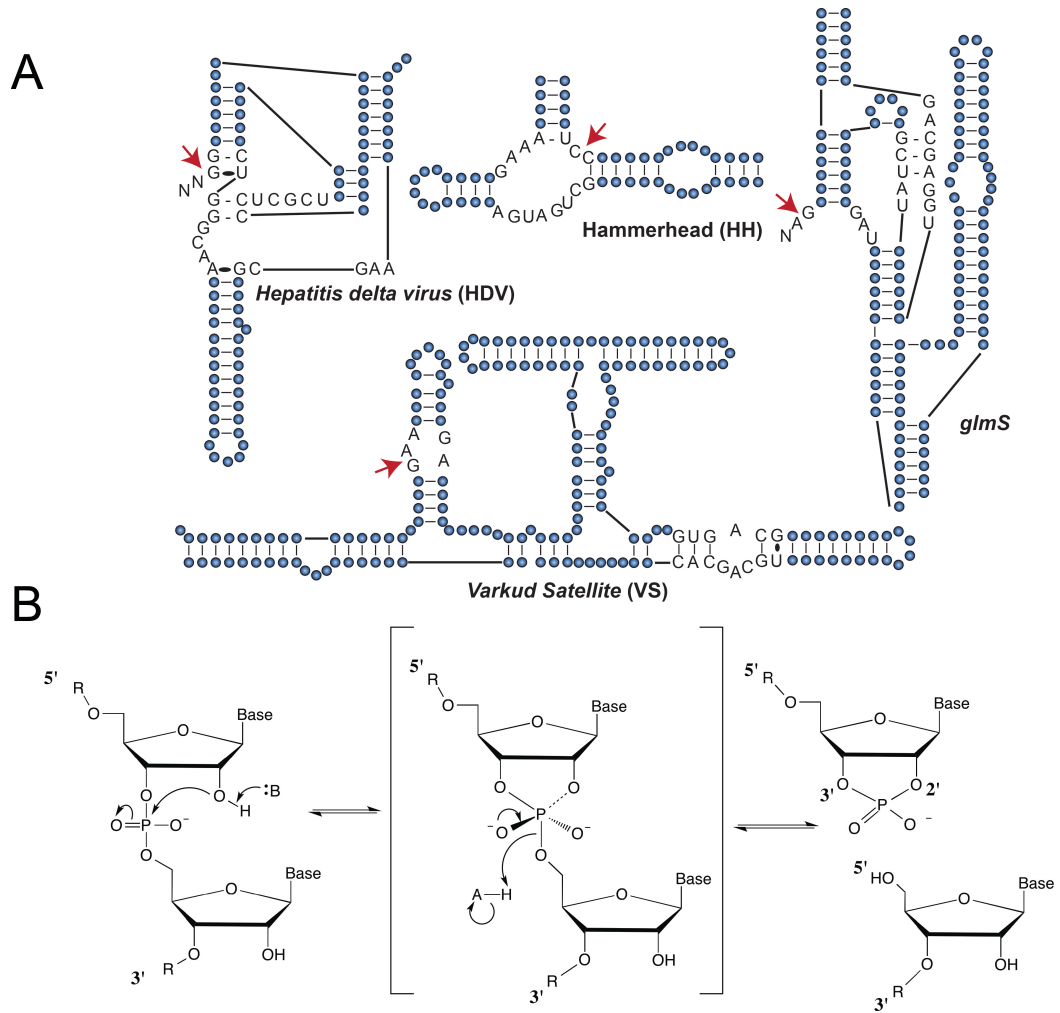


Figure 1.2 : Présentation générale de ribozymes nucléolytiques. (A) Structures secondaires des ribozymes HDV, HH, glmS et VS. Les nucléotides importants pour le clivage sont indiqués par des lettres sur les structures alors que les autres nucléotides sont illustrés d'un point. Le site de clivage de chaque ribozyme est indiqué par une flèche. (Adaptée de Cochrane et Strobel, 2008²⁵) (B) Mécanisme général de clivage des ribozymes nucléolytiques. Ce mécanisme est entre autre utilisé par les ribozymes illustrés en A (Adapté de Ferré-d'Amaré et Scott, 2010 et de Viladoms et Fedor, 2012^{26,29}).

En conclusion, parmi les deux méthodes existantes pour préparer de l'ARN *in vitro*, la synthèse enzymatique *in vitro* par la polymérase à ARN du phage T7 est la plus répandue. Cependant, malgré son utilisation fréquente, cette méthode produit plusieurs autres ARN que celui désiré. D'autres outils, tels que les ribozymes, peuvent aider à la préparation d'un ARN. Dans tous les cas, il est généralement important de bien choisir la méthode de purification lors de la préparation de l'ARN désiré.

1.3 – Purification d'ARN

Le besoin grandissant d'ARN très pur a augmenté dans les dernières années avec la découverte de nouvelles fonctions de l'ARN dans les processus biologiques. Par exemple, pour des études structurales, de grandes quantités d'ARN très pur sont requises; la moindre hétérogénéité, chimique ou conformationnelle nuit considérablement à l'obtention de structures. Pour les études fonctionnelles, la purification d'un ARN sous forme native et sans hétérogénéité est également importante. Plusieurs méthodes de purification existent : la méthode traditionnelle de purification d'ARN par gel dénaturant de polyacrylamide, les méthodes de chromatographie et les nouvelles méthodes de purification par affinité. Chacune de ces méthodes de purification d'ARN comporte ses avantages et ses inconvénients. Le choix de la méthode est primordial et dépend des moyens disponibles dans un laboratoire, de l'utilisation faite avec l'ARN suite à cette purification et de la facilité de purification de l'ARN. En effet, certains ARN sont plus faciles à obtenir sous forme native due à la stabilité de leur structure, alors que d'autres, aux structures plus complexes et moins stables, seront plus sensibles à toutes conditions dénaturantes^{30,31}.

1.3.1 – Méthode de purification par gels dénaturants de polyacrylamide

La méthode traditionnelle de purification d'ARN repose sur l'électrophorèse par gel dénaturant de polyacrylamide^{30,32} (Figure 1.3). La méthode consiste à faire migrer l'ARN sur des gels de polyacrylamide contenant de 7 M à 8 M urée et, après avoir coupé la bande d'ARN désirée, extraire l'ARN du gel. Cette méthode permet de produire des ARN de 50 nt ou moins avec un bon

rendement et une très grande pureté. Plus un ARN est long, plus il sera difficile à obtenir avec une grande pureté, car les bandes d'ARN contaminants avec hétérogénéité en 5' et 3' se retrouveront près de la bande d'ARN d'intérêt et seront moins bien séparées sur un gel de polyacrylamide. De plus, la méthode peut devenir très longue et fastidieuse lorsque de très grandes quantités d'ARN sont nécessaires puisque le nombre de gels se multiplie avec la quantité d'ARN désirée³². D'autres problèmes souvent rencontrés lors de purifications par gel sont la présence d'oligomères de polyacrylamide dans l'échantillon d'ARN purifié et les dommages à l'ARN dus aux rayons UV utilisés lors de la révélation des bandes sur le gel^{33,34}. Un autre désavantage important de cette méthode est que l'ARN est systématiquement dénaturé, ce qui peut être un inconvénient majeur pour les études structurales et fonctionnelles. L'ARN peut être replié sous sa forme native en utilisant la méthode de « *heat and snap cool* » qui consiste à chauffer l'ARN afin de défaire toute forme d'interaction et ensuite le mettre rapidement sur glace pour qu'il se replie dans sa forme native. Par contre, cette méthode fonctionne seulement pour les petits ARN, principalement les petites tiges boucles, n'ayant pas de structures secondaires complexes³⁵. Pour toutes ces raisons, plusieurs méthodes de purification alternatives ont été développées.

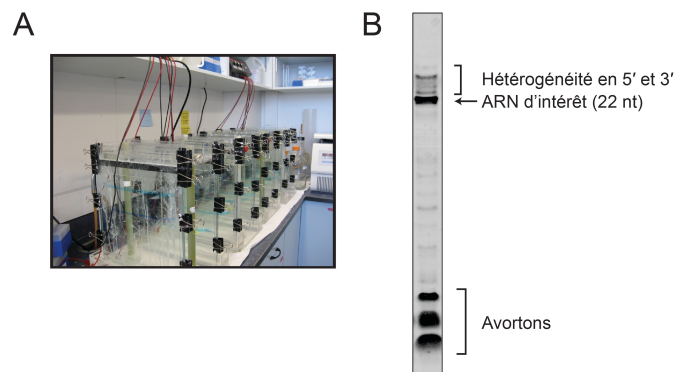


Figure 1.3 : Purification d'ARN par gel dénaturant de polyacrylamide. (A) Photo illustrant la purification de plusieurs milligrammes d'ARN par gels dénaturants de polyacrylamide. (B) Gel de polyacrylamide dénaturant illustrant les produits de transcription *in vitro* par la polymérase à ARN du phage T7 d'un ARN de 22 nt (Bouchard et Legault, résultats non publiés).

1.3.2 – Méthodes de purification d'ARN par chromatographie

D'autres méthodes de purification ont vu le jour pour répondre aux besoins d'ARN purifiés sous une forme native. Une des voies explorées consiste à l'utilisation de la chromatographie, soit échangeuse d'ions, soit SEC (*size exclusion chromatography*)³³. Ces techniques sont moins laborieuses que la méthode de purification par gels dénaturants lorsque de grandes quantités d'ARN (de l'ordre du milligramme) sont nécessaires, car elles peuvent être facilement automatisées. De plus, la méthode SEC permet la séparation entre les monomères et les multimères (ou agrégats) d'ARN, ce qui n'est pas possible avec toutes les méthodes de purification. La faiblesse majeure des méthodes de chromatographie est la faible résolution, soit la capacité à bien séparer les différents d'ARN retrouvés dans la transcription. Par exemple, les ARN différant par seulement un nucléotide de la séquence d'intérêt ne pourront pas être séparés avec ces méthodes. Par contre, ce problème peut être contourné en ajoutant des étiquettes d'ARN clivables, comme le ribozyme hammerhead, aux extrémités de l'ARN d'intérêt. Ainsi, ces méthodes représentent une bonne alternative pour l'obtention d'un ARN natif conformationnellement et chimiquement pur. Par contre, la méthode est plus chronophage si plusieurs ARN doivent être purifiés avec un même système HPLC. De plus, pour assurer un environnement complètement dépourvu de RNase, un système HPLC doit être dédié exclusivement à l'ARN, ce qui occasionne des coûts élevés.

1.3.3 – Purification d'ARN par affinité

Méthodes existantes pour la purification de protéines par affinité

La purification par affinité se fait beaucoup pour la purification de protéines³⁶. Une approche bien connue consiste à ajouter une étiquette protéique à l'extrémité C-terminale ou N-terminale de la protéine à purifier. Cette étiquette sert à lier une matrice afin d'immobiliser la protéine d'intérêt pour la purifier. Par exemple, l'ajout de six résidus histidine à l'extrémité N-terminale ou C-terminale de la protéine d'intérêt permettra d'immobiliser la protéine de fusion résultante sur une colonne contenant des ions de nickel immobilisés. Après plusieurs lavages

servant à se débarrasser des impuretés, l'élution se fait dans un tampon approprié afin que la protéine de fusion se détache de la macromolécule à laquelle elle est rattachée. Dans le cas de l'étiquette hexa-histidine, une concentration élevée d'imidazole (200 mM ou plus) est utilisée pour permettre l'élution de la protéine de fusion. Une autre étiquette très utilisée pour les protéines est la GST (Glutathion S-Transférase) qui permettra à la protéine de fusion résultante de se lier spécifiquement à la résine GSH-sepharose (le GSH est la forme réduite du glutathion) puis d'être éluée par l'ajout de glutathion. Dans le cas des deux méthodes décrites ci-dessus, la protéine éluée est une protéine de fusion, ce qui ne nuit pas toujours aux expériences faites avec la protéine d'intérêt. Par contre, dans les cas où une protéine sans étiquette est nécessaire, l'étiquette peut être clivée avec la thrombine ou la protéase TEV, tout dépendant du site de clivage incorporé entre l'étiquette et la protéine d'intérêt³⁷.

Pour les protéines, les méthodes de purification par affinité sont simples et facilement applicables. Il est possible de trouver des kits de purification de protéines sur le marché. Ces méthodes peuvent également être facilement jumelées afin d'obtenir une plus grande pureté de la protéine. Par contre, même si les méthodes de purification par affinité pour l'ARN sont fortement inspirées des méthodes de purification de protéines, ce domaine n'a pas encore été très exploré. Certains protocoles existent, mais il est impossible de trouver des systèmes disponibles commercialement pour la purification d'ARN par affinité comme il est possible d'en trouver pour les protéines.

Purification d'ARN par affinité : Immobilisation de l'ARN

Bien qu'aucun système commercial ne soit disponible, différentes méthodes de purification d'ARN par affinité ont été développées dans les dernières années. Le principe de purification dépend principalement de deux étapes simples, soit l'immobilisation de l'ARN d'intérêt sur une matrice et l'élution de cet ARN. Une des méthodes pour immobiliser l'ARN utilise une étiquette d'ARN en 3' ayant une séquence spécifique pour l'hybridation à des oligonucléotides d'ADN³⁸. Il s'agit d'une méthode très simple utilisée depuis plusieurs années pour isoler les ARNm polyadénylés à l'aide d'une résine

oligo(dT). D'autres méthodes ont été mises au point qui utilisent une étiquette d'ARN en 3' contenant un aptamère qui reconnaît spécifiquement une matrice d'affinité telle que la streptavidine³⁹ ou le Sephadex⁴⁰. Ces deux stratégies d'immobilisation sont très semblables, mais celle utilisant la streptavidine semble plus efficace en raison de la plus forte affinité de l'aptamère streptavidine à la matrice. Le protocole de purification utilisant la streptavidine a même été adapté pour l'isolation de complexes ARN-protéines⁴¹, ce qui rend ce protocole très polyvalent. Des étiquettes d'ARN liant des protéines ou des domaines protéiques sont inclus parmi les méthodes de purification par affinité d'ARN les plus développées. La stratégie consiste à utiliser une protéine de fusion ayant un domaine de liaison à l'étiquette d'ARN ainsi qu'un domaine de liaison à une matrice donnée. Parmi les protéines utilisées pour lier une étiquette d'ARN on retrouve la protéine MS2 qui lie spécifiquement des petites tiges-boucles⁴².

Purification d'ARN par affinité : Éluéon de l'ARN

Pour l'éluéon de l'ARN, plusieurs options sont disponibles. Par exemple, pour la méthode utilisant des ARNm polyadénylés et purifiés à l'aide d'une résine oligo(dT), utiliser de la chaleur et un tampon de faible force ionique permet l'éluéon de l'ARN³⁸. Une méthode différente consiste à utiliser un DNase afin de cliver l'ARN spécifiquement en *trans*⁴³. L'inconvénient majeur de cette méthode est la nécessité d'ajouter des étapes à la purification afin d'éliminer l'enzyme qui co-éluera avec l'ARN d'intérêt. Une autre solution particulièrement intéressante pour éluer l'ARN est l'utilisation d'une étiquette avec ribozymes activables⁴⁴. Ces ribozymes s'autoclivent une fois en présence d'un activateur spécifique, généralement une petite molécule chimique, ce qui rend possible l'éluéon de l'ARN d'intérêt par autoclivage suite à l'ajout d'un activateur.

Deux ribozymes activables ont été utilisés pour la purification d'ARN par affinité. Le premier est le ribozyme *hepatitis delta virus* (HDV) mutant qui est actif seulement en présence d'imidazole⁴⁵. Le second est le ribozyme *glmS*, d'origine bactérienne et naturellement clivable en présence de glucosamine-6-phosphate (GlcN6P)⁴⁶ (Figure 1.2). Ces deux ribozymes possèdent un site de clivage avec une séquence de reconnaissance minimale à leur extrémité

5'. Ils peuvent donc être positionnés en 3' de l'ARN d'intérêt et n'ont pas à être modifiés pour être compatibles avec un ARN donné. Le ribozyme HDV mutant activable par l'imidazole nécessite une haute concentration d'imidazole (environ 500 mM) et un long temps d'incubation (environ 2 h à 37 °C) afin de pouvoir cliver l'ARN d'intérêt. Ces conditions de clivage peuvent occasionner de la dégradation pour les ARN plus fragiles ce qui n'est donc pas idéal pour l'obtention d'un ARN très pur. Le ribozyme *glmS* quant à lui peut être clivé avec une concentration relativement basse de GlcN6P (environ 1 mM) et un temps du ribozyme *glmS* est qu'il se clive en présence de tampon Tris, ce dernier ne doit donc pas être utilisé lors des premières étapes de purification.

1.4 – Purification d'ARN par affinité utilisant l'étiquette ARiBo

Malgré la diversité des méthodes existantes de purification d'ARN par affinité, aucune des méthodes disponibles n'a été développée en mettant l'accent sur à la fois la maximisation du rendement et de la pureté de l'ARN. Une nouvelle méthode de purification par affinité a été mise au point dans notre laboratoire pour purifier l'ARN sous sa forme native et obtenir une grande quantité d'ARN très pur⁴⁷.

1.4.1 – Description de la méthode

La méthode de purification par affinité développée dans notre laboratoire consiste à utiliser une étiquette ARiBo (*Activatable Ribozyme* avec une séquence d'ARN λ *BoxB*) contenant le ribozyme *glmS* et l'ARN *BoxB* du phage λ se liant à une protéine de fusion λ N⁺-L⁺-GST⁴⁷. Cette méthode utilise la technologie GST/GSH-sepharose pour l'immobilisation de l'ARN et le ribozyme *glmS* pour l'élution de l'ARN d'intérêt.

La méthode ARiBo consiste à transcrire, grâce à la polymérase à ARN du phage T7, l'ARN d'intérêt avec l'étiquette ARiBo en 3' de manière à obtenir un ARN de fusion (Figure 1.4). Cet ARN de fusion, lorsque mis en présence de la protéine de fusion λ N⁺-L⁺-GST, s'y lie avec une très grande affinité. Le complexe ARN-protéine peut ensuite être immobilisé sur une résine commerciale, la GSH-

sepharose. L'avantage d'utiliser la résine GSH-sepharose est qu'elle est relativement peu dispendieuse, facilement accessible et est couramment utilisée pour la purification de protéine. La GlcN6P est ensuite ajoutée pour activer le ribozyme *glmS* de l'étiquette ARiBo et ainsi cliver l'ARN d'intérêt qui est récupéré dans l'élution. Contrairement aux méthodes de purification par chromatographie utilisant une colonne, cette méthode a été développée en mode type « batch », c'est-à-dire qu'elle ne nécessite pas une colonne, peut être réalisée dans n'importe quel type de tube sans limitation au niveau des quantités d'ARN produit. Un autre avantage de l'utilisation de la résine GSH-sepharose est qu'elle peut être régénérée suite à la purification de l'ARN et réutilisée afin de réduire les coûts reliés à la purification d'un ARN.

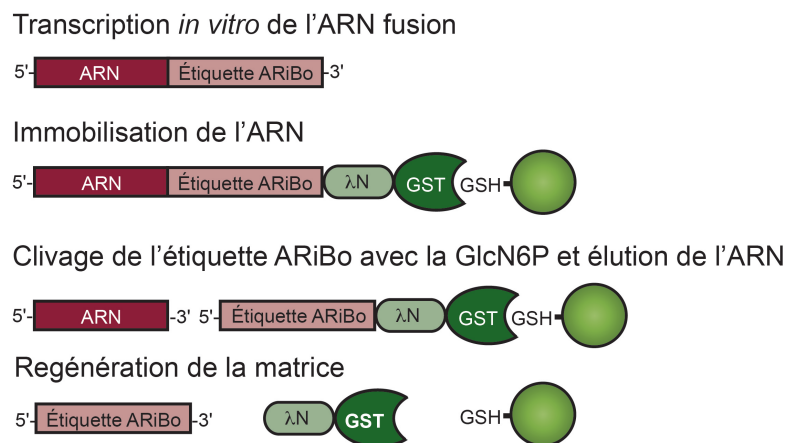


Figure 1.4 : Représentation schématique des étapes de la purification d'ARN par affinité utilisant l'étiquette ARiBo. Les grandes étapes de purification sont illustrées, de la transcription de l'ARN jusqu'à la régénération de la résine GSH-sepharose après avoir récupéré l'ARN d'intérêt (Adaptée de Di Tomasso et al, 2011)⁴⁷.

L'étiquette ARiBo

L'étiquette ARiBo est une étiquette d'ARN ajoutée en 3' de l'ARN d'intérêt^{47,48}. Cette étiquette est formée du ribozyme *glmS* de *Bacillus anthracis* activable avec la GlcN6P (Figure 1.5A). Pour l'étiquette ARiBo, la séquence du ribozyme *glmS* a été modifiée au niveau de sa tige-boucle P1 pour y substituer la séquence d'ARN *BoxB* provenant du bactériophage λ. L'ARN *BoxB* se lie avec

une très haute affinité (de l'ordre du picomolaire) au peptide λN^+ dérivé de la protéine N^+ du phage λ ⁴⁷.

Pour le clivage du ribozyme *glmS* et l'éluion de l'ARN d'intérêt, la réaction de clivage débute en utilisant le 2'-OH d'un ribose (une adénosine) comme nucléophile qui sera déprotoné par une guanosine du site de liaison de la GlcN6P (Figure 1.5B)⁴⁶. De son côté, le groupe partant sera protoné par la GlcN6P qui agit comme acide général dans la réaction. Tel que mentionné précédemment, il est particulièrement important de ne pas inclure de Tris dans la transcription. Le tampon Tris stimule le clivage du ribozyme *glmS*, sa présence est donc à éviter pour empêcher un clivage prématuré d'une fraction de l'ARN d'intérêt directement dans la transcription, ce qui diminuerait le rendement⁴.

Le ribozyme *glmS* n'a pas énormément de restriction quant à la séquence de l'ARN d'intérêt. Pour avoir un clivage optimal, c'est-à-dire un clivage complet et rapide avec la plus basse concentration possible de GlcN6P, une adénine doit être positionnée à l'extrémité 3', de l'ARN d'intérêt. Si un résidu autre qu'une adénine se retrouve à l'extrémité 3' il sera généralement possible de cliver cet ARN d'intérêt à plus de 95%, mais en utilisant un temps de clivage plus long et une quantité de GlcN6P plus élevée. Il est néanmoins possible d'utiliser n'importe quels résidus en 3' de l'ARN d'intérêt si cela est nécessaire⁴⁸.

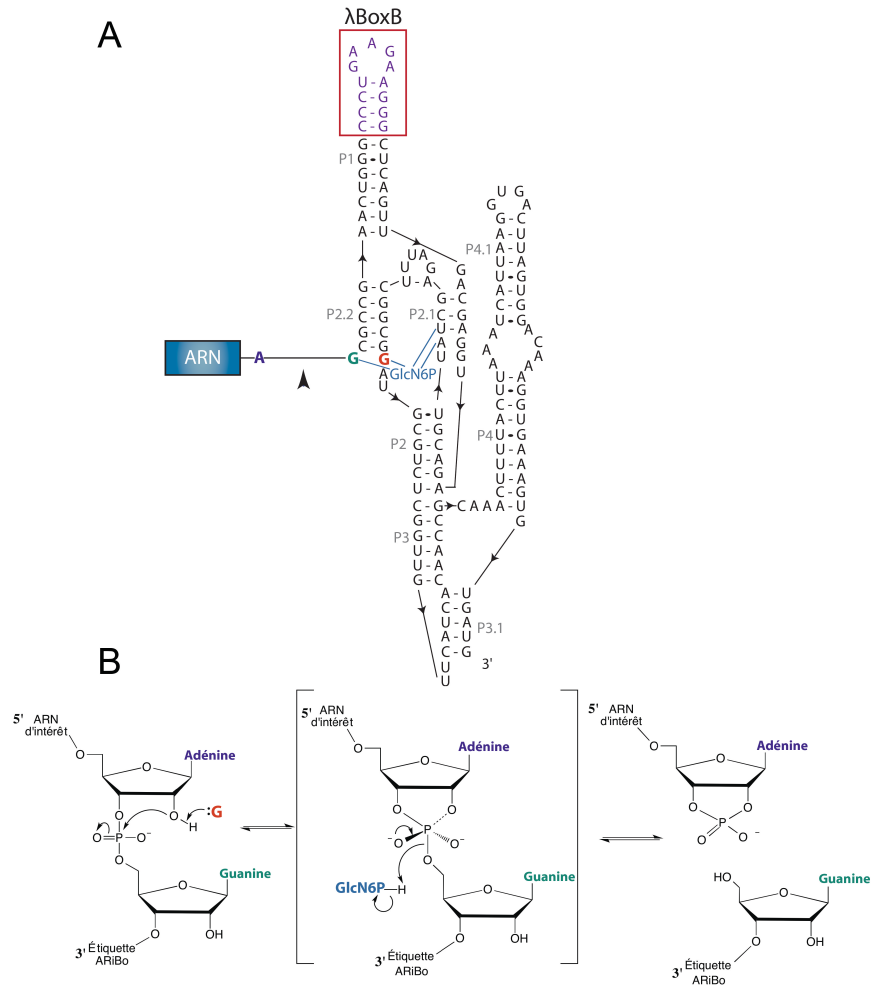


Figure 1.5 : Structure secondaire et réaction de clivage de l'étiquette ARiBo. (A) La structure secondaire de l'étiquette ARiBo est présentée avec le site de clivage à la GlcN6P indiqué avec une flèche et le domaine λ BoxB entouré d'une boîte rouge. L'étiquette ARiBo est le ribozyme *glmS* auquel quelques modifications ont été apportées, notamment, l'ajout du domaine λ BoxB. Le site de liaison à la GlcN6P est indiqué en bleu. (Adaptée de Di Tomasso et al, 2012)⁴⁸. (B) La réaction d'autoclivage propre au ribozyme *glmS* en présence de GlcN6P est illustrée, les nucléotides directement impliqués dans la réaction de clivage sont de la même couleur que sur la structure secondaire de l'étiquette ARiBo (Adaptée de Viladom, 2012)²⁹.

La protéine $\lambda N^+ - L^+ - GST$

La protéine $\lambda N^+ - L^+ - GST$ a été créée dans le cadre du développement de la méthode de purification ARiBo (Figure 1.6)⁴⁹. Elle contient deux domaines

distincts intercalés d'une suite de 10 dipeptides glycine-alanine : un domaine GST standard à son extrémité C-terminale et un petit domaine de seulement 22 acides aminés, λN^+ , à son extrémité N-terminale. Le domaine GST permet de lier la résine GSH-sepharose, ce qui rend la méthode très accessible à cause de la disponibilité et la facilité d'utilisation de la résine GSH-sepharose. Le domaine λN^+ , quant à lui, permet à la protéine de fusion de se lier à l'ARN *BoxB* de l'étiquette ARiBo avec une affinité de l'ordre du picomolaire.

Lors de la purification de la protéine de fusion, il est particulièrement important de s'assurer qu'elle soit dépourvue de RNase. Des tests peuvent être faits afin de s'assurer de la stabilité et de la pureté de la protéine en plus de s'assurer que la présence de la protéine ne dégrade pas les ARN mis en sa présence⁴⁹.



Figure 1.6 : Protéine de fusion $\lambda N^+ -L^+ -GST$. Protéine de fusion utilisée dans la purification d'ARN par affinité utilisant une étiquette ARiBo. Le schéma de la protéine λN^+ , dont la séquence d'acides aminés est indiquée sur la figure, fusionnée avec la protéine GST via une séquence répétitive de dix Gly-Ala (Adaptée de Di Tomasso et al, 2012)⁴⁷.

1.4.2 – Avantages et inconvénients de la méthode ARiBo

Comme d'autres méthodes de purification d'ARN par affinité, l'ARN obtenu avec la méthode ARiBo est récupéré sous sa forme native. Ensuite, cette méthode est simple et surtout très rapide (environ 3 h). Finalement, cette méthode a été optimisée pour obtenir des niveaux de pureté élevés (plus de 99%) et des rendements semblables à la méthode traditionnelle de purification par gel dénaturant de polyacrylamide. Le système GST/GSH-sepharose utilisé pour l'immobilisation de l'ARN lors de la purification est peu coûteux et simple d'utilisation. Cette méthode est idéale pour la purification à petite échelle de nombreux ARN en parallèle due à la rapidité et à la facilité d'exécution de la méthode une fois l'ARN de fusion synthétisé. Un autre avantage considérable de

la méthode est qu'elle pourrait être adaptée pour l'isolation de complexes ARN-protéines, ce qui représente un avantage par rapport à d'autres méthodes de purification d'ARN tel que la méthode traditionnelle par gel dénaturant de polyacrylamide ou les méthodes par chromatographie.

Malgré ces nombreux avantages, la méthode de purification d'ARN par affinité utilisant l'étiquette ARiBo présentait également quelques limites et désavantages lorsque ce projet de maîtrise a débuté. Tout d'abord, même si l'hétérogénéité en 3' est automatiquement éliminée par le clivage de l'étiquette en 3', l'hétérogénéité en 5' reste présente, ce qui est problématique pour de nombreux ARN. De plus, même si plusieurs ARN (ribozymes, riboswitch, pre-miARN, etc.) ont été purifiés avec succès grâce à la méthode ARiBo, ceci représente un éventail restreint de séquences et la méthode n'avait pas encore été testée avec de très longs ARN. Il est donc important de mieux démontrer la versatilité de la méthode et confirmer qu'elle ne fonctionne pas seulement pour quelques ARN spécifiques.

1.5 – Description de l'hétérogénéité de séquences en 5' et 3'

Dans les cas de purification par affinité, étant donné que l'étiquette ARiBo se trouve en 3' de l'ARN, l'hétérogénéité générée en 3' lors de la transcription de l'ARN de fusion se retrouve au niveau de l'étiquette. Au moment du clivage de l'étiquette, la coupure de l'ARN de fusion produit une extrémité 2'-3'-cyclique phosphate en 3' de l'ARN d'intérêt²⁵. Il n'y a donc aucune hétérogénéité de séquence au niveau de l'extrémité 3' de l'ARN d'intérêt.

L'hétérogénéité en 5' est quant à elle plus problématique notamment parce que la méthode ARiBo n'assure pas l'homogénéité en 5' comme elle assure l'homogénéité en 3'. L'hétérogénéité en 5' est problématique pour plusieurs études structurales et fonctionnelles. Par exemple, en RMN, les nucléotides supplémentaires ou manquants causent des différences significatives dans les spectres et rendent l'analyse RMN beaucoup plus complexe que si l'ARN était pur.

1.6 – Perspectives pour réduire le problème d'hétérogénéité en 5' avec la méthode ARiBo

Il existe déjà des méthodes pour éliminer l'hétérogénéité en 5'. La première consiste à choisir la séquence en 5' de l'ARN d'intérêt. Il a été démontré que la séquence de l'extrémité 5' influence la quantité d'hétérogénéité produite par la polymérase à ARN du phage T7. En effet, des études ont démontré un plus haut niveau d'hétérogénéité en 5' lorsque les ARN débutent pas GGG¹⁸ ou GAG²⁸, alors qu'une séquence débutant pas GCG¹⁸ est beaucoup plus homogène en 5'. Par contre, aucune étude systématique n'a été faite sur ce sujet jusqu'à maintenant.

Une méthode a été mise au point qui permet l'élimination de l'hétérogénéité en 5'. Il s'agit de l'ajout du ribozyme hammerhead (HH) à l'extrémité 5' qui s'autoclive à un site précis durant la transcription de l'ARN (section 1.6.1)²⁸. D'autres types d'étiquettes en 5' pourraient être utilisés, tels que l'ARN CRISPR qui peut être clivé par l'endonucléase Cse3 (section 1.6.2).

1.6.1 – Utilisation du ribozyme hammerhead pour l'élimination de l'hétérogénéité en 5'

Le ribozyme hammerhead (HH) est bien caractérisé et a déjà été utilisé pour éliminer l'hétérogénéité en 5' d'ARN transcrits *in vitro*²⁸. Étant donné que ce ribozyme est autoclivable, il est donc clivé directement durant la transcription, ce qui présente un avantage car aucune étape de clivage supplémentaire n'est nécessaire lors de la purification. La difficulté majeure liée à l'utilisation du ribozyme HH est qu'il doit être légèrement modifié afin d'être compatible avec l'ARN à purifier (Figure 1.7). Comme le clivage du ribozyme HH requiert la formation d'une tige très stable avec l'ARN d'intérêt, l'extrémité 5' du ribozyme doit être adaptée pour s'apparier à l'extrémité 5' de l'ARN d'intérêt sans restriction quant à la nature des résidus. Il peut être ardu d'adapter le ribozyme HH pour obtenir un autoclivage complet. Par exemple, si la structure de l'ARN désiré est trop stable, son extrémité 3' ne formera pas, avec le ribozyme, la tige nécessaire au clivage de cet ARN. De plus, des conditions non dénaturantes

peuvent empêcher la dissociation du ribozyme HH de l'ARN d'intérêt et ce malgré un clivage complet. Donc, il peut être difficile, dans certains cas, de se débarrasser du ribozyme HH en conditions natives.

Jumelé à l'étiquette ARiBo, l'ajout d'un ribozyme hammerhead ne devrait rien changer au protocole de purification par affinité. Aucune étape de clivage supplémentaire ne serait nécessaire et le ribozyme HH clivé pourrait être automatiquement éliminé lors des lavages avant l'éluion de l'ARN d'intérêt. Le ribozyme hammerhead est déjà utilisé pour éliminer l'hétérogénéité en 5', par contre, sa compatibilité avec la méthode ARiBo n'a jamais été testée.

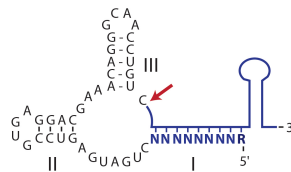


Figure 1.7 : Structure secondaire du ribozyme hammerhead. Le ribozyme HH doit être adapté pour chaque ARN avec lequel il est couplé. La suite de N en 5' peut varier en longueur, mais doit être complémentaire à l'extrémité 5' de l'ARN choisi et débiter obligatoirement par une purine pour permettre la transcription avec polymérase à ARN du phage T7. La flèche indique le site de l'autoclivage du ribozyme hammerhead (Adaptée de Di Tomasso et al, 2014).

1.6.2 – Utilisation potentielle du CRISPR comme étiquette à l'extrémité 5' d'ARN

Dans le but d'éradiquer complètement l'hétérogénéité en 5' des ARN purifiés par affinité avec l'étiquette ARiBo, une étiquette autre que le ribozyme hammerhead pourrait être envisagée telle que l'ARN non-codant CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeat*) qui se retrouve dans les bactéries et sert de moyen de défense contre les phages⁵⁰ (Figure 1.8). Le CRISPR est formé de plusieurs tiges-boucles identiques de 28 nt. Ces tiges-boucles sont intercalées de courtes séquences non répétitives dans lesquelles on retrouve parfois des séquences d'ADN dérivées de celles des phages.

Dans le génome bactérien, en amont du CRISPR on retrouve le gène CAS, codant pour plusieurs protéines Cas. Bien que le mécanisme exact ne soit pas

encore bien compris, il est connu que lorsqu'une bactérie subit une attaque par un phage, et si cette bactérie possède dans son génome un dérivé de l'ADN du même phage, la transcription du gène CAS est activée⁵⁰. Les tiges-boucles CRISPR sont clivées par une protéine Cas, produisant des fragments de CRISPR, soit de petites séquences non répétitives encadrées de deux fragments dérivés des tiges-boucles CRISPR (Figure 1.8). De ces petits fragments dérivés du clivage du CRISPR, ceux étant complémentaires à l'ARNm ou à l'ADN du phage, lié avec un complexe protéique incluant la protéine Cas, reconnaissent la séquence du phage, ce qui mènera à sa dégradation. Ainsi, le phage ne proliférera pas dans la bactérie, ce qui empêchera l'infection au sein de celle-ci.

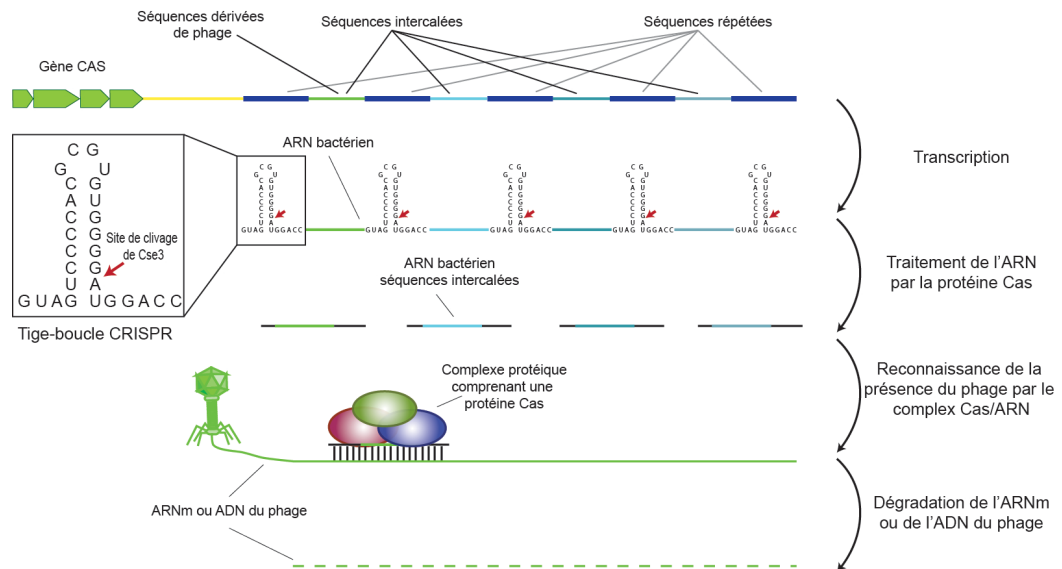


Figure 1.8 : Mode de fonctionnement du CRISPR. La structure et la séquence du CRISPR sont illustrées ainsi que l'utilisation que la bactérie en fait pour se défendre contre les phages. La séquence est clivée par une protéine Cas, comme par exemple la protéine Cse3 de *Thermus thermophilus*. Les séquences qui intercalent les petites tiges-boucles répétées du CRISPR sont dérivées de phages, ce qui permet à la bactérie de reconnaître l'attaque de phages grâce à une protéine Cas et ainsi produire de petits ARN contenant deux fragments provenant des tiges-boucles aux extrémités de la séquence intercalée semblable à celle du phage. En s'associant sous forme de complexe avec la protéine Cas, ces petits ARN guident la dégradation des acides nucléiques du phage (Adaptée de Sorek et al, 2008 et de Gesner et al, 2011)^{50,51}.

Cse3 fait partie des protéines Cas de *Thermus thermophilus*. Cette protéine clive la tige-boucle du CRISPR après son 21^e nucléotide (Figure 1.8). Une étude a démontré que la séquence suivant le site de clivage de Cse3 (soit après le 21^e nt) pouvait être modifiée sans que le clivage ne soit trop affecté, ce qui inclut les mutations A22U, U23A et G24C⁵². Cette étude suggère qu'après le 21^e nt seulement les résidus 22, 23 et 24, peu importe leur nature, sont nécessaires au clivage du CRISPR par Cse3. Par contre, aucune étude n'a été effectuée pour tester systématiquement l'importance de l'identité de chacun des nucléotides après le 21^e nucléotide du CRISPR. Le système CRISPR/Cse3 pourrait donc être un système intéressant, car, si les nucléotides en 3' du site de clivage ne sont pas essentiels, seuls les 21 premiers nucléotides de sa séquence pourraient être utilisés comme étiquette. Le clivage de cette étiquette CRISPR pourrait se faire par l'endonucléase Cse3 directement après ce 21^e nucléotide et donc ne laisser aucune séquence résiduelle à l'extrémité 5' de l'ARN d'intérêt, ce qui permettrait l'élimination de l'hétérogénéité en 5'⁵².

1.6 – Hypothèses et Objectifs

Un des objectifs du laboratoire est de poursuivre le développement de la méthode ARiBo pour permettre l'obtention d'ARN homogènes en 5' et en 3', de purifier un plus grand éventail d'ARN et d'isoler des complexes ribonucléoprotéiques. L'objectif de mon mémoire est de contribuer au développement de la méthode ARiBo en poursuivant les trois objectifs spécifiques décrits ci-dessous.

Élimination de l'hétérogénéité en 5'

Le premier objectif est d'éradiquer l'hétérogénéité en 5' d'ARN transcrits *in vitro* par la polymérase à ARN du phage T7. Deux hypothèses seraient testées pour atteindre ce premier objectif.

Notre première hypothèse est que l'hétérogénéité en 5' dépend de la nature de la séquence à l'extrémité 5' de l'ARN. Je voulais donc identifier toutes les combinaisons de nucléotides en 5' produisant des ARN homogènes. Le but est de

pouvoir sélectionner la séquence en 5' de l'ARN et, sans avoir à recourir à d'autres méthodes, permettre la synthèse d'un ARN parfaitement homogène en 5'. Deux polymérases à ARN du phage T7 seront testées, celle de type sauvage et le mutant P266L. Nous pensons que le mutant P266L pourrait aider à diminuer l'hétérogénéité en 5', étant donné qu'il aide à réduire les avortons lors de la transcription¹⁹. Une difficulté de ce premier objectif est d'identifier la provenance exacte de l'hétérogénéité, soit de démontrer d'où provient l'hétérogénéité dans la séquence, de l'extrémité 5' ou du reste de la séquence. Pour élucider la provenance de l'hétérogénéité, l'ARN SLI, un substrat non-naturel du ribozyme VS, a été choisi afin qu'il puisse être clivé après son 5^e nucléotide en 5' par le ribozyme VS. Ainsi, il sera possible de comparer l'extrémité 5' et le reste de l'ARN et d'observer d'où provient l'hétérogénéité.

Ensuite, pour les expériences nécessitant une séquence particulière, nous voulons mettre au point une autre stratégie qui incorpore l'ajout d'une étiquette en 5' de l'ARN. Notre deuxième hypothèse est que l'hétérogénéité en 5' peut être éliminée par l'utilisation d'une étiquette d'ARN clivable en 5' de l'ARN d'intérêt. Le plus grand défi avec l'ajout d'une seconde étiquette est que le clivage de cette étiquette doit être effectué à 100% avant l'élution sinon l'étiquette non-clivée se retrouvera dans l'élution créant ainsi des impuretés. Nous pensons qu'en plus de l'élimination de l'hétérogénéité en 5', l'ajout de cette étiquette permettra de faire débiter la séquence de l'ARN d'intérêt par n'importe quel résidu (soit un A, un C, un G ou un U) et non obligatoirement par une purine. L'ARN choisi pour servir d'étiquette en 5' est la séquence du CRISPR clivable en présence de l'endonucléase Cse3. Cette première étude est détaillée dans l'article du Chapitre 2.

Purification d'un ARN de plus de 600 nt

Mon second objectif est de permettre la purification d'ARN plus longs que ceux purifiés jusqu'à maintenant avec la méthode ARiBo. Nous croyons que la méthode de purification ARiBo pourrait permettre de purifier des longs ARN de

plus de 600 nt. J'ai donc choisi de purifier miR106b-25, un polycistron de miARN d'une longueur de 614 nt. Des études préliminaires avec un ARN plus court que miR-106b-25, mais plus long que tous les autres ARN purifiés jusqu'à maintenant (le ribozyme VS d'une longueur de 138 nt) a démontré que d'utiliser du sel lors de l'élution de l'ARN permet d'obtenir de meilleurs rendements. Ces études préliminaires laissent présager que des modifications, telles que l'ajout de sel, devront être apportées au protocole original de la méthode ARiBo afin de purifier miR-106b-25. Les résultats de cette étude sont présentés dans la discussion de ce mémoire et ont servi à supporter l'élaboration du protocole présenté dans l'article de Di Tomasso *et al.* 2014⁵³.

Isolation de complexes ARN-protéines

Enfin, nous pensons que la méthode de purification avec une étiquette ARiBo pourrait être adaptée pour isoler des protéines associées à un ARN donné. Des extraits cellulaires seront utilisés pour repêcher des protéines se liant à l'ARN immobilisé par une étiquette ARiBo qu'il sera ensuite possible d'identifier par immunobuvardage de type Western et par spectrométrie de masse. Dans le cadre du développement de cette méthode, j'ai effectué des études préliminaires avec le précurseur de miARN, pre-miR153-2 qui sont présentés dans la discussion de ce mémoire.

Ainsi, atteindre mes trois objectifs contribuerait à démontrer la compatibilité de la méthode ARiBo avec l'obtention d'ARN homogènes en 5' et en 3', avec la purification d'un long ARN et avec l'isolation de complexes ARN-protéines.

**Chapitre 2 – Élimination d'hétérogénéité d'ARN
transcrits avec la polymérase à ARN du phage T7 et
purifiés par affinité en utilisant les étiquettes ARiBo et
CRISPR**

2.1 – Mise en contexte de l'article

La production d'ARN est une étape très importante dans de nombreuses sphères de recherche impliquant de l'ARN. La méthode de purification par affinité d'ARN utilisant une étiquette ARiBo développée par le laboratoire Legault permet d'obtenir un ARN très pur sous sa forme native en un très court laps de temps. Par contre, des erreurs de lecture de la polymérase à ARN du phage T7 utilisée pour la transcription *in vitro* de l'ARN occasionne une hétérogénéité de séquence en 5' qui n'est pas idéale pour plusieurs types d'études. Cette hétérogénéité n'est pas toujours évidente à observer, car sur un gel de polyacrylamide, pour des gros ARN, la séparation entre une séquence N et N+1 ne se voit pas bien. Ainsi, nous avons choisi comme modèle un petit ARN de 29 nucléotides, SLI, le substrat non-naturel du ribozyme VS de *Neurospora*. Nous avons choisi SLI comme modèle non seulement pour sa petite taille, mais également parce qu'il s'agit d'une tige-boucle très stable qui est facilement purifiable avec l'étiquette ARiBo. De plus, il peut facilement être muté et il peut être clivé après son cinquième nucléotide en 5' par le ribozyme VS. SLI a donc été mutées en 32 différentes séquences débutant soit par un A soit par un G et dont les deuxièmes et troisièmes paires de bases de la tige boucle ont été muté en 16 différentes séquences. Ces 32 séquences ont été transcrites par deux différentes polymérases à ARN du phage T7, la polymérase de type sauvage et la polymérase avec la mutation P266L. Les ARN obtenus ont été purifiés avec l'étiquette ARiBo et l'hétérogénéité a pu être quantifié. Cet article propose des séquences à utiliser telles quelles, c'est-à-dire que naturellement ces séquences ne présentent peu ou pas d'hétérogénéité en 5'. Dans les cas où un ARN présente trop d'hétérogénéité pour être utilisé, nous proposons une approche utilisant une seconde étiquette clivable, ajoutée en 5' de l'ARN, le CRISPR. Cette nouvelle approche est simple, facile à utiliser et permet également d'obtenir un ARN homogène.

Contribution des auteurs à l'article :

Affinity purification of T7 RNA transcripts with homogeneous ends using ARiBo and CRISPR tags

Salvail-Lacoste A, Di Tomasso G, Piette B. L, Legault P, *RNA*, 2013, May; **19**, 1003-1014, doi:10.1261/rna.037432.112.

Alix Salvail-Lacoste : a réalisé, quantifié et analysé les expériences sur les séquences transcrites avec le promoteur de classe III, quantifié et analysé les expériences sur les séquences transcrites avec le promoteur de classe II ϕ 2.5 et participé à la rédaction de l'article. (Tableaux I et II, Figures 2.1, 2.2 et S2.2)

Geneviève Di Tomasso : a réalisé, quantifié et analysé les expériences avec l'étiquette CRISPR et participé à la rédaction de l'article. (Figures 2.4, 2.5, S2.1, S2.3, S2.4, S2.5 et S2.6)

Benjamin L Piette : a réalisé, quantifié et analysé les expériences sur les séquences transcrites avec le promoteur de classe II ϕ 2.5 et participé à la rédaction de l'article. (Tableau II et Figure 2.3)

Pascale Legault : a conçu et dirigé le projet, participé à l'analyse des résultats et rédigé l'article.

2.2 – Article

Affinity purification of T7 RNA transcripts with homogeneous ends using ARiBo and CRISPR tags

Alix Salvail-Lacoste¹, Geneviève Di Tomasso¹, Benjamin L. Piette and Pascale Legault

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

¹These authors contributed equally to this work.

Running head: Affinity purification of RNA with homogeneous ends

Keywords: Affinity purification of RNA; ARiBo tag; CRISPR tag; T7 RNA polymerase; 5'-heterogeneity; Cse3 endoribonuclease;

Abstract

Affinity purification of RNA using the ARiBo tag technology currently provides an ideal approach to quickly prepare RNA with 3'-homogeneity. Here, we explored strategies to also ensure 5'-homogeneity of affinity-purified RNAs. First, we systematically investigated the effect of starting nucleotides on the 5'-heterogeneity of a small SLI RNA substrate from the *Neurospora* VS ribozyme purified from an SLI-ARiBo precursor. A series of 32 SLI RNA sequences with variations in the +1 to +3 region were produced from two T7 promoters (class III consensus and class II ϕ 2.5) using either the wild-type T7 RNA polymerase or the P266L mutant. Although the P266L mutant helps decrease the levels of 5'-sequence heterogeneity in several cases, significant levels of 5'-heterogeneity ($\geq 1.5\%$) remain for transcripts starting with GGG, GAG, GCG, GGC, AGG, AGA, AAA, ACA, AUA, AAC, ACC, AUC and AAU. To provide a more general approach to purifying RNA with 5'-homogeneity, we tested the suitability of using a small CRISPR RNA stem-loop at the 5'-end of the SLI-ARiBo RNA. Interestingly, we found that complete cleavage of the 5'-CRISPR tag with the Cse3 endoribonuclease can be achieved quickly from CRISPR-SLI-ARiBo transcripts. With this procedure, it is possible to generate SLI-ARiBo RNAs starting with any of the four standard nucleotides (G, C, A or U) involved in either a single- or a double-stranded structure. Moreover, the 5'-CRISPR-based strategy can be combined with affinity purification using the 3'-ARiBo tag for quick purification of RNA with both 5'- and 3'-homogeneity.

Introduction

In vitro synthesis of RNA has become an indispensable tool for biological sciences. This is most commonly achieved via transcription with bacteriophage RNA polymerases like T7, T3 and SP6 (Chamberlin and Ryan 1982; Melton et al. 1984; Milligan et al. 1987; Krupp 1988), which allow specific promoter-dependent transcription of diverse RNA sequences. In vitro transcription with the T7 RNA polymerase is particularly widespread since it can produce milligram quantities of RNAs for a variety of in vitro and in vivo applications.

Although the T7 RNA polymerase (T7 RNAP) accurately transcribes most sequences starting with a purine (Dunn and Studier 1983; Milligan et al. 1987; Imburgio et al. 2000), it often produces transcripts with heterogeneous 5' and 3' ends. The T7 RNAP uses an 18-nucleotide (nt) double-stranded promoter to initiate transcription and then transcribes a single- or double-stranded DNA template until the end of the complementary DNA strand is reached (Milligan et al. 1987; Krupp 1988). The T7 RNAP may terminate prematurely, but often adds a few extra bases at the 3'-end that are not encoded by the template (Milligan et al. 1987; Krupp 1988; Martin et al. 1988) and, in certain cases, may produce RNAs that are much longer than expected (Cazenave and Uhlenbeck 1994). In addition, transcriptions from the G-initiating T7 class III consensus promoter [residues -17 to +1; (Chamberlin and Ryan 1982)] can yield 5'-sequence heterogeneity, particularly for sequences starting with GGG (Martin et al. 1988; Cunningham et al. 1991; Pleiss et al. 1998; Imburgio et al. 2000; Sherlin et al. 2001) and GAG (Ferré-D'Amaré and Doudna 1996). 5'-sequence heterogeneity is also observed when the +1 G nucleotide is substituted by A, U or C (Cunningham et al. 1991; Helm et al. 1999; Imburgio et al. 2000). It was reported that RNA sequences transcribed from the A-initiating T7 class II 2.5 promoter have superior 5'-sequence homogeneity (Coleman et al. 2004). However this promoter has not been extensively exploited for in vitro synthesis of RNA (Milligan et al. 1987; Cunningham et al. 1991; Imburgio et al. 2000).

To help produce RNA with homogeneous ends, one may rely on existing solutions. Incorporation of 2'-methoxyls in the first two 5'-nucleotides of the complementary DNA strand can help reduce 3'-heterogeneity (Kao et al. 1999; Sherlin et al. 2001). Furthermore, purification by denaturing gel electrophoresis can be used for size fractionation of RNAs with nucleotide resolution (Milligan et al. 1987; Wyatt et al. 1991; Doudna 1997). However, this approach is only useful for small RNAs (< 30-50 bases) and will likely not separate RNAs of similar size with both 5'- and 3'-heterogeneity. Endoribonuclease processing of an RNA precursor represents a more general solution to producing an RNA target with homogeneous ends. Many endonucleases with minimal sequence requirements

have been exploited, including RNase H (Lapham and Crothers 1996), DNazymes (Santoro and Joyce 1997) and ribozymes (Dzianott and Bujarski 1988; Grosshans and Cech 1991; Taira et al. 1991; Price et al. 1995; Ferré-D'Amaré and Doudna 1996; Schürer et al. 2002; Walker et al. 2003; Kieft and Batey 2004; Batey and Kieft 2007; Pereira et al. 2010; Nelissen et al. 2012). Ribozymes are particularly convenient because they can be added in cis or co-transcribed to achieve RNA cleavage during transcription. For example, the well-characterized hammerhead ribozyme has been utilized for 5' and 3' cleavage of RNA transcripts (Dzianott and Bujarski 1988; Grosshans and Cech 1991; Taira et al. 1991; Price et al. 1995; Ferré-D'Amaré and Doudna 1996; Walker et al. 2003). However, its cleavage requires base pairing between ribozyme residues and the RNA target, such that it may be incomplete in cases where the RNA target structure competes with the folding of the active ribozyme (Grosshans and Cech 1991; Price et al. 1995; Walker et al. 2003). Moreover, it may be cumbersome in such cases to find ideal ribozyme sequence and cleavage conditions (Walker et al. 2003). Thus, endoribonucleases that do not rely on structural reorganization of the target RNA present a distinct advantage for trimming the ends of RNA transcripts. For example, the VS, HDV and *glmS* ribozymes have little sequence or structure requirements 5' of the cleavage site and can be incorporated at the 3'-end of an RNA to ensure its 3'-sequence homogeneity (Ferré-D'Amaré and Doudna 1996; Kieft and Batey 2004; Batey and Kieft 2007; Di Tomasso et al. 2011). We are unaware of an equivalent ribozyme (or DNzyme) with minimal sequence and structure requirements 3' of the cleavage site that has been used to achieve 5'-sequence homogeneity. RNase H cleavage can fulfill these criteria, but it requires a long incubation with chemically modified oligonucleotides that makes it less practical (Lapham and Crothers 1996).

Several recently discovered *Cas endoribonucleases* display minimal sequence requirement 3' of the cleavage site and have potential as biochemical tools for generating RNA with 5'-homogeneity (Brouns et al. 2008; Haurwitz et al. 2010; Gesner et al. 2011; Sashital et al. 2011; Garside et al. 2012). *Cas endoribonucleases* specifically cleave at cognate CRISPR (Clustered regularly

interspaced short palindromic repeats) RNA sequences found in prokaryotes as part of an adaptive immune system against bacteriophages and plasmids [for recent reviews see (Horvath and Barrangou 2010; Karginov and Hannon 2010; Marraffini and Sontheimer 2010; Al-Attar et al. 2011; Makarova et al. 2011; Terns and Terns 2011; Wiedenheft et al. 2012)]. The CRISPR RNA sequences often comprise small stable RNA hairpins that are specifically cleaved at the 3'-end of the stem by the CRISPR-associated (Cas) endoribonuclease. For example, Cse3, the Cas endoribonuclease from *Thermus thermophilus* (*T. thermophilus*), binds a 21-nt RNA hairpin and cleaves directly after G21 (Brouns et al. 2008; Gesner et al. 2011; Sashital et al. 2011).

Recently, we reported a quick procedure for affinity batch purification of in vitro transcribed RNA using a 3'-ARiBo tag (Di Tomasso et al. 2011; Di Tomasso et al. 2012a) comprised of an Activatable Ribozyme, the *glmS* ribozyme, and the *BoxB* RNA. The *boxB* RNA allows immobilization on glutathione-Sepharose (GSH-Sepharose) resin via a N-GST fusion protein, whereas the *glmS* ribozyme is activated with glucosamine-6-phosphate (GlcN6P) for RNA elution. This procedure yields RNA of high purity and ensures 3'-sequence homogeneity. However, it does not guarantee 5'-sequence homogeneity, which is important for several applications, particularly when the identity of the starting residue is critical. To ensure 5'-homogeneity of affinity-purified RNAs, we investigated the effect of starting nucleotides on the 5'-heterogeneity of a small SLI RNA substrate derived from the *Neurospora* VS ribozyme and purified by affinity from an SLI-ARiBo precursor. We also examined the potential of incorporating a small CRISPR tag at the 5'-end of SLI-ARiBo RNAs to achieve 5'-homogeneity using the *T. thermophilus* Cse3 endoribonuclease. As detailed below, RNA sequence selection can help ensure 5'-homogeneity, but the use of the Cse3/CRISPR system provides a more general approach to purifying RNA with 5'-homogeneity.

Results

Evidence for 5'-heterogeneity of an affinity-purified T7 transcript starting with GAG

As part of our structural and functional studies of the *Neurospora* VS ribozyme, we have prepared several variants of its SLI RNA hairpin substrate (Hoffmann et al. 2003; Bouchard et al. 2008; Lacroix-Labonte et al. 2012) by in vitro transcription with T7 RNAP followed by purification with gel electrophoresis. Given that gel purification is a tedious and long procedure, we were interested in speeding up the process using batch affinity purification with an ARiBo tag (Di Tomasso et al. 2011; Di Tomasso et al. 2012a). The first SLI RNA tested was SLI(2) (Fig. 2.1A), which starts with the GAG sequence. SLI(2) was synthesized with a 3'-ARiBo tag and affinity purified on a λ N⁺-L⁺-GST/GSH-Sepharose matrix (Fig. 2.1B). The ARiBo tag sequence used here (ARiBo4; Supplementary Fig. S2.1) is slightly different from our original ARiBo1 tag (Di Tomasso et al. 2011) in that the *glmS* ribozyme sequence was modified to help reduce misfolding and improve GlcN6P-induced cleavage of ARiBo-fusion RNAs. Using this ARiBo tag, the SLI RNA of the correct size was successfully detected in the elution fractions. Analysis by denaturing gel electrophoresis demonstrated that this SLI RNA is highly pure with respect to the most likely contaminant, the ARiBo tag. However, there are a few low intensity bands close to the predominant (most intense) band on the gel (lane E1 on Fig. 2.1B) that represents around 15% heterogeneity within the population of purified SLI RNA. This level of heterogeneity is too high for biophysical and structural investigations, thus we sought to investigate the effect of starting nucleotides on the heterogeneity of T7 transcripts.

5'-heterogeneity of T7 transcripts synthesized from the G-initiating T7 class III promoter

To investigate the effect of the first nucleotides on the 5'-heterogeneity of T7 transcripts, we affinity purified a given RNA in which we systematically modified the first few nucleotides. The SLI hairpin substrate derived from the VS

ribozyme serves as an ideal model for these studies (Fig. 2.1A), because its small size (29 nt) facilitates detection of 5'- and 3'- heterogeneities on denaturing gels. Furthermore, SLI can be cleaved by the VS ribozyme to yield products (Fig. 2.1A) that can be analyzed on denaturing gels to pinpoint the source (e.g. 5' or 3') of heterogeneity.

In addition to the SLI(2) RNA that starts with GAG, other SLI RNA hairpins with sequence variations of nucleotides +2/+3 [SLI(1) to SLI(16); Fig. 2.2A] were synthesized with a 3'-ARiBo tag from the G-initiating T7 class III promoter using the wild-type T7 RNAP. These SLI RNAs also incorporate sequence variations of nucleotides +26/+27 to maintain Watson-Crick base pairing in stem Ia (Figs. 2.1A and 2.2A). The SLI RNAs were affinity purified, and the eluted fractions were separated on denaturing gels (Fig. 2.2B). The gel conditions are not fully denaturing because gels were run at low voltage to increase resolution, thus the mobility of the predominant band varies from one SLI sequence to the other; those with a higher content of G-C base pairs in stem Ia systematically migrate faster, as previously observed (Lehrach et al. 1977; Frank and Köster 1979). For some SLI RNAs, there are, in addition to the most intense band, lower intensity bands indicative of transcriptional heterogeneity (see for example lanes 1 and 2 in Fig. 2.2B). This heterogeneity likely originates from the 5'-end, since *glmS* ribozyme cleavage should ensure 3'-homogeneity of the purified RNA. Nevertheless, we confirmed the absence of 3'-heterogeneity by performing VS ribozyme cleavage of the 16 SLI RNAs prior to elution. The VS ribozyme cleaves SLI RNAs between G5 and A6, such that a short 5'-product is released during the VS cleavage step and a 24-nt 3'-product is eluted along with the 29-nt substrate following activation of *glmS* cleavage. In contrast, all 24-nt 3'-products migrate as single homogeneous bands on a denaturing gel (Supplementary Fig. S2.2).

To verify that the observed heterogeneity is not due to a partially dephosphorylated 5'-triphosphate group, affinity purifications were also performed in which the E1 elution fractions were treated with calf alkaline phosphatase. For most SLI RNAs, this phosphatase treatment reduced the number of low intensity

bands on the gel (compare Fig. 2.2B with Fig. 2.2C). Interestingly, after alkaline phosphatase treatment very few RNAs display low intensity bands.

Affinity purifications with alkaline phosphatase treatment were performed at least three times to determine average percentages of heterogeneity with experimental errors (Table I). As previously reported (Pleiss et al. 1998; Coleman et al. 2004), a high percentage of heterogeneity is detected for transcripts starting with the GGG and GAG sequences, respectively 45% and 17%. We also find lower but significant levels of 5'-sequence heterogeneity (1.5-3%) for SLI RNAs starting with the GCG, GGC and GCC sequences. However, no significant heterogeneity (<1.5%) is observed for SLI RNAs (11 out of 16) starting with GUG, GGA, GAA, GCA, GUA, GAC, GUC, GGU, GAU, GCU and GUU.

The T7 RNAP mutant P266L is known to greatly reduce the relative levels of abortive transcripts in favor of run off transcription (Guillerez et al. 2005; Ramírez-Tapia and Martin 2012), and we reasoned that it could also help reduce levels of 5'-heterogeneity. Thus, SLI-ARiBo RNAs were also synthesized using the P266L mutant, and the SLI RNAs were affinity purified with alkaline phosphatase treatment. For these G-initiating transcripts, the P266L mutant produces similar levels of 5'-sequence heterogeneity compared to the wild-type polymerase (Fig. 2.2D and Table I), with a significant reduction only for the SLI(11) RNA that starts with GCC (from $2.2 \pm 0.6\%$ to $0.7 \pm 0.5\%$).

5'-heterogeneity of T7 transcripts synthesized from the A-initiating class II ϕ 2.5 promoter

The T7 class II 2.5 promoter allows synthesis of RNA sequences that start with an adenine. It has not been widely used, most likely because T7 class II promoters are considered weaker promoters both in vitro and in vivo [(Chamberlin and Ryan 1982; Dunn and Studier 1983; Milligan et al. 1987; Imburgio et al. 2000) and references therein]. However, it was shown that similar RNA yields can be obtained in vitro when using the T7 class II 2.5 and class III promoters, and that the class II 2.5 promoter can produce RNA with significantly less 5'-heterogeneity (Coleman et al. 2004). Given that only a few sequences were tested, we were

interested to systematically investigate 5'-sequence heterogeneity originating from this promoter.

SLI RNA hairpins with sequence variations of the +2/+3 nucleotides [SLI(17) to SLI(32); Fig. 2.3A] were synthesized with a 3'-ARiBo tag from the A-initiating T7 class II 2.5 promoter using the wild-type T7 RNAP. After affinity purification, the eluted fractions were treated with alkaline phosphatase and analyzed by denaturing gels (Fig. 2.3B and Table II). In contrast to our observations with the class III promoter, many RNAs synthesized from the class II 2.5 promoter display a high percentage ($\geq 5\%$) of 5'-sequence heterogeneity, namely those starting with AGG, AGA, AAA, ACA, AUA, AAC, ACC, AUC and AAU. The percentage of heterogeneity of SLI(17) starting with AGG ($11 \pm 2\%$) is somewhat higher but similar to what has been previously reported for transcripts starting with AGG [$6 \pm 2\%$; (Coleman et al. 2004)]. Consistent with a prior study (Cunningham et al. 1991), transcripts starting with AAA ($84 \pm 2\%$) and AAC ($68 \pm 1\%$) are so heterogeneous that the most intense RNA band does not represent the expected RNA product (see below), contrary to what is generally assumed. Notably, the highest percentages of heterogeneity obtained for the class II 2.5 and class III promoters are with the AAA- and GGG-initiating transcripts, respectively, suggesting that this heterogeneity arises from similar mechanisms, likely slippage within an unstable initiation complex and/or priming of abortive initiation products (Martin et al. 1988; Cunningham et al. 1991; Moroney and Piccirilli 1991; Pleiss et al. 1998). Lower but still significant levels of heterogeneity (1.5-3%) are observed for SLI transcripts starting with AUG and AGU, whereas no significant heterogeneity ($< 1.5\%$) is observed for those starting with AAG, ACG, AGC, ACU and AUU. In contrast to a prior study (Coleman et al. 2004), our systematic investigation demonstrates that the class III promoter generally produces transcripts with superior 5'-homogeneity compared with the class II 2.5 promoter when using the wild-type T7 RNAP.

Although the P266L mutant has little effect on 5'-sequence heterogeneity for transcripts synthesized from the class III promoter, it does have a significant effect on several transcripts synthesized from the class II 2.5 promoter (Fig. 2.3C

and Table II). Significant reductions in 5'-heterogeneity are observed with P266L for transcripts starting with AGG (from $11 \pm 2\%$ to $3.3 \pm 0.5\%$), AUG (from $2.3 \pm 0.4\%$ to $0.4 \pm 0.3\%$), AGA (from $7 \pm 1\%$ to $4.4 \pm 0.5\%$), AAC (from $68 \pm 1\%$ to $50 \pm 1\%$), AUC (from $12 \pm 1\%$ to $9 \pm 1\%$), AGU (from $2.1 \pm 0.9\%$ to $0.3 \pm 0.2\%$) and AAU (from $15 \pm 2\%$ to $6 \pm 2\%$). In contrast, a significant increase in heterogeneity is observed for the transcript starting with AUA (from $9 \pm 1\%$ to $18 \pm 1\%$, assuming that the most intense band represent the accurately-transcribed RNA). Nevertheless, the P266L mutant generally helps reduce the level of 5'-heterogeneity for A-initiating transcripts. Compared to the wild-type T7 RNAP, it reduces the level of 5'-heterogeneity to insignificant levels for transcripts starting with AUG and AGU, such that it increases the pool of A-initiating transcripts that can be synthesized with 5'-homogeneity from 5 to 7 out of the 16 tested sequences, namely those starting with AAG, ACG, AUG, AGC, AGU, ACU and AUU.

Ensuring 5'-homogeneity by Cse3 cleavage of a CRISPR RNA tag

With the goal of developing a reliable system for ensuring 5'-homogeneity of RNA, we investigated cleavage of CRISPR RNA tags by the *T. thermophilus* Cse3 endoribonuclease. First, a CRISPR-SLI(2)-ARiBo double-fusion RNA was synthesized with a 21-nt CRISPR RNA stem-loop tag at the 5'-end of the RNA of interest [here SLI(2)-ARiBo; Fig. 2.4A]. The CRISPR RNA tag contains a mutant 5'-GCAGA tail sequence to allow multiple turnover kinetics and the AUG linker sequence 3' of the cleavage site (Sashital et al. 2011). According to previous studies, such CRISPR RNA tag should be quickly cleaved by Cse3 (Sashital et al. 2011). Not surprisingly, complete cleavage ($\geq 99\%$) is obtained in as little as 5 min when incubating 1 μ M CRISPR-SLI(2)-ARiBo double-fusion RNA with 1-2 μ M Cse3 at the optimum enzyme temperature of 70°C (Fig. 2.4B). Interestingly, similar results are obtained at 37°C, a lower temperature that helps reduce non-enzymatic RNA hydrolysis (Fig. 2.4B).

A variant of this original CRISPR sequence (Fig. 2.4A) was designed to eliminate the 3'AUG linker sequence and allow production of the SLI(2)-ARiBo RNA, without extra nucleotides at its 5'-end. As expected, the cleavage efficiency

of this "no linker" variant is significantly reduced, particularly at 37°C where only 25% cleavage is reached using 2 μM Cse3 for 15 min. Nevertheless, complete cleavage ($\geq 99\%$) is obtained when incubating the CRISPR-SLI(2)-ARiBo RNA for 15 min at 70°C with 1-2 μM Cse3 (Fig. 2.4B).

To further investigate the CRISPR requirements for Cse3 cleavage, several other variants of the CRISPR-SLI(2)-ARiBo RNA were tested for Cse3 cleavage at 37°C (Supplementary Fig. S2.3) and 70°C (Fig. 2.4). First, variants similar to the "no linker" variant were tested that contain different Watson-Crick base pairs involving the 5'-nucleotide of SLI (no linker GC→AU, no linker GCCG, no linker GCUA; Fig. 2.4A). Cse3 cleavage of these variants was tested at 37°C with 1-2 μM Cse3 for 15-30 min, but only the SLI starting with an A (no linker GC→AU) is completely cleaved under these conditions (Supplementary Fig S2.3B). Interestingly, the SLI RNA starting with G (no linker), A (no linker GC→AU) and C (no linker GC→CG) are all completely cleaved when incubated at 70°C with 2 μM Cse3 for 15 min (Fig. 2.4C). For the SLI RNA starting with a U (no linker GC→UA), the 15-min incubation at 70°C provides only 95% cleavage (Fig. 2.4C). To improve cleavage of this variant (no linker GC→UA), we transcribed it again with a reversion of the 5'-GCAGA tail to the wild-type 5'-GUAGU sequence, assuming that it may help destabilize an interaction between the A5 of the CRISPR and the U1 of SLI that likely inhibits cleavage (Sashital et al. 2011). As anticipated, this variant (no linker GC→UA with 5'-GUAGU tail) is a better substrate for Cse3 than the parental construct (no linker GC→UA) and yields complete cleavage when incubated at 70°C with 2 M Cse3 for 15 min (Supplementary Fig. S2.4). Thus, complete Cse3 cleavage ($\geq 99\%$) can be quickly obtained between a 5'-CRISPR tag and the desired RNA in which the first nucleotide is involved in any of the four standard Watson-Crick base pair.

Next, we tested the effect of short linkers on Cse3 cleavage efficiency. Single-nucleotide linkers, containing A, C, G or U, do not allow complete cleavage of the CRISPR-SLI-ARiBo RNA when incubated at 37°C with 1-2 μM Cse3 for 15-30 min, however the A linker provides the most efficient cleavage (up

to 69%) under these conditions (Supplementary Fig. S2.3C). Thus, dinucleotide linkers starting with an A were tested, and 98-100% Cse3 cleavage of CRISPR-SLI-ARiBo RNAs carrying these linkers is obtained at 37°C with 2 μ M Cse3 for 30 min (Supplementary Fig. S2.3D). Again, cleavage of these variants is more efficient at 70°C, with the single-nucleotide A, C and G linkers being cleaved to completion when incubated with 2 μ M Cse3 for 45 min or less (Fig. 2.4D) and the four tested dinucleotide linkers (AA, AC, AG and AU linkers) being cleaved to completion when incubated with 1 M Cse3 for 5 min (Fig. 2.4E). Cse3 cleavage of the CRISPR-SLI-ARiBo RNA with the single-nucleotide U linker is still very low (51%) under stringent conditions (70°C, 2 μ M Cse3 for 45 min). However, in the context of the 5'-GUAGU reversion variant, the CRISPR-SLI-ARiBo RNA with the single-nucleotide U linker (U linker with 5'-GUAGU tail) is a better substrate for Cse3, which achieves complete cleavage of the CRISPR tag when incubated at 70°C with 2 μ M Cse3 for 90 min (Supplementary Fig. S2.5). So far, this 5'-GUAGU tail reversion improves cleavage of the CRISPR tag to yield two SLI RNA variants starting with a U (no linker GC \rightarrow UA and U linker), in agreement with this U forming an A-U base pair in the 5'-GCAGA tail context that interferes with Cse3 cleavage (Sashital et al. 2011). With the goal of further improving cleavage conditions for these two SLI variants, we also tested the Cse3 N102A mutant, which is known to increase enzyme turnover (Sashital et al. 2011). However, this Cse3 mutant did not improve the cleavage results of the two SLI variants with the 5'-GUAGU tail reversion (data not shown). Nevertheless, complete Cse3 cleavage (\geq 99%) can be obtained for all variants tested containing linkers of 0-3 nucleotides between a 5'-CRISPR tag and the desired RNA. As validated here with the SLI RNA, the Cse3/CRISPR system constitutes a reliable molecular tool to ensure 5'-homogeneity of in vitro transcribed RNA.

Affinity purification of RNA with homogeneous ends using CRISPR and ARiBo tags

To investigate the compatibility of the Cse3/CRISPR system with affinity purification of RNA, we modified our ARiBo-based affinity purification procedure (Di Tomasso et al. 2011) by incorporating a Cse3 cleavage step after the transcription and before affinity immobilization (Fig. 2.5A). This new procedure was tested with CRISPR-SLI-ARiBo precursors (Fig. 2.5B) containing SLI sequences that give the highest levels of 5'-sequence heterogeneity when purified from SLI-ARiBo precursors (Fig. 2.5B). As expected, affinity purification of these SLI RNAs from CRISPR-SLI-ARiBo transcripts yields in each case a single SLI product, and thus completely resolves the problem of 5'-sequence heterogeneity observed previously for these SLI RNAs (Fig. 2.5C). In addition, the alkaline phosphatase step can be bypassed because Cse3 cleavage produces a 5'-hydroxyl termini (Gesner et al. 2011; Jore et al. 2011). Furthermore, affinity purifications from these double-fusion RNAs are fairly efficient with 13-38% SLI RNA eluted and low levels of contaminants ($\leq 1\%$ ARiBo tag and $\leq 0.5\%$ CRISPR-SLI; Supplementary Fig. S2.6). The level of cleaved CRISPR tag contaminants varies depending on the SLI RNAs; those purification requiring the most stringent Cse3 cleavage conditions produce 1-4% cleaved CRISPR tag contaminant, whereas that of SLI(2) starting with GAG was essentially free of such contaminant ($\leq 1\%$; Supplementary Fig. S2.6). These results clearly demonstrate the compatibility of the ARiBo affinity purification with Cse3 cleavage of a 5'-CRISPR tag.

Discussion

Several strategies are available to ensure 5'-sequence homogeneity of RNA transcribed in vitro with the T7 RNAP. First, if the 5'-sequence composition of the RNA of interest is not restricted, it can be selected among those that yield very low 5'-heterogeneity. The percentages of 5'-sequence heterogeneity obtained here for the SLI RNAs may provide a useful reference for future RNA syntheses. When using the wild-type T7 RNAP, low levels of 5'-sequence heterogeneity ($\leq 1.5\%$) were observed only for 5'-GNN sequences starting with GUG, GGA, GAA, GCA,

GUA, GAC, GUC, GGU, GAU, GCU or GUU and 5'-ANN sequences starting with AAG, ACG, AGC, ACU and AUU. The pool of starting sequences that yields acceptable levels of 5'-homogeneity is slightly expanded when using the P266L mutant, since it also includes those starting with GCC, AUG and AGU.

For producing RNA with homogeneous 5'-end, the Cse3/CRISPR system provides a more general approach. Interestingly, we found that the Cse3 cleavage of the 5'-CRISPR tag can be achieved quickly and efficiently at 70°C from CRISPR-SLI-ARiBo RNA transcripts to generate SLI-ARiBo RNAs starting with any of the four standard nucleotides (G, C, A or U) within either a single- or double-stranded structure. The compatibility of the Cse3/CRISPR system with our batch affinity purification protocol using an ARiBo tag makes it possible to quickly purify RNA with homogeneity at both the 5' and the 3'-ends. For a given CRISPR-RNA-ARiBo precursor, Cse3 cleavage conditions can be optimized to ensure complete CRISPR cleavage generally within 45 min or less. The selection of a 5'-sequence for which complete Cse3 cleavage can be obtained quickly at 37°C should be considered, if possible, to prevent RNA conformational changes and minimize non-enzymatic RNA hydrolysis that may occur at higher temperatures. Complete Cse3 cleavage can be obtained in 30 min or less at 37°C for RNA sequences starting with a paired adenine or with a single-stranded AUG, AU, AC or AG sequence. With such starting sequences, it should be straightforward to purify RNA with homogeneous 5'- and 3'-ends under non-denaturing conditions from a CRISPR-RNA-ARiBo precursor cleaved with Cse3 and then purified by affinity using the ARiBo tag.

Preparation of RNA with 5'-homogeneity using the Cse3/CRISPR system presents one minor limitation but several important advantages, particularly when compared with commonly-used cis-cleaving hammerhead ribozymes (Taira et al. 1991; Price et al. 1995; Ferré-D'Amaré and Doudna 1996; Walker et al. 2003). The minor limitation is the requirement for large quantities of purified Cse3, which is most effective at protein:RNA ratios of 1:1, 2:1 or higher. Nevertheless, the Cse3 protein can be purified easily with high yields (200 mg/L of media). Despite this minor limitation, there are several important advantages of the Cse3/CRISPR

system that are worth considering. The CRISPR tag is relatively small, which may be desirable when transcribing RNA in the presence of limiting, modified and/or expensive NTPs. Like the hammerhead ribozymes, the Cse3 endoribonuclease is sequence specific, thus there should be little concern with undesirable cleavage outside the tag. Cse3 has also little sequence requirement 3' of the cleavage site, and thus usage of a 5'-CRISPR tag expands the sequence possibilities for the first nucleotide of the RNA target beyond that permitted by the T7 RNAP. Moreover, unlike ribozymes used as 5'-tags, the CRISPR tag folds independently of the RNA of interest and there is no need to systematically tailor the tag sequence for each new RNA sequence. The fact that complete cleavage of the 5'-tag can be obtained quickly is also a distinct advantage of the Cse3/CRISPR system. Evidently, complete cleavage helps ensure high yields of purified RNA. More importantly, it ensures high RNA purity when combined with affinity purification, as incomplete cleavage would yield CRISPR-RNA contaminants in the final RNA sample and thus reduce the advantage of ARiBo-based affinity purification. Incomplete cleavage can be problematic for ribozyme tags that require folding of an active structure that is incompatible with that of the RNA target (Grosshans and Cech 1991; Price et al. 1995; Walker et al. 2003). In contrast, the Cse3/CRISPR system is well suited for batch affinity purification using the ARiBo tag, and the combination of the CRISPR and ARiBo tags allows for rapid production of RNA with homogeneity at both the 5' and the 3' ends.

The practicality of the ARiBo-based affinity purification has already been demonstrated for several RNAs, including the small SLI RNA hairpin (29 nt; this manuscript), the *let-7g* pre-miRNA [46 nt; (Di Tomasso et al. 2012a)], a purine riboswitch variant [69 nt; (Di Tomasso et al. 2011)] and the *Neurospora* VS ribozyme (138 nt; this manuscript). Here, we described two approaches to insure that the affinity-purified RNA has a homogeneous 5'-end, either choosing an appropriate GNN/ANN starting sequence or incorporating the Cse3/CRISPR system in the purification scheme. The later option, although more elaborate, may be particularly advantageous for preparation of RNAs starting with a GNN/ANN sequence that do not allow 5'-homogeneity or that start with a YNN sequence.

Furthermore, the Cse3/CRISPR system provides a novel tool for RNA purification that can be either used by itself or combined with other purification methods, such as the ARiBo-based affinity purification. Future studies will help define its range of applicability for different RNA sequences and purification strategies, and creative applications may reveal additional advantages of this system.

Materials and methods

Expression and purification of the $\lambda N^+ - L^+ - GST$ protein

The expression and purification of the $\lambda N^+ - L^+ - GST$ protein was performed as described earlier (Di Tomasso et al. 2011; Di Tomasso et al. 2012b).

Expression and purification of the *T. thermophilus* Cse3 endoribonuclease

The pET-30a(+) vector containing the *T. thermophilus* Cse3 gene was a kind gift from A.M. MacMillan [University of Alberta; (Gesner et al. 2011)]. The plasmid was transformed into Rosetta (DE3) cells (Novagen). The bacteria were grown at 37°C in Luria-Bertani (LB) media, and protein expression was induced with 1 mM isopropyl--D-1-thiogalactopyranoside (IPTG) for 16 h at 25°C. The cells were harvested by centrifugation and resuspended in Cse3 buffer (20 mM Tris pH=7.4, 1 M NaCl, 1 mM DTT and 20 mM imidazole) supplemented with 0.15% w/v protease inhibitor (Sigma-Aldrich). The cells were lysed by French press, sonicated 10 s and centrifuged at 12 000 g for 15 min at 4°C. The supernatant was then heated at 55°C for 30 min and ultracentrifuged at 138000 g for 1 h at 4°C. The supernatant was incubated for 1 h at 4°C with Ni-charged IMAC Sepharose 6 Fast Flow (GE Healthcare). The resin was then washed two times with the Cse3 buffer. The bound His₆-tagged Cse3 fusion protein was eluted from the resin by two 10-min incubations at room temperature with elution buffer (Cse3 buffer + 180 mM imidazole). The supernatant was dialyzed overnight at 4°C in FPLC-A buffer (20 mM sodium phosphate pH=7.4, 1 mM EDTA and 1 mM DTT) and then applied to an SP Sepharose high-performance column (GE Healthcare; 100 mL bed volume) equilibrated with FPLC-A buffer. The proteins were eluted from the column using a gradient (from 0 to 100% over 525 mL) of

FPLC-B buffer (FPLC-A buffer with 2 M NaCl). The pooled fractions containing the protein of interest were dialyzed overnight at 4°C in Cse3 storage buffer (50 mM HEPES pH=8.0, 100 mM NaCl, 2 mM DTT and 20% glycerol). The purity ($\geq 95\%$) and correct mass of the protein were verified by SDS-PAGE and mass spectroscopy, respectively, whereas its RNase-free state was verified as previously described (Di Tomasso et al. 2012b).

Expression and purification of the wild-type and mutant T7 RNAP

The wild-type T7 RNAP with an N-terminal His₆ tag was prepared from *Escherichia coli* strain BL21 carrying the plasmid pT7-911Q. The P266L mutant clone was prepared from the wild-type T7 clone using the QuikChangeII site-restricted mutagenesis procedure (Stratagene) and verified by DNA sequencing. The bacteria were grown at 37°C in LB media, and protein expression was induced with 1 mM IPTG for 4 h at 30°C. The cells were harvested by centrifugation and resuspended in T7 buffer (50 mM Tris pH=8.0, 100 mM NaCl, 1 mM imidazole, 1 mM DTT and 5% glycerol) supplemented with Complete EDTA-free protease inhibitor tablets (Roche). The cells were lysed by French press, sonicated 10 s and centrifuged at 138 000 g for 1 h at 4°C. The supernatant was incubated for 1 h at 4°C with Ni-charged IMAC Sepharose 6 Fast Flow beads (GE Healthcare). The resin was then washed once with T7 buffer supplemented with protease inhibitor, three times with T7 buffer and four times with Ni wash buffer (T7 buffer + 9 mM imidazole). The bound T7 RNAP was eluted from the resin by three 5-min incubations at room temperature with elution buffer (T7 buffer + 199 mM imidazole). The supernatant was dialyzed at 4°C in S7 nuclease buffer (50 mM Tris pH=8.0, 50 mM NaCl, 5 mM CaCl₂ and 1 mM DTT) and then incubated for 30 min at 37°C with 75 U of S7 nuclease (Roche) per liter of bacteria. The sample was then dialyzed overnight at 4°C in Cobalt buffer (50 mM sodium phosphate pH=7.0, 300 mM NaCl, 1 mM imidazole and 1 mM DTT), and the dialyzed sample was applied on a Co-charged IMAC Sepharose 6 Fast Flow column (GE Healthcare). The resin was washed with 3 column volumes of Cobalt buffer and with 3 column volumes of Cobalt wash buffer (Cobalt buffer + 49 mM imidazole).

The bound T7 RNAP was eluted from the resin with elution buffer (Cobalt buffer + 299 mM imidazole). The pooled fractions containing the protein of interest were dialyzed overnight at 4°C in T7 storage buffer (20 mM KH₂PO₄, 0.1 mM EDTA, 1 mM DTT and 50% glycerol, adjusted to pH=8.0). The high purity (≥ 98%) of the protein was verified by SDS-PAGE.

Clonig of the pSLI-ARiBo and pCRISPR-SLI-ARiBo plamids

The pARiBo4 plasmid expressing the ARiBo4 tag (Supplementary Fig. S2.1) was generated by mutagenesis of the pARiBo1 plasmid (Di Tomasso et al. 2011). All pSLI-ARiBo4 plasmids expressing the SLI-ARiBo4 fused RNAs (Figs. 2.1A, 2.2A and 2.3A) were obtained by mutagenesis of the pARiBo4 plasmid. The pCRISPR-SLI-ARiBo plasmids expressing the CRISPR-SLI-ARiBo RNAs were generated by mutagenesis of pSLI(2)-ARiBo4. The pCRISPR-SLI-ARiBo plasmids contain two restriction sites for subcloning: the BstXI site (5'-CCANNNNNNTGG-3') located within the CRISPR sequence and the ApaI site located within the *boxB* hairpin (5'-GGGCC-3') at the 5'-end of the ARiBo tag. All mutagenesis were performed using the QuikChangeII site-restricted mutagenesis procedure, and all sequences were verified by DNA sequencing.

In vitro transcription of ARiBo-fused RNAs

Medium-scale preparations (~0.5 mg) of plasmid DNA template were typically obtained by growing 0.15 L of plasmid-transformed DH5α cells (Invitrogen), purifying the plasmid using the QIAGEN Plasmid Maxi Kit and linearizing it with EcoRI (New England Biolabs). The SLI-ARiBo fusion RNAs were transcribed at 37°C for 3 h using the following reaction conditions: 40 mM HEPES pH=8.0, 50 mM DTT, 0.1% Triton X-100, 1 mM spermidine, 4 mM ATP, 4 mM CTP, 4 mM UTP, 4 mM GTP, 20 mM MgCl₂, 60 µg/mL His-tagged T7 RNAP, 3 U/mL RNasin Ribonuclease Inhibitor (Promega) and 80 µg/mL of linearized plasmid DNA template. The CRISPR-SLI-ARiBo RNAs were transcribed using the same conditions, although 25 mM MgCl₂ and a template concentration of 120 µg/mL were used for some variants. Transcription reactions

were stopped by adding 20-25 mM EDTA pH=8.0 and stored at 20°C.

Small-scale affinity batch purification of SLI RNAs from SLI-ARiBo fusion RNAs

Small-scale affinity batch purifications were performed as described earlier (Di Tomasso et al. 2011; Di Tomasso et al. 2012b). Briefly, 17.5 nmol of λ N⁺-L⁺-GST fusion protein was first added to an aliquot of the transcription reaction (75-200 μ L) containing 3.5 nmol of ARiBo-fused SLI, and the volume was completed to 400 μ L with equilibration buffer (50 mM HEPES pH=7.5). After a 15-min incubation, the RNA-protein mix was added to a spin cup column (Pierce) containing 125 μ L of washed GSH-Sepharose 4B resin (from 163 μ L of slurry; GE Healthcare) and incubated for 15 min. The load eluate was collected by centrifugation 1 min at 5000 g. The resin was washed three times with 400 μ L of equilibration buffer. All washes include incubation for 5 min and centrifugation for 1 min at 5000 g. For elution, the resin was incubated at 37°C for 30 min in 400 μ L of elution buffer (20-40 mM Tris buffer pH=7.6, 10 mM MgCl₂ and 1-10 mM of GlcN6P) and at room temperature for 5 min prior to centrifugation. The resin was then washed two times with 400 μ L of equilibration buffer. Finally, the resin was washed with 400 μ L of 2.5 M NaCl. The load eluates (LE), wash eluates (W1, W2, W3), elution (E1), elution-washes (E2 and E3) and NaCl eluates (NaCl) were kept for quantitative analysis.

VS ribozyme cleavage during affinity purification

The small-scale affinity batch purification was slightly modified by incorporating a VS ribozyme cleavage step and two additional wash steps between the third wash (W3) and the RNA elution step (E1). For VS ribozyme cleavage, the Avapl ribozyme was prepared by in vitro T7 RNAP transcription from the pAvapl-ARiBo1 plasmid derived from pAvapl (Bouchard et al. 2008; Di Tomasso et al. 2012a), linearized with EcoRI and purified by affinity, as previously described (Di Tomasso et al. 2011; Di Tomasso et al. 2012a). The VS ribozyme cleavage step was performed for 30 min at 37°C in 400 μ L of VS cleavage buffer

(50 mM HEPES pH=7.5, 25 mM MgCl₂, 50 mM KCl, 0.5 U RNasin (Promega) and 9.75 nM Avapl ribozyme) followed by an incubation of 5 min at room temperature prior to centrifugation.

Calf alkaline phosphatase cleavage of affinity-purified SLI RNAs

For calf alkaline phosphatase cleavage, aliquots (corresponding to ~0.125 nmol SLI RNA) of E1 elutions from small-scale affinity batch purifications were incubated at 37°C for 3 h in a total volume of 50 µL containing CIP buffer (50 mM HEPES pH=8.0 and 0.1 mM EDTA), 0.5 U RNasin Ribonuclease Inhibitor (Promega) and 10 U of calf intestinal alkaline phosphatase (Roche).

Cse3 cleavage assay

For Cse3 cleavage, 1 µM of unpurified CRISPR-SLI-ARiBo RNA from the transcription reaction was incubated at 37°C or 70°C in the Cse3 reaction buffer (20 mM HEPES pH=7.5 and 150 mM KCl) with either 1 or 2 µM of Cse3 protein. Aliquots (0.75 µL) were removed from the 25-µL reaction mixture at specific times, and the reaction was stopped by addition of 10 µL of loading dye (25 mM EDTA and bromophenol blue in formamide), transferred to 4°C and separated on a 10% denaturing polyacrylamide gel.

Affinity batch purification of SLI RNAs from CRIPSR-SLI-ARiBo fusion RNAs

The CRISPR-SLI-ARiBo double-fusion RNAs were synthesized from the class III T7 promoter using the wild-type T7 RNAP. The small-scale affinity batch purification of these RNAs was performed as for SLI-ARiBo fusion RNAs, except that, prior to starting, a small volume of the transcription reaction corresponding to 1.86 nmol of CRISPR-SLI-ARiBo RNA was incubated with either 3.72 nmol [SLI(2)] or 7.44 nmol [SLI(1), SLI(22) and SLI(26)] of Cse3 protein at 70°C in Cse3 reaction buffer supplemented with 1.67 U of RNasin (Promega) for a total reaction volume of 200 µL. A Cse3 inactivation step, which consists of a 2-min incubation at 95°C and 5min cooling on ice, was added to improve sample migration on the gel, but has no effect on the performance of the method. The

affinity batch purification was then resumed as described above, but downscaled by a factor of two to account for the smaller quantity of starting RNA (1.75 nmol). In addition, the first wash buffer was supplemented with 10 mM of MgCl₂ to reduce the amount of CRISPR tag impurity in the elution fractions.

Quantitative analysis from denaturing gels stained with SYBR Gold

Aliquots from the various steps of small-scale affinity batch purifications and Cse3 cleavage were analyzed by denaturing gel electrophoresis using polyacrylamide gels containing 7 M urea. Care was taken to load gels with sample volumes corresponding to precise amounts of SLI-ARiBo or CRISPR-SLI-ARiBo fusion RNA present in the transcription reaction. The gels were pre-run at 400-600 V for 20 min prior to loading the samples, and then they were run at 400-600 V for 2-4 hrs. The gels were stained for 10 min in a SYBR Gold (Invitrogen) solution [1:10000 dilution in TBE buffer (50 mM Tris-Base, 50 mM boric acid and 1 mM EDTA)] and scanned on a ChemiDoc MP imaging system (Bio-Rad). The band intensities were analyzed using the Image Lab software (version 4.0.1 from Bio-Rad).

The percentage of 5'-heterogeneity was calculated from the E1 elution fraction of affinity purifications treated with calf alkaline phosphatase using the following equation:

$$\left[\left(\sum B_{RNA} - B_{RNA^*} \right) / \sum B_{RNA} \right] \times 100\%$$

where $\sum B_{RNA}$ represents the sum of the intensities of all the SLI RNA bands and B_{RNA^*} represents the intensity of the reference SLI band only .

For Cse3 cleavage assays, the percentage of Cse3 cleavage was determined using the following equation:

$$\left\{ (B_C / nt_C) / [(B_C / nt_C) + (B_U / nt_U)] \right\} \times 100\%$$

where B_C and B_U correspond to the band intensities of Cse3-cleaved and uncleaved CRISPR-SLI-ARiBo fusion RNA, respectively, whereas nt_C and nt_U represent the number of nucleotides for the SLI-ARiBo and CRISPR-SLI-ARiBo fusion RNAs, respectively.

Quantitative analysis of small-scale affinity batch purifications from CRISPR-SLI-ARiBo precursors was achieved based on our published procedure for quantification of affinity purification from RNA-ARiBo precursors⁴⁷. First, control lanes were loaded with known amounts of RNA to derive standard curves that were used to determine the quantity (in ng) of SLI (N_{RNA}), CRISPR tag (N_{CRISPR}), ARiBo tag (N_{ARiBo}), and SLI-ARiBo fusion RNA (N_{Fusion}) at each purification step. For the SLI and CRISPR tag standard curves, known quantities of purified CRISPR RNA and a control RNA derived from SLI (5'-GAG CGA AGG CUG GAC CAC CAG CCG AGC UC-3') were loaded on the gel, whereas other standard curves were calibrated as previously described⁴⁷. The percentage of RNA eluted was calculated using the following equation:

$$\left[\left(\sum N_{RNA} \right) / I_{RNA} \right] \times 100\%$$

where $\sum N_{RNA}$ represents the total amount of SLI (ng) detected in lanes E1 and E2, and I_{RNA} represents the calculated amount of SLI expected from an equivalent volume of transcription (100 ng). In addition, the percentages of ARiBo tag and CRISPR tag impurities were calculated from the E1 lane using the equations $[N_{ARiBo} / (N_{RNA} + N_{ARiBo})] \times 100\%$ and $[N_{CRISPR} / (N_{RNA} + N_{CRISPR})] \times 100\%$ respectively.

Acknowledgments

We thank Andrew MacMillan for the expression vector of *T. thermophilus* His-tagged Cse3, Julie Lacroix-Labonté for the control SLI RNA and James G. Omichinski for critical reading of the manuscript. This work was supported by grants to PL from the Canadian Institutes for Health Research (CIHR; HOP-83068 and MOP-86502) and the Natural Sciences and Engineering Council of Canada (NSERC). ASL and BLP respectively hold graduate and summer scholarships from the CDMC-CREATE (Cellular Dynamics of Macromolecular Complexes-Collaborative Research and Training Experience) program funded by NSERC. PL holds a Canada Research Chair in Engineering and Structural Biology of RNA.

Tables

Table I. 5'-sequence heterogeneity^a of affinity-purified SLI RNAs synthesized as SLI-ARiBo precursors from the T7 class III promoter

5' sequences	5'-sequence heterogeneity (%)	
	Wild-type T7 RNAP	P266L T7 RNAP
1-GGG	45 ± 4	37 ± 4
2-GAG	17 ± 3	15 ± 3
3-GCG	2.1 ± 0.7	1.8 ± 0.3
4-GUG	0.6 ± 0.4	0.3 ± 0.3
5-GGA	0.6 ± 0.4	0.2 ± 0.1
6-GAA	0.2 ± 0.2	0.2 ± 0.2
7-GCA	0.8 ± 0.5	0.11 ± 0.03
8-GUA	0.11 ± 0.04	0.09 ± 0.03
9-GGC	2.0 ± 0.5	2.7 ± 0.6
10-GAC	0.2 ± 0.2	0.13 ± 0.07
11-GCC	2.2 ± 0.6	0.7 ± 0.5
12-GUC	0.3 ± 0.2	0.13 ± 0.06
13-GGU	1.3 ± 0.5	0.8 ± 0.4
14-GAU	0.11 ± 0.04	0.2 ± 0.2
15-GCU	0.3 ± 0.3	0.2 ± 0.1
16-GUU	0.2 ± 0.1	0.14 ± 0.04

^aThe percentage of sequence heterogeneity is given with respect to the most intense band on the gel, which likely represents the accurately-transcribed RNA product based on the relative migration distances (see text). All experiments were performed at least three times such that the average value and its error respectively represent the statistical mean and standard deviation from these multiple experiments.

Table II. 5'-sequence heterogeneity^a of affinity-purified SLI RNAs synthesized as SLI-ARiBo precursors from the T7 class II ϕ 2.5 promoter

5' sequences	5'-sequence heterogeneity (%)	
	Wild-type T7 RNAP	P266L T7 RNAP
17-AGG	11 ± 2	3.3 ± 0.5
18-AAG	0.4 ± 0.4	0.6 ± 0.7
19-ACG	0.2 ± 0.2	0.3 ± 0.2
20-AUG	2.3 ± 0.4	0.4 ± 0.3
21-AGA	7 ± 1	4.4 ± 0.5
22-AAA	84 ± 2	87 ± 1
23-ACA	5 ± 1	5.4 ± 0.4
24-AUA	9 ± 1	18 ± 1
25-AGC	0.12 ± 0.07	0.3 ± 0.2
26-AAC	68 ± 1	50 ± 1
27-ACC	11.5 ± 0.7	11 ± 2
28-AUC	12 ± 1	9 ± 1
29-AGU	2.1 ± 0.9	0.3 ± 0.2
30-AAU	15 ± 2	6 ± 2
31-ACU	0.15 ± 0.08	0.2 ± 0.2
32-AUU	0.2 ± 0.2	0.14 ± 0.09

^aThe percentage of sequence heterogeneity is given with respect to the most intense band on the gel, except for SLI RNAs 22 (AAA) and 26 (AAC), where the percentage is given with respect to the band that contains the accurately-transcribed RNA (see Fig. 2.5C). All experiments were performed at least three times such that the average value and its error respectively represent the statistical mean and standard deviation from these multiple experiments.

Figures

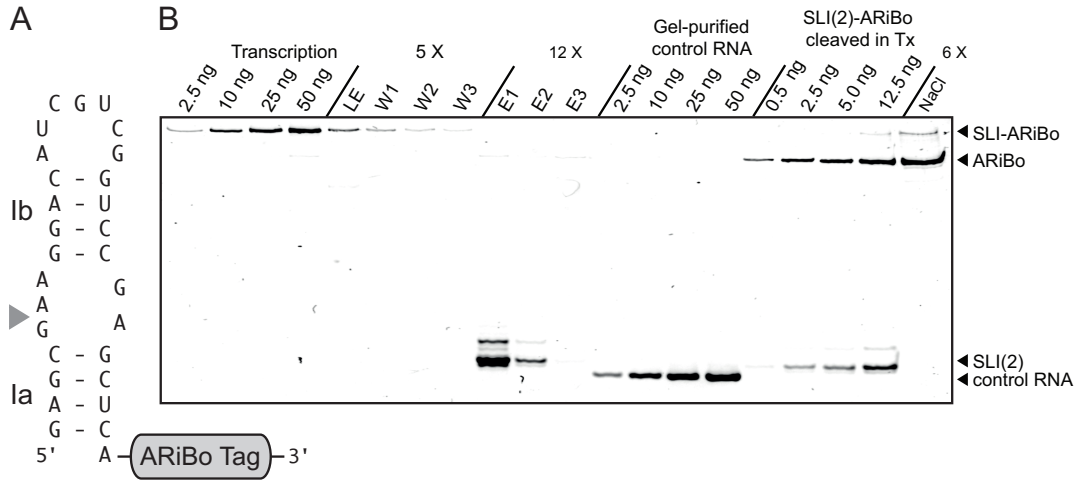


Figure 2.1 : Evidence of 5'-heterogeneity revealed from affinity purification of an SLI RNA transcribed from an SLI-ARiBo precursor. (A) Schematic representation of SLI(2)-ARiBo fusion RNA. The grey arrowhead points to the VS ribozyme cleavage site in the internal loop between stem Ia and stem Ib. (B) Small-scale affinity batch purification of SLI(2) analyzed on a 20% denaturing polyacrylamide gel stained with SYBR Gold. The SLI(2) RNA was transcribed as an ARiBo-fusion RNA [SLI(2)-ARiBo] and purified by affinity. Aliquots from each purification step were loaded on the gel (LE: load eluate; W1–3: washes; E1–3: elutions; and NaCl: matrix regeneration with 2.5 M NaCl) in amounts shown, where 1X correspond to approximately 50 ng of SLI(2)-ARiBo precursor present in the transcription reaction or the equivalent of 8.23 ng of SLI(2) to be purified. In addition, standard quantities of SLI(2)-ARiBo from the transcription reaction, gel-purified control RNA (29 nt) and SLI(2) cleaved in the transcription reaction were loaded as controls. Bands corresponding to the SLI(2)-ARiBo (176 nt), the ARiBo tag (147 nt) and SLI(2) (29 nt) RNAs are indicated on the right side of the gel.

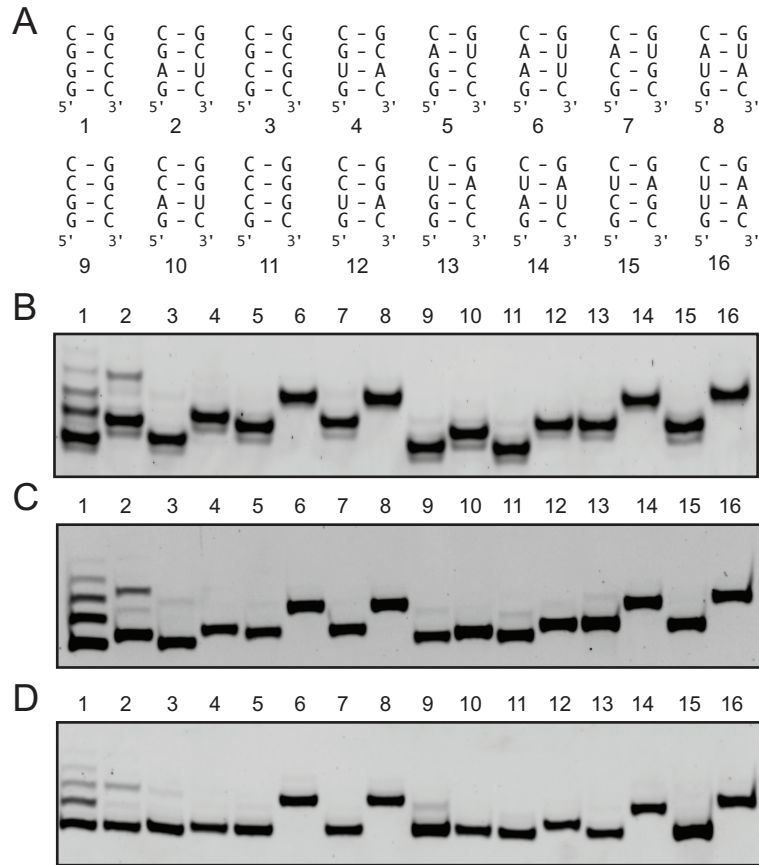


Figure 2.2 : Effect of the 5' sequence on the heterogeneity of SLI RNAs transcribed as SLI-ARiBo precursors from the consensus T7 class III promoter. (A) Sequence of stem Ia and numbering of SLI RNAs with different 5' (and 3') sequences. (B) Small-scale affinity batch purifications of each of the 16 SLI RNAs transcribed as SLI-ARiBo precursors from the T7 class III promoter using the wild-type T7 RNAP and analyzed on a 20% denaturing polyacrylamide gel stained with SYBR Gold. Only the E1 elution fractions are shown, and from the 400- μ L elution volumes, 1.2- μ L aliquots were loaded on the gel. (C)(D) Small-scale affinity batch purifications of each of the 16 SLI RNAs transcribed as SLI-ARiBo RNAs from the T7 class III promoter using the wild-type (C) or P266L mutant (D) T7 RNAP. These E1 elution fractions were treated with calf alkaline phosphatase to remove phosphate heterogeneity at the 5'-end (in C and D only). From the 50- μ L phosphatase reaction mixture, 6.3- μ L aliquots were analyzed on 20% denaturing polyacrylamide gels stained with SYBR Gold. In (B)-(D), gel lanes match the SLI numbering given in (A).

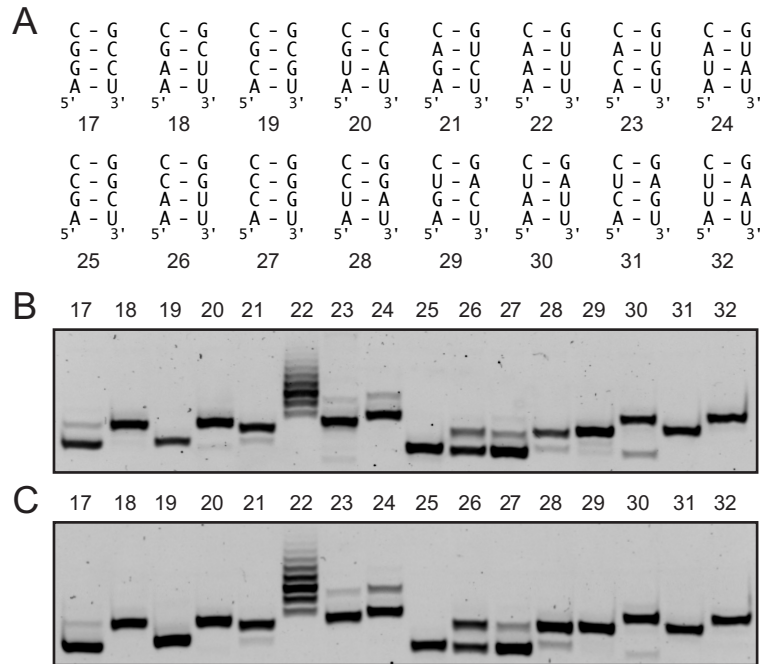


Figure 2.3 : Effect of the 5' sequence on the heterogeneity of SLI RNAs transcribed as SLI-ARiBo precursors from the T7 class II ϕ 2.5 promoter. (A) Sequence of stem Ia and numbering of SLI RNAs with different 5' (and 3') sequences. (B)(C) Small-scale affinity batch purifications of each of the 16 SLI RNAs transcribed as SLI-ARiBo RNAs from the T7 class II ϕ 2.5 promoter using the wild-type (B) or P266L mutant (C) T7 RNAP. The E1 elution fractions were treated with calf alkaline phosphatase to remove phosphate heterogeneity at the 5'-end. From the 50- μ L phosphatase reaction mixture, 6.3- μ L aliquots were analyzed on 20% denaturing polyacrylamide gels stained with SYBR Gold. In (B) and (C), gel lanes match the SLI numbering given in (A).

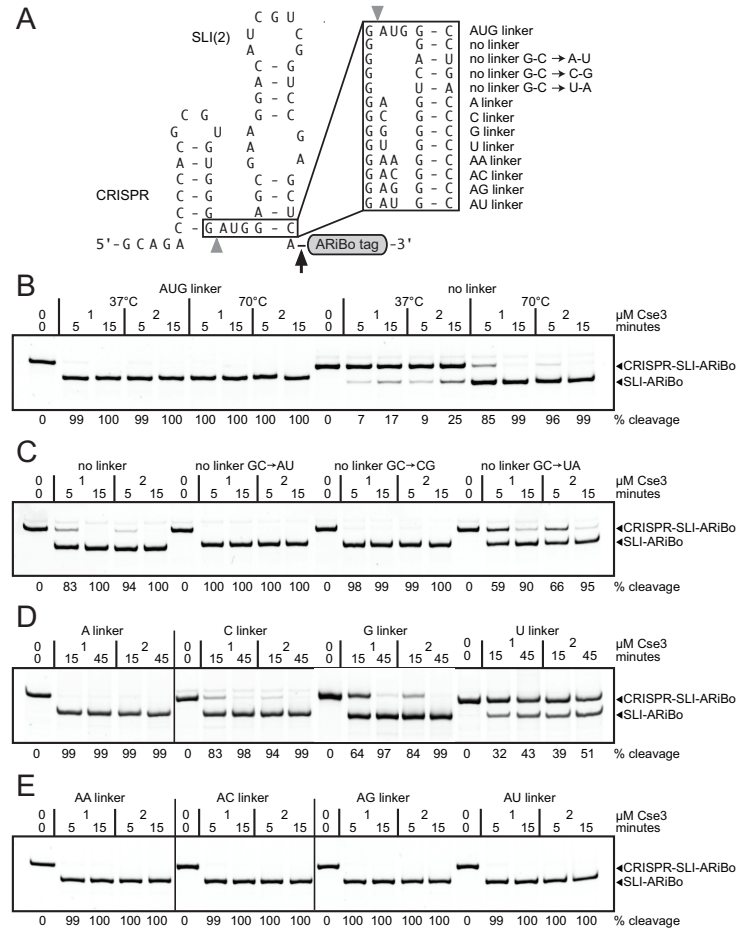


Figure 2.4 : Effect of CRISPR-RNA junction sequence on Cse3 cleavage. (A) Schematic representation of CRISPR-SLI(2)-ARiBo double-fusion RNAs with the original CRISPR sequence (AUG linker) or related variants with sequence changes at the CRISPR-RNA junction (boxed area). The grey arrowhead points to the Cse3 cleavage site, whereas the black arrow points to the glmS cleavage site. (B) to (E) Cse3 cleavage of CRISPR-SLI-ARiBo RNAs analyzed on 10% denaturing polyacrylamide gels stained with SYBR Gold. Cse3 cleavage was performed using aliquots from the transcription reactions (~1 μM RNA), 20 mM HEPES pH=7.5, 150 mM KCl, either 1 or 2 μM Cse3 and different incubation times, as indicated above each lane. For experiments reported in (B), Cse3 cleavage was performed at either 37°C or 70°C, as indicated, whereas for those reported in (C)-(E), Cse3 cleavage was performed at 70°C. The gel mobility of the RNA precursor (CRISPR-SLI-ARiBo) and the Cse3 cleavage product (SLI-ARiBo) are indicated with arrows on the right side of the gels. The percentages of Cse3 cleavage are given below the gels.

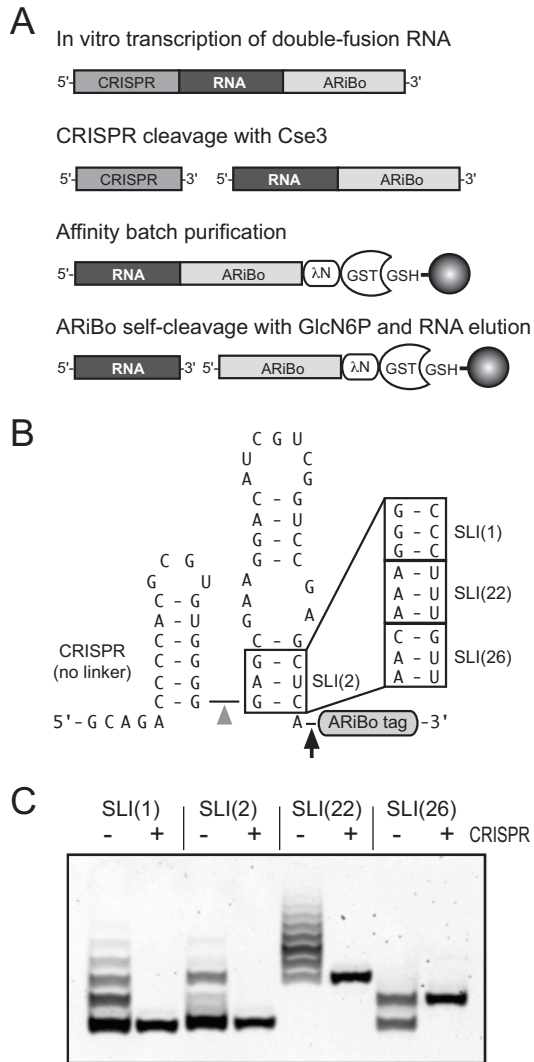


Figure 2.5 : Affinity purification of RNA with homogeneous ends using CRISPR and ARiBo tags. (A) General strategy for affinity batch purification of RNA from a CRISPR-RNA-ARiBo precursor, a double-fusion RNA with a 5'-CRISPR tag and a 3'-ARiBo tag. After transcription, Cse3 cleavage of the CRISPR tag yields the ARiBo-fusion RNA, which is bound to a λ N-GST fusion protein and immobilized on GSH-Sepharose beads. After several washes, RNA elution is triggered by addition of GlcN6P, which activates the *glmS* ribozyme of the ARiBo tag. (B) Schematic representation of CRISPR-SLI-ARiBo double-fusion RNAs with an AUG deletion (no linker) just 3' from the Cse3 cleavage site. (C) Small-scale affinity batch purifications of SLI RNAs from SLI-ARiBo (- lanes) and CRISPR-SLI-ARiBo (+ lanes) precursors analyzed on a 20% denaturing polyacrylamide gel stained with SYBR Gold. For purification of SLI-ARiBo precursors, the E1 elution

fractions were treated with calf alkaline phosphatase prior to loading on the gel; they correspond to samples shown in Figs. 2.2D and 2.3C. For CRISPR-SLI-ARiBo precursors, CRISPR cleavage was performed for 15 min at 70°C using an RNA:Cse3 ratio of 1:2 [SLI(2)] or for 30 min at 70°C using an RNA:Cse3 ratio of 1:4 [SLI(1), SLI(22) and SLI(26)]. Aliquots of the E1 elution fractions were loaded on the gel.

References

- Al-Attar S, Westra ER, van der Oost J, Brouns SJ. 2011. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biol Chem* 392(4): 277-289.
- Batey RT, Kieft JS. 2007. Improved native affinity purification of RNA. *RNA* 13(8): 1384-1389.
- Bouchard P, Lacroix-Labonté J, Desjardins G, Lampron P, Lisi V, Lemieux S, Major F, Legault P. 2008. Role of SLV in SLI substrate recognition by the *Neurospora* VS ribozyme. *RNA* 14: 736-748.
- Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321(5891): 960-964.
- Cazenave C, Uhlenbeck OC. 1994. RNA template-directed RNA synthesis by T7 RNA polymerase. *Proc Natl Acad Sci U S A* 91(15): 6972-6976.
- Chamberlin M, Ryan T. 1982. Bacteriophage DNA-dependent RNA polymerases. In *The Enzymes*, Vol XV (ed. PD Boyer), pp. 87-108. Academic Press.
- Coleman TM, Wang G, Huang F. 2004. Superior 5' homogeneity of RNA from ATP-initiated transcription under the T7 phi 2.5 promoter. *Nucleic Acids Res* 32(1): e14.
- Cunningham PR, Weitzmann CJ, Ofengand J. 1991. SP6 RNA polymerase stutters when initiating from an AAA... sequence. *Nucleic Acids Res* 19(17): 4669-4673.

- Di Tomasso G, Dagenais P, Desjardins A, Rompré-Brodeur A, Delfosse V, Legault P. 2012a. Affinity purification of RNA using an ARiBo tag. In *Recombinant and in vitro RNA synthesis: Methods and Protocols*, Vol 941, p. 306. Humana Press.
- Di Tomasso G, Lampron P, Dagenais P, Omichinski JG, Legault P. 2011. The ARiBo tag: a reliable tool for affinity purification of RNAs under native conditions. *Nucleic Acids Res* 39(3): e18.
- Di Tomasso G, Lampron P, Omichinski JG, Legault P. 2012b. Preparation of λ N-GST fusion protein for affinity immobilization of RNA. In *Recombinant and in vitro RNA synthesis: Methods and Protocols*, Vol 941, p. 306. Humana Press.
- Doudna JA. 1997. Preparation of homogeneous ribozyme RNA for crystallization. *Methods Mol Biol* 74: 365-370.
- Dunn JJ, Studier FW. 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J Mol Biol* 166(4): 477-535.
- Dzianott AM, Bujarski JJ. 1988. An in vitro transcription vector which generates nearly correctly ended RNAs by self-cleavage of longer transcripts. *Nucleic Acids Res* 16(22): 10940.
- Ferré-D'Amaré AR, Doudna JA. 1996. Use of cis- and trans-ribozymes to remove 5' and 3' heterogeneities from milligrams of in vitro transcribed RNA. *Nucleic Acids Res* 24(5): 977-978.
- Frank R, Köster H. 1979. DNA chain length markers and the influence of base composition on electrophoretic mobility of oligodeoxyribonucleotides in polyacrylamide-gels. *Nucleic Acids Res* 6(6): 2069-2087.
- Garside EL, Schellenberg MJ, Gesner EM, Bonanno JB, Sauder JM, Burley SK, Almo SC, Mehta G, Macmillan AM. 2012. Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. *RNA* 18(11): 2020-2028.

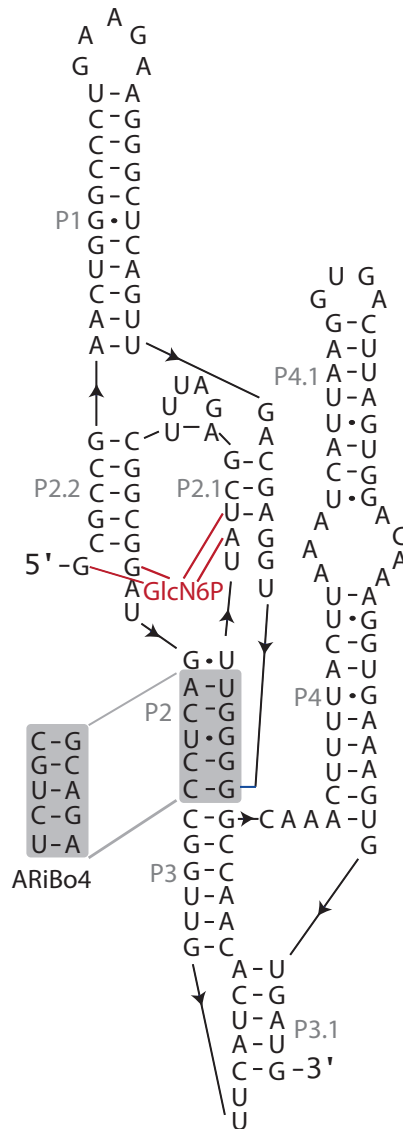
- Gesner EM, Schellenberg MJ, Garside EL, George MM, Macmillan AM. 2011. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat Struct Mol Biol* 18(6): 688-692.
- Grosshans CA, Cech TR. 1991. A hammerhead ribozyme allows synthesis of a new form of the Tetrahymena ribozyme homogeneous in length with a 3' end blocked for transesterification. *Nucleic Acids Res* 19(14): 3875-3880.
- Guillerez J, Lopez PJ, Proux F, Launay H, Dreyfus M. 2005. A mutation in T7 RNA polymerase that facilitates promoter clearance. *Proc Natl Acad Sci U S A* 102(17): 5958-5963.
- Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA. 2010. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329(5997): 1355-1358.
- Helm M, Brule H, Giege R, Florentz C. 1999. More mistakes by T7 RNA polymerase at the 5' ends of in vitro-transcribed RNAs. *RNA* 5(5): 618-621.
- Hoffmann B, Mitchell GT, Gendron P, Major F, Andersen AA, Collins RA, Legault P. 2003. NMR structure of the active conformation of the Varkud satellite ribozyme cleavage site. *Proc Natl Acad Sci U S A* 100(12): 7003-7008.
- Horvath P, Barrangou R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327(5962): 167-170.
- Imburgio D, Rong MQ, Ma KY, McAllister WT. 2000. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry* 39(34): 10419-10430.
- Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R et al. 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* 18(5): 529-536.

- Kao C, Zheng M, Rudisser S. 1999. A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* 5(9): 1268-1272.
- Karginov FV, Hannon GJ. 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol Cell* 37(1): 7.
- Kieft JS, Batey RT. 2004. A general method for rapid and nondenaturing purification of RNAs. *RNA* 10(6): 988-995.
- Krupp G. 1988. RNA synthesis: strategies for the use of bacteriophage RNA polymerases. *Gene* 72(1-2): 75-89.
- Lacroix-Labonte J, Girard N, Lemieux S, Legault P. 2012. Helix-length compensation studies reveal the adaptability of the VS ribozyme architecture. *Nucleic Acids Res* 40(5): 2284-2293.
- Lapham J, Crothers DM. 1996. RNase H cleavage for processing of in vitro transcribed RNA for NMR studies and RNA ligation. *RNA* 2(3): 289-296.
- Lehrach H, Diamond D, Wozney JM, Boedtke H. 1977. RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry* 16(21): 4743-4751.
- Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF et al. 2011. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9(6): 467-477.
- Marraffini LA, Sontheimer EJ. 2010. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* 11(3): 181-190.
- Martin CT, Muller DK, Coleman JE. 1988. Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry* 27(11): 3966-3974.
- Melton DA, Krieg PA, Rebagliati MR, Maniatis T, Zinn K, Green MR. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res* 12(18): 7035-7056.

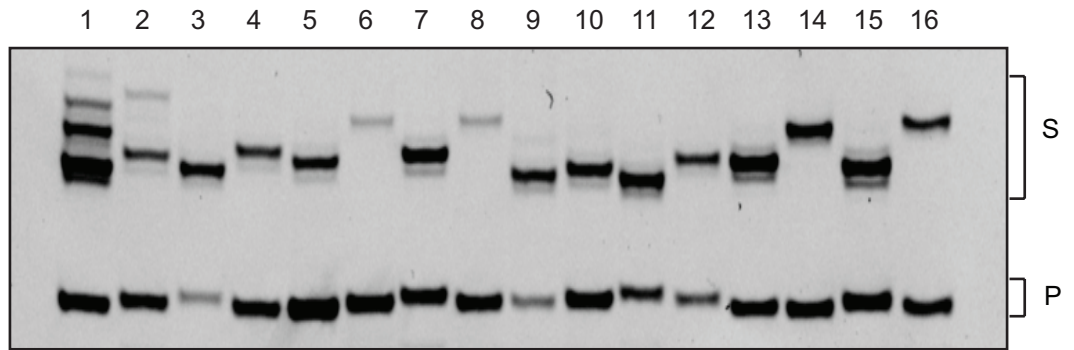
- Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* 15: 8783-8798.
- Moroney SE, Piccirilli JA. 1991. Abortive products as initiating nucleotides during transcription by T7 RNA polymerase. *Biochemistry* 30: 10343-10349.
- Nelissen FH, Leunissen EH, van de Laar L, Tessari M, Heus HA, Wilmenga SS. 2012. Fast production of homogeneous recombinant RNA--towards large-scale production of RNA. *Nucleic Acids Res* 40(13): e102.
- Pereira MJ, Behera V, Walter NG. 2010. Nondenaturing purification of co-transcriptionally folded RNA avoids common folding heterogeneity. *Plos One* 5(9): e12953.
- Pleiss JA, Derrick ML, Uhlenbeck OC. 1998. T7 RNA polymerase produces 5' end heterogeneity during in vitro transcription from certain templates. *RNA* 4(10): 1313-1317.
- Price SR, Ito N, Oubridge C, Avis JM, Nagai K. 1995. Crystallization of RNA-protein complexes I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J Mol Biol* 249(2): 398-408.
- Ramírez-Tapia LE, Martin CT. 2012. New insights into the mechanism of initial transcription: The T7 RNA polymerase mutant P266L transitions to elongation at longer RNA lengths than wild type *J Biol Chem* 287(44): 37352-37361.
- Santoro SW, Joyce GF. 1997. A general purpose RNA-cleaving DNA enzyme. *Proc Natl Acad Sci U S A* 94: 4262-4266.
- Sashital DG, Jinek M, Doudna JA. 2011. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat Struct Mol Biol* 18(6): 680-687.
- Schürer H, Lang K, Schuster J, Mörl M. 2002. A universal method to produce in vitro transcripts with homogeneous 3' ends. *Nucleic Acids Res* 30(12): e56.
- Sherlin LD, Bullock TL, Nissan TA, Perona JJ, Lariviere FJ, Uhlenbeck OC, Scaringe SA. 2001. Chemical and enzymatic synthesis of tRNAs for high-throughput crystallization. *RNA* 7(11): 1671-1678.

- Taira K, Nakagawa K, Nishikawa S, Furukawa K. 1991. Construction of a novel RNA-transcript-trimming plasmid which can be used both in vitro in place of run-off and (G)-free transcriptions and in vivo as multi-sequences transcription vectors. *Nucleic Acids Res* 19(19): 5125-5130.
- Terns MP, Terns RM. 2011. CRISPR-based adaptive immune systems. *Curr Opin Microbiol* 14(3): 321-327.
- Walker SC, Avis JM, Conn GL. 2003. General plasmids for producing RNA in vitro transcripts with homogeneous ends. *Nucleic Acids Res* 31(15).
- Wiedenheft B, Sternberg SH, Doudna JA. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482(7385): 331-338.
- Wyatt JR, Chastain M, Puglisi JD. 1991. Synthesis and purification of large amounts of RNA oligonucleotides. *Biotechniques* 11(6): 764-769.

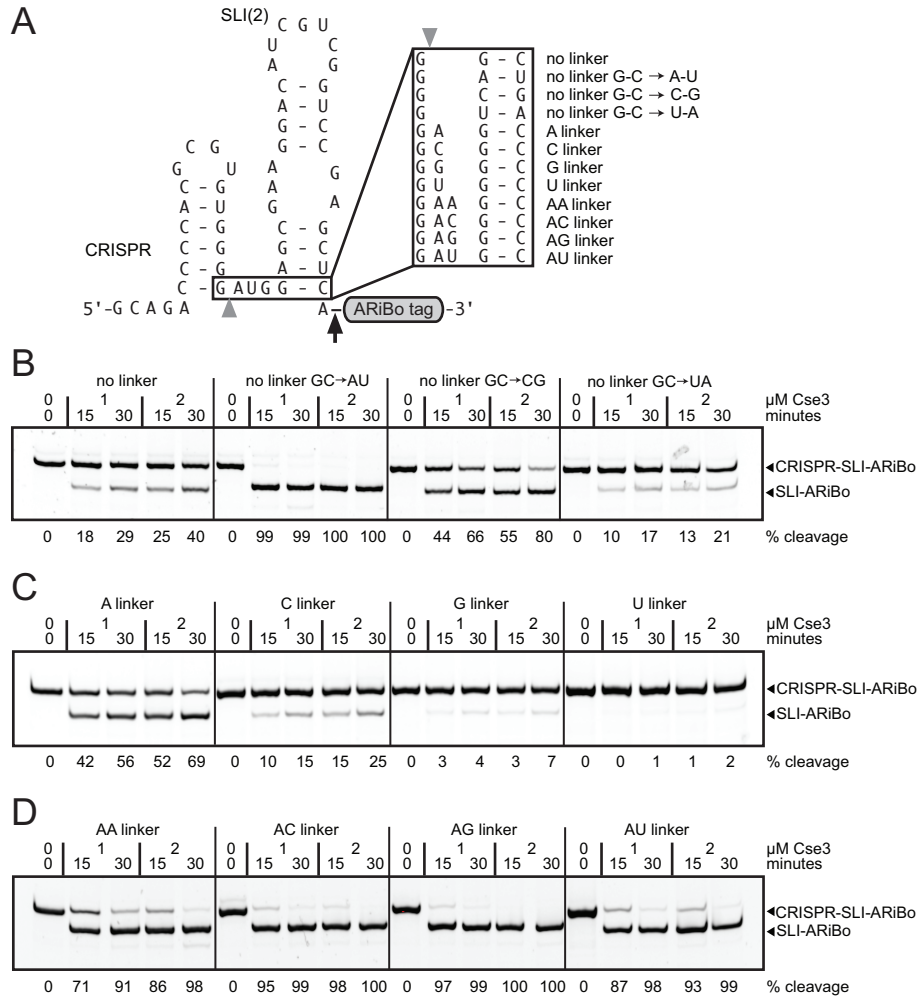
Supplemental material



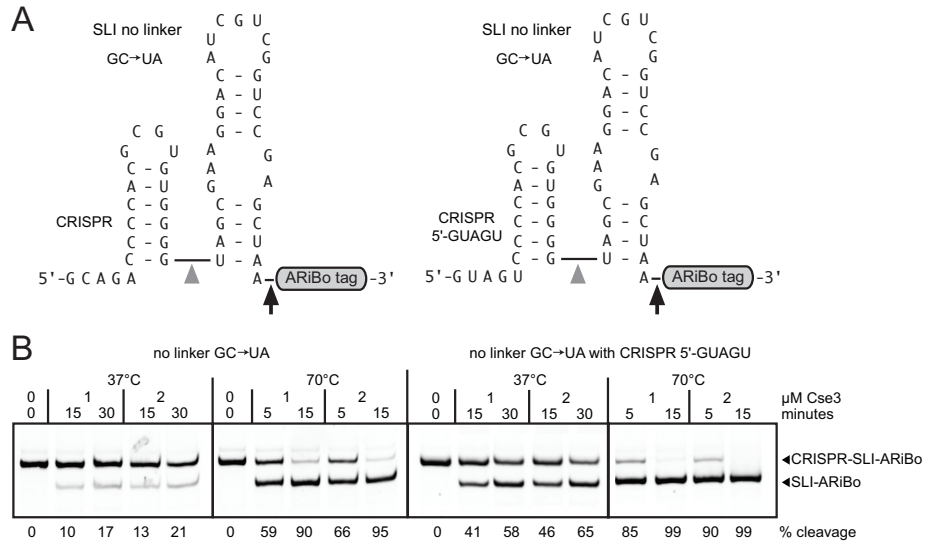
Supplementary Figure 2.1 : The ARiBo4 tag used in this study is a P2 helix mutant of the ARiBo1 tag previously employed for affinity purification (Di Tomasso et al. 2011; Di Tomasso et al. 2012). The primary and proposed secondary structures of the ARiBo1 tag are shown with the original P2 region and the corresponding region in ARiBo4, both boxed in grey, as indicated.



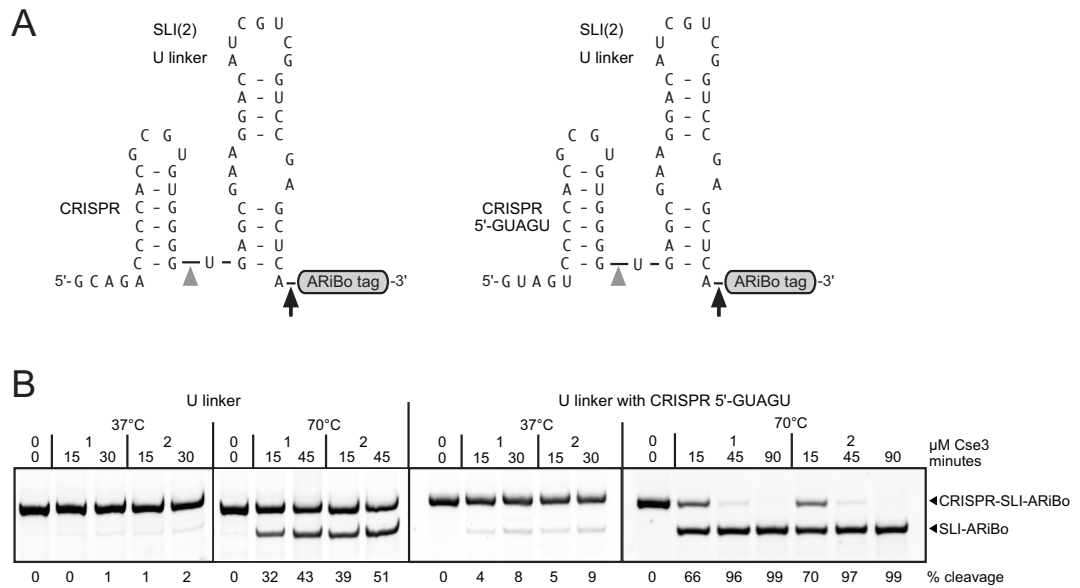
Supplementary Figure 2.2 : Evidence of 5'-heterogeneity from VS ribozyme cleavage of affinity-purified SLI RNAs transcribed as SLI-ARiBo precursors with the wild-type T7 RNAP from the T7 class III promoter. The E1 elution fractions of each of the 16 small-scale affinity purifications are analyzed on a 20% denaturing polyacrylamide gel stained with SYBR Gold. Prior to elution, the bound ARiBo-fusion SLI RNAs were incubated with the VS ribozyme. The 29-nucleotide SLI substrates (S) and 24-nucleotide products (P) are detected on the gel. Gel lanes match the SLI numbering given in Fig. 2.2A. From the 400- μ L elution volumes, 2 μ L aliquots were loaded on the gel.



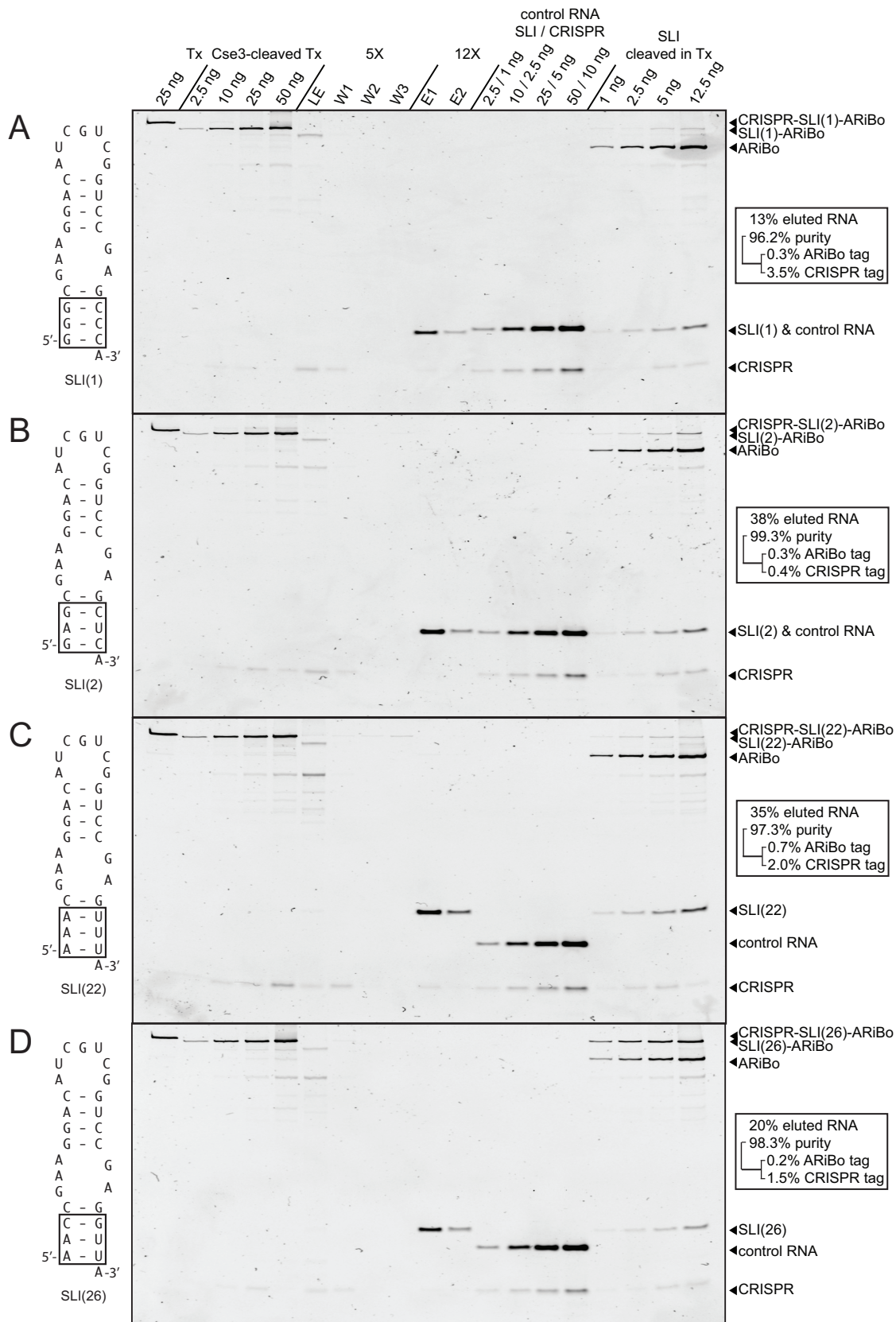
Supplementary Figure 2.3 : Effect of CRISPR-RNA junction sequence on Cse3 cleavage at 37°C. (A) Schematic representation of CRISPR-SLI(2)-ARiBo double-fusion RNAs with the original CRISPR sequence (AUG linker) or related variants with sequence changes at the CRISPR-RNA junction (boxed area). The grey arrowhead points to the Cse3 cleavage site, whereas the black arrow points to the glmS cleavage site. (B), (C) and (D) Cse3 cleavage of CRISPR-SLI-ARiBo RNAs analyzed on 10% denaturing polyacrylamide gels stained with SYBR Gold. Cse3 cleavage was performed at 37°C using aliquots from the transcription reactions (~1 μM RNA), 20 mM HEPES pH=7.5, 150 mM KCl, either 1 or 2 μM Cse3 and different incubation times, as indicated above each lane. The gel mobility of the RNA precursor (CRISPR-SLI-ARiBo) and the Cse3 cleavage product (SLI-ARiBo) are indicated with arrows on the right side of the gel. The percentages of Cse3 cleavage are given below the gels.



Supplementary Figure 2.4 : Reversion of the CRISPR 5'-tail back to the wild-type sequence favors Cse3 cleavage of a CRISPR-SLI-ARiBo double-fusion RNA with a U-A base pair closing the SLI stem. (A) Schematic representation of CRISPR-SLI-ARiBo double-fusion variant RNAs with no linker between the CRISPR and the SLI RNA and a G-C to U-A base-pair change at the bottom of the SLI stem (no linker GC→UA). The original sequence with the 5'-GCAGA CRISPR tail (no linker GC→UA) was changed to a 5'-GUAGU tail (no linker GC→UA with CRISPR 5'-GUAGU). The grey arrowhead points to the Cse3 cleavage site, whereas the black arrow points to the glmS cleavage site. (B) Cse3 cleavage of CRISPR-SLI-ARiBo RNAs analyzed on 10% denaturing polyacrylamide gels stained with SYBR Gold. Cse3 cleavage was performed at 37°C or 70°C using aliquots from the transcription reactions (~1 μM RNA), 20 mM HEPES pH=7.5, 150 mM KCl, either 1 or 2 μM Cse3 and different incubation times, as indicated above each lane. The gel mobility of the RNA precursor (CRISPR-SLI-ARiBo) and the Cse3 cleavage product (SLI-ARiBo) are indicated with arrows on the right side of the gel. The percentages of Cse3 cleavage are given below the gels.



Supplementary Figure 2.5 : Reversion of the CRISPR 5'-tail back to the wild-type sequence favors Cse3 cleavage of a CRISPR-SLI-ARiBo double-fusion RNA with a U linker between the CRISPR and the SLI(2) RNA. (A) Schematic representation of CRISPR-SLI(2)-ARiBo double-fusion variant RNAs with a single-nucleotide U linker between the CRISPR and the SLI(2) RNA. The original sequence with the 5'-GCAGA CRISPR tail (U linker) was changed to a 5'-GCAGU tail (U linker with CRISPR 5'-GUAGU). The grey arrowhead points to the Cse3 cleavage site, whereas the black arrow points to the glmS cleavage site. (B) Cse3 cleavage of CRISPR-SLI-ARiBo RNAs analyzed on a 10% denaturing polyacrylamide gel stained with SYBR Gold. Cse3 cleavage was performed at 37°C or 70°C using aliquots from the transcription reactions (~1 μM RNA), 20 mM HEPES pH=7.5, 150 mM KCl, either 1 or 2 μM Cse3 and different incubation times, as indicated above each lane. The gel mobility of the RNA precursor (CRISPR-SLI-ARiBo) and the Cse3 cleavage product (SLI-ARiBo) are indicated with arrows on the right side of the gel. The percentages of Cse3 cleavage are given below the gels.



Supplementary Figure 2.6 : Small-scale affinity batch purifications of (A) SLI(1), (B) SLI(2), (C) SLI(22) and (D) SLI(26) RNAs analyzed on 15% denaturing polyacrylamide gels stained with SYBR Gold. The SLI RNAs (left panels) were transcribed as CRISPR-SLI-ARiBo double-fusion RNAs and purified by affinity as described in Fig. 2.5A. Aliquots from each purification step were loaded on the gel (right panels; LE: load eluate; W1–3: washes; and E1–2: elutions) in amounts shown, where 1X correspond to approximately 50 ng of ARiBo-fusion SLI present in the transcription reaction or the equivalent of 8.23 ng of SLI to be purified. In addition, standard quantities of CRISPR-SLI-ARiBo RNA from the transcription reaction (Tx), SLI-ARiBo RNA from the Cse3-cleavage transcription reaction (Cse3-cleaved Tx), a mixture containing known amounts of a control RNA (29 nt) and the CRISPR RNA (Control RNA/CRIPSR) as well as standard amounts of SLI from CRISPR-SLI-ARiBo cleaved with Cse3 and GlcN6P in the transcription reaction (SLI cleaved in Tx) were loaded as controls. Bands corresponding to the CRISPR-SLI-ARiBo RNA, the SLI-ARiBo, the ARiBo tag, the control RNA, SLI RNA [SLI(1), SLI(2), SLI(22) or SLI(26)] and CRISPR RNA are indicated on the right side of the gels.

References (for supplemental material only)

- Di Tomasso G, Dagenais P, Desjardins A, Rompré-Brodeur A, Delfosse V, Legault P. 2012. Affinity purification of RNA using an ARiBo tag. In *Recombinant and in vitro RNA synthesis: Methods and Protocols*, Vol 941, p. 306. Humana Press.
- Di Tomasso G, Lampron P, Dagenais P, Omichinski JG, Legault P. 2011. The ARiBo tag: a reliable tool for affinity purification of RNAs under native conditions. *Nucleic Acids Res* 39(3): e18.

Chapitre 3 – Discussion

3.1 – Retour sur les résultats

L'article du Chapitre 2 présente deux différentes méthodes visant à éliminer l'hétérogénéité en 5' d'ARN transcrits *in vitro* par la polymérase à ARN du phage T7. L'hétérogénéité en 5' de l'ARN est problématique pour différents types d'expériences qui requièrent un ARN très pur, particulièrement pour les études structurales telles que la RMN et la cristallographie. Pour remédier à cette problématique, deux méthodes pour éliminer la présence d'hétérogénéité en 5' ont été élaborées; la première consiste à choisir les séquences en 5' des ARN et la seconde à utiliser le système CRISPR/Cse3 en 5' de l'ARN d'intérêt. Couplées à la méthode de purification ARiBo, ces deux solutions procurent un ARN très pur et avec des extrémités homogènes.

3.1.1 – Sélection des séquences en 5'

La première stratégie consiste à sélectionner une séquence en 5' d'un ARN qui ne produira pas d'hétérogénéité en 5'. En effet, nous pensions qu'en modifiant les trois premiers nucléotides de la séquence SLI afin d'obtenir 32 différentes séquences, nous parviendrions à identifier des séquences ne présentant aucune hétérogénéité en 5'. Afin de confirmer cette hypothèse, une étude systématique de purification par affinité a été effectuée sur des séquences d'ARN débutant par différentes combinaisons de séquence en 5'. Chacun des ARN a été transcrit par la polymérase à ARN du phage T7 et purifié avec la méthode ARiBo. L'ARN récupéré dans l'élution lors de la purification a été séparé sur un gel dénaturant de polyacrylamide coloré au SYBR Gold. Ensuite, le degré d'hétérogénéité en 5' a été quantifié pour chacune des séquences. Le pourcentage d'hétérogénéité a été calculé à partir de l'intensité des bandes d'ARN mesurés sur le gel avec le système ChemiDoctm MP selon la formule illustrée à la Figure 3.1. Un seuil de 1.5% d'hétérogénéité a été jugé le seuil acceptable d'impureté des ARN.

Pour cette étude, 32 séquences dérivées de la tige boucle SLI (29 nt) (Figure 3.2B), un substrat non-naturel du ribozyme VS de *Neurospora*, ont été purifiées avec la méthode ARiBo. SLI a été choisi comme modèle pour l'étude car

il s'agit d'un petit ARN, et pour un tel ARN, une variation de taille de seulement un nucléotide est généralement observable sur gel. De plus, la séquence de SLI peut être facilement mutée et sa purification est facile avec la méthode ARiBo. Finalement, ce substrat non-naturel du VS peut tout de même être clivé après son cinquième nucléotide en 5', ce qui nous a été utile dans cette étude.

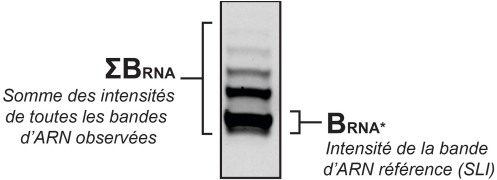
$$\text{Hétérogénéité} = \left[\frac{(\sum B_{RNA} - B_{RNA^*})}{\sum B_{RNA}} \right]$$


Figure 3.1 : Formule du calcul de l'hétérogénéité de séquence en 5' d'un ARN transcrit par la polymérase à ARN du phage T7. La formule prend en compte deux paramètres. Le premier est la somme de l'intensité de chacune des bandes observées dans l'éluion. Le second est l'intensité de la bande de référence, c'est-à-dire de la bande correspondant à l'ARN d'intérêt.

Les 32 séquences SLI ont été synthétisées à l'aide des promoteurs de classe III et de classe II ϕ 2.5 de la polymérase à ARN du phage T7, et débutant respectivement par un G ou par un A. Seulement les deuxièmes et troisièmes paires de base à l'extrémité de la tige-boucle SLI ont été modifiées afin d'obtenir toutes les combinaisons de nucléotides possibles; soit 16 séquences débutant en 5' par GNN et 16 séquences débutant en 5' par ANN (N correspondant à un A, un C, un G ou un U). Chacune de ces séquences ont été transcrites par la polymérase à ARN du phage T7 de type sauvage et par une polymérase mutée (P266L). La polymérase mutée P266L est reconnue pour faciliter le passage de la polymérase au niveau du promoteur⁷. Nous avons démontré que l'utilisation de cette polymérase au lieu de la polymérase de type sauvage diminue le pourcentage d'hétérogénéité en 5' dans plusieurs cas. Cette polymérase élimine complètement l'hétérogénéité (<1.5%) pour certains ARN spécifiques, particulièrement ceux utilisant le promoteur de classe II ϕ 2.5 (Tableaux I et II).

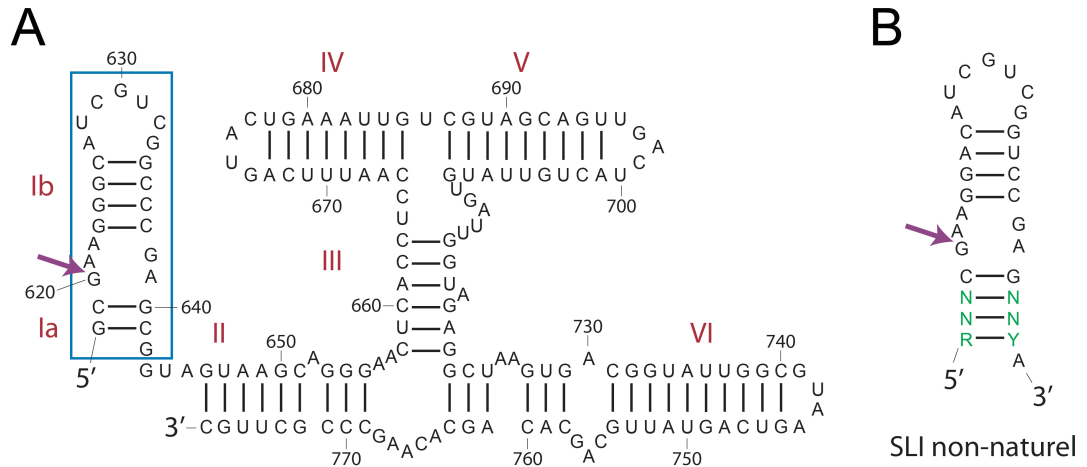


Figure 3.2 : Structure secondaire du ribozyme VS de *Neurospora* et de son substrat SLI. (A) Le substrat du ribozyme VS, SLI, est encadré en bleu sur la figure. La flèche mauve représente le site de clivage du ribozyme (Adaptée de Desjardins et al, 2011)⁵⁴. (B) La tige-boucle SLI modifiée pour des études d'hétérogénéité. Le site de clivage du ribozyme VS est indiqué par une flèche mauve et les nucléotides qui ont été modifiés pour l'étude sont en vert (R pour purine, Y pour pyrimidine et N pour n'importe quel résidu).

Des 16 séquences débutant par GNN transcrites avec la polymérase de type sauvage, 11 présentent moins d'hétérogénéité que le seuil de 1.5% fixé (Figure 2.2), ce qui était le cas pour seulement 5 des 16 séquences débutant en 5' par ANN (Figure 2.3). En résumé, il y a 16 séquences transcrites avec la polymérase de type sauvage, ayant un pourcentage d'hétérogénéité acceptable et celle-ci débute par GUG, GGA, GAA, GCA, GUA, GAC, GUC, GGU, GAU, GCU, GUU, AAG, ACG, AGC, ACU et AUU. L'utilisation de la polymérase mutée P266L, permet l'ajout des séquences débutant par GCC, AUG et AGU parmi les séquences présentant une hétérogénéité de 1.5% ou moins (Tableaux I et II). En générale, les séquences débutants par un A sont plus susceptibles de présenter une hétérogénéité en 5', mais certaines présentent néanmoins aucune hétérogénéité lorsqu'elles sont purifiées avec l'étiquette ARiBo.

Dans le but de s'assurer que l'hétérogénéité observée sur gel provenait exclusivement de l'extrémité 5' de l'ARN, les ARN SLI ont été clivés par le ribozyme VS de *Neurospora*. Ces ARN peuvent être clivés par le ribozyme VS

après leur 5^e nucléotide en 5' (Figure 3.2). Le produit de 5 nucléotides étant élué lors des lavages, seulement le produit de 24 nucléotides est visible sur le gel (Supplementary Figure 2.1). Les produits de 24 nucléotides observés sont tous parfaitement purs et homogènes. Ces résultats confirment que la séquence en 5' affecte seulement l'homogénéité en 5' de l'ARN.

En résumé, cette étude a permis de générer une liste de séquences ne présentant pas d'hétérogénéité en 5' et qui peut servir de guide pour de futures synthèses d'ARN. L'utilisation d'une des séquences énumérées plus haut (Tableaux I et II) pour la purification par affinité utilisant l'étiquette ARiBo permet ainsi de purifier un ARN sous sa forme native avec une homogénéité de séquence en 5' et en 3'.

3.1.2 – Utilisation du système CRISPR/Cse3 pour assurer l'homogénéité en 5'

Pour éliminer l'hétérogénéité en 5', le choix d'une séquence optimale pour la purification est la méthode la plus rapide et la plus simple. Cependant, une méthode alternative doit être envisagée pour les cas où la séquence choisie est transcrite avec de l'hétérogénéité en 5'. La stratégie adoptée dans l'article présenté au Chapitre 2 consiste à ajouter une étiquette CRISPR clivable par l'endonucléase Cse3 de *Thermus thermophilus*. La courte séquence du CRISPR facilite le clonage et ne risque pas d'affecter les rendements de la transcription. De plus, il s'agit d'une petite tige-boucle suffisamment stable pour que l'ARN se replie bien sur lui-même. Le CRISPR peut être clivé par la Cse3 directement après la réaction de transcription, suite à quoi la purification avec l'étiquette ARiBo peut être réalisée selon le protocole original⁵⁵.

Un clivage complet (100%) est primordial à l'obtention d'un ARN pur. Si l'étiquette CRISPR n'est pas entièrement clivée dans la transcription, les ARN non clivés contaminants co-élueront et des ARN de différentes séquences seront mélangés dans l'élution. Des tests de clivage avec Cse3 ont été effectués avec différentes séquences connectrices impliquant l'ajout d'aucun, un, deux ou trois nucléotides reliant le CRISPR à l'ARN d'intérêt (Figure 2.4, Supplementary Figure 2.3, Supplementary Figure 2.4 et Supplementary Figure 2.6). En faisant

varier légèrement les temps et températures d'incubation de l'ARN double-fusion avec l'endonucléase Cse3, un clivage du CRISPR de plus de 99% a pu être obtenu pour chacune des séquences connectrices. À titre d'exemple, les ARN débutants par une adénine sont généralement clivés à plus de 99% en moins de 30 min à 37°C. Cependant, certains ARN nécessitent de plus longues incubations (de 30 min à 90 min) et de plus hautes températures (70°C) pour être clivés, ce qui tend à augmenter les risques de dégradation de l'ARN. Malgré ces risques, il est important de noter qu'en cas de nécessité, l'étiquette CRISPR peut être clivée à plus de 99%, et ce, quelle que soit la nature de la séquence suivant le site de clivage. L'utilisation d'une étiquette CRISPR permet donc non seulement l'élimination complète de l'hétérogénéité en 5', mais elle permet aussi le choix du nucléotide en 5' de la séquence d'ARN désirée⁸.

Les résultats obtenus pour la purification par affinité à partir d'un ARN double fusion CRISPR-SLI-ARiBo démontrent l'efficacité de la méthode pour obtenir un ARN homogène en 5' et en 3'. Les deux ARN débutant par GNN (GGG et GAG) et les deux ARN débutant par ANN (AAA et AAC) présentant le plus d'hétérogénéité lorsque purifiés avec la méthode ARiBo sans CRISPR ont été transcrits avec une étiquette CRISPR en plus. Les séquences débutant par GGG, GAG, AAA et AAC ne présentaient alors plus d'hétérogénéité en 5' après le clivage du CRISPR et la purification avec l'étiquette ARiBo. Dans certains cas (AAA et AAC), la bande de plus forte intensité obtenue lors de la purification à partir de l'ARN fusion SLI-ARiBo sans étiquette CRISPR s'est avérée ne pas être la séquence d'ARN désirée (Figure 2.5). L'ajout d'une étiquette CRISPR en 5' permet donc de remédier facilement aux problèmes d'hétérogénéité en 5' reliés à la transcription *in vitro* par la polymérase à ARN du phage T7.

3.1.3 – Utilisation du ribozyme HH pour éliminer l'hétérogénéité en 5'

Suite à des essais supplémentaires de purification par affinité à partir d'ARN de fusion CRISPR-ARN-ARiBo, il a été observé que l'utilisation du système CRISPR/Cse3 n'était pas idéale pour tous les ARN⁵³. Certaines séquences d'ARN étaient clivées en présence de l'endonucléase Cse3, donc dans ces cas

l'utilisation du système CRISPR/Cse3 n'est pas une solution viable pour régler le problème d'hétérogénéité en 5'. Nous avons donc décidé d'explorer l'utilisation d'une étiquette composée du ribozyme hammerhead (Figure 1.7)⁵³. Le ribozyme HH est autoclivable, il est donc clivé directement dans la transcription, ce qui présente un avantage, car une étape de clivage supplémentaire n'est pas nécessaire contrairement à l'utilisation de l'étiquette CRISPR (Figure 3.3).

En effet, dans le cas de pre-miR-153-2, un ARN de 88 nt associé à la maladie de Parkinson⁵⁶, l'ajout d'une étiquette CRISPR en 5' et le clivage avec l'endonucléase Cse3 étaient incompatibles avec la purification ARiBo. Plusieurs séquences non identifiées apparaissaient sur le gel de suivi de la purification dans le puits contenant un échantillon de l'élution. Ceci est probablement dû au clivage non spécifique par l'endonucléase Cse3. Par contre, le même ARN couplé à une étiquette ribozyme HH en 5' a pu être parfaitement purifié avec la méthode ARiBo⁵³.

L'étiquette HH en 5' a également été testée pour purifier une séquence SLI à partir d'un ARN de fusion HH-SLI-ARiBo, mais dans ce cas l'étiquette se retrouvait dans l'élution lors de la purification et ce, même si le clivage était effectué à 98%⁵³. Ceci est vraisemblablement dû aux conditions non dénaturantes de la purification avec l'étiquette ARiBo qui empêchent la dissociation du ribozyme HH de l'ARN d'intérêt. Pour l'ARN SLI, l'utilisation du système CRISPR/Cse3 est donc préférable à celle de l'étiquette HH pour obtenir un ARN pur et homogène.

Nos résultats démontrent que ni l'étiquette CRISPR, ni l'étiquette HH en 5' de l'ARN de fusion ARiBo représente une solution universelle pour générer un ARN avec une extrémité 5' homogène. Avoir la possibilité d'utiliser l'une de ces deux méthodes est donc avantageux pour assurer l'homogénéité des ARN transcrits par la polymérase à ARN du phage T7 et purifiés avec la méthode ARiBo. Malgré tout, l'utilisation du système CRISPR/Cse3 semble une meilleure alternative que le ribozyme HH pour l'élimination de l'hétérogénéité en 5'. En effet, l'endonucléase Cse3 de *T. thermophilus* se lie à la tige boucle spécifique de 21 nt et clive directement après le 21^e nt⁵¹. Aucune modification de la séquence

n'est nécessaire pour le clivage comme il est nécessaire avec le ribozyme HH. De plus, le ribozyme HH n'a pas une aussi grande compatibilité avec la méthode ARiBo à cause des conditions natives qui doivent être utilisées. Il en résulte que, dans certains cas, le ribozyme HH reste lié à l'ARN d'intérêt et ce même après un clivage complet. Par contre, dans les cas où la Cse3 clive la séquence de l'ARN d'intérêt, le ribozyme HH représente une alternative intéressante. Ainsi, il existe maintenant trois méthodes pour éliminer l'hétérogénéité en 5'. Idéalement, la séquence en 5' de l'ARN peut être choisie pour réduire l'hétérogénéité. Si la séquence de l'ARN désiré n'est pas idéale et qu'il est impossible de la modifier, deux différentes étiquettes, HH et CRISPR, peuvent être testées pour remédier au problème de l'hétérogénéité en 5'.

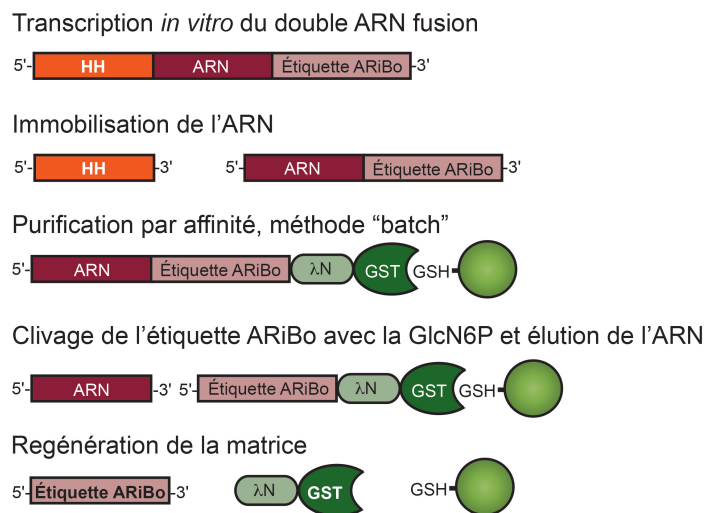


Figure 3.3 : Représentation schématique des étapes de la méthode de purification ARiBo jumelée à l'utilisation d'une étiquette hammerhead pour éliminer l'hétérogénéité en 5'. Les grandes étapes de purification sont illustrées de la transcription de l'ARN de fusion jusqu'à la régénération de la résine GSH-sepharose. (Adaptée de Di Tomasso et al. 2014).

3.2 – Suite du projet

La méthode ARiBo a été développée afin de fournir un outil de purification d'ARN facile et rapide mais surtout adaptable pour n'importe quel type ou longueur d'ARN. Cette méthode a récemment été adaptée pour différents types de

projets tel que la purification de longs ARN et l'isolation de complexes ARN-protéines. J'ai moi-même contribué à ces projets dans le cadre de ma maîtrise tel que décrit ci-dessous.

3.2.1 – Purification d'un long ARN, le polycistron miR-106b-25

Démontrer la fiabilité, l'utilité et l'adaptabilité de la méthode de purification par affinité d'ARN en utilisant une étiquette ARiBo était primordiale pour le laboratoire. Pour ce faire, une large variété d'ARN a été purifiée en utilisant la méthode ARiBo par plusieurs membres du laboratoire dans le cadre de différents projets. Différents ARN ont donc été transcrits avec l'étiquette ARiBo et purifiés par la suite. Ainsi, la méthode a pu être adaptée pour la purification d'une variété d'ARN avec des séquences, structures secondaires et longueurs différentes⁵³.

La complexité de purification par affinité des ARN est, en général, proportionnelle à la taille de ceux-ci. Pour purifier de longs ARN (>300 nt), de nombreux problèmes peuvent survenir et empêcher le bon déroulement d'une purification avec l'étiquette ARiBo. Tout d'abord, il est impératif que l'ARN d'intérêt et l'ARN de l'étiquette ARiBo soient tous deux repliés de façon indépendante, puisqu'un mauvais repliement peut empêcher l'immobilisation de l'ARN sur la résine et peut mener à une inhibition partielle ou complète du clivage. Pour cette raison, des modifications peuvent être apportées à la séquence de l'étiquette ARiBo (un exemple est illustré à la Supplementary Figure 2.1), dans les cas où la séquence d'ARN d'intérêt ne serait pas compatible avec une étiquette ARiBo donné. Des modifications peuvent également être apportées au niveau de la séquence connectrice entre l'ARN d'intérêt et l'étiquette ARiBo.

Le polycistron miR-106b-25 est un long transcrit primaire de miARN, retrouvé dans le noyau des cellules humaines. Ces ARN polycistroniques sont pris en charge exactement comme le sont les miARN monocistroniques (Figure 3.4)⁵⁷; le précurseur de miARN (pre-miARN) est généré par le clivage du transcrit primaire par le complexe protéique Drosha/DGCR8 dans le noyau. Le pre-miARN est ensuite exporté dans le cytoplasme par l'exportine 5-RanGTP où il est pris en

charge par Dicer en complexe avec Argonaute et PACT ou TRBP qui clivera le pre-miARN en un simple brin d'ARN, le miARN mature. La particularité du polycistron de miARN est que le transcrit primaire contient plusieurs pre-miARN (Figure 3.4). L'ARN polycistronique, ou polycistron, purifié pour cette expérience est miR-106b-25, un ARN d'environ 600 nucléotides qui contient trois pre-miARN, pre-miR-106b, pre-miR-93 et pre-miR25 (Figure 3.5).

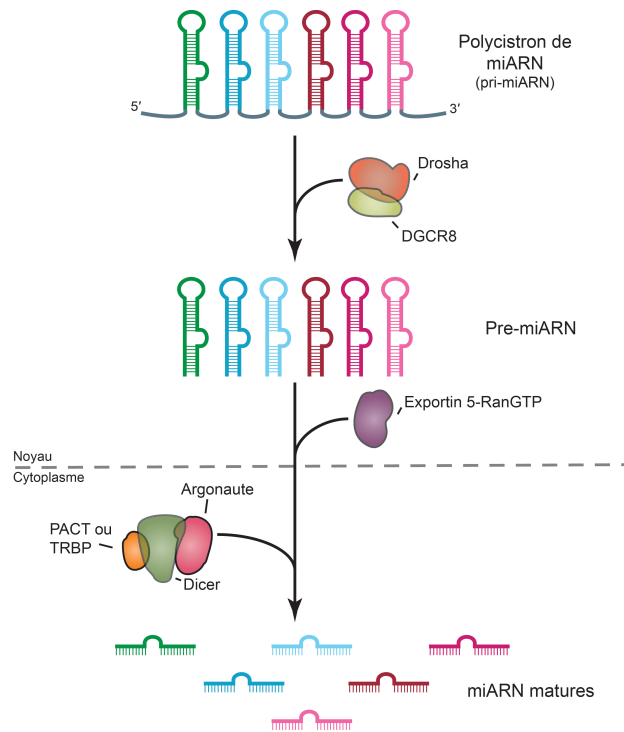


Figure 3.4 : Représentation schématique simplifiée des grandes étapes de la biogénèse d'un polycistron de miARN. On retrouve l'ARN sous forme polycistronique uniquement au niveau du pri-miARN. Suite au clivage du transcrit primaire par Droscha/DGCR8, la biogénèse de chaque pre-miARN est poursuivie de façon individuelle (Adaptée de Chaulk et al. 2011)⁵⁸.

Pour faciliter la purification, il est possible de prédire les structures secondaires d'ARN, grâce à des logiciels de prédiction de structure, tels que le logiciel Sfold⁵⁹. Ces prédictions ne sont pas entièrement fiables, notamment pour les ARN aux structures complexes, mais permettent d'avoir une idée de la

compatibilité entre les séquences de l'étiquette ARiBo et l'ARN d'intérêt. Au niveau des structures secondaires prédites, une attention particulière doit être portée aux différents constituants de l'ARN de fusion: l'ARN d'intérêt et l'étiquette ARiBo doivent être repliés de façon indépendante, le site de clivage du ribozyme *glmS* doit être bien dégagé pour un bon clivage et la séquence *λBxB* doit être accessible pour une bonne liaison à la protéine fusion $\lambda N^+ - L^+ - GST$ (Figure 3.6).

La première étape, avant la purification, est de s'assurer que l'ARN transcrit peut être clivé en présence de GlcN6P. Afin de faire le choix le plus judicieux possible pour sélectionner l'ARN de fusion optimal à la purification du polycistron miR-106b-25, les prédictions ont été faites à l'aide du logiciel Sfold⁵⁹. Après avoir conceptualisé plusieurs différentes structures secondaires d'ARN de fusion, certaines ont été rejetées directement sans être synthétisées. Par exemple, l'étiquette ARiBo1, une étiquette très utilisée dans le laboratoire, n'a jamais été synthétisée avec miR-106b-25 car l'étiquette et l'ARN ne se repliaient pas indépendamment l'un de l'autre (Figure 3.6A). La simulation Sfold a été utilisée pour d'autres ARN de fusion incluant le polycistron miR-106b-25 et différentes versions de l'étiquette ARiBo. Seuls les ARN de fusion dont le miR-106b-25 et l'étiquette ARiBo semblaient bien se replier ont été choisis pour être synthétisés, tels que miR-106b-25-15A-ARiBo4 (Figure 3.6B). Les légères modifications apportées à l'étiquette ARiBo4 dans la boucle P2.2, par rapport à l'étiquette ARiBo1 ont permis un meilleur repliement. En effet, avec cet ARN de fusion, l'ARN d'intérêt miR-106b-25 (599 nt) est replié indépendamment de l'étiquette ARiBo; le site de clivage du ribozyme *glmS* est bien dégagé et le site *λBxB* est bien replié et est accessible.

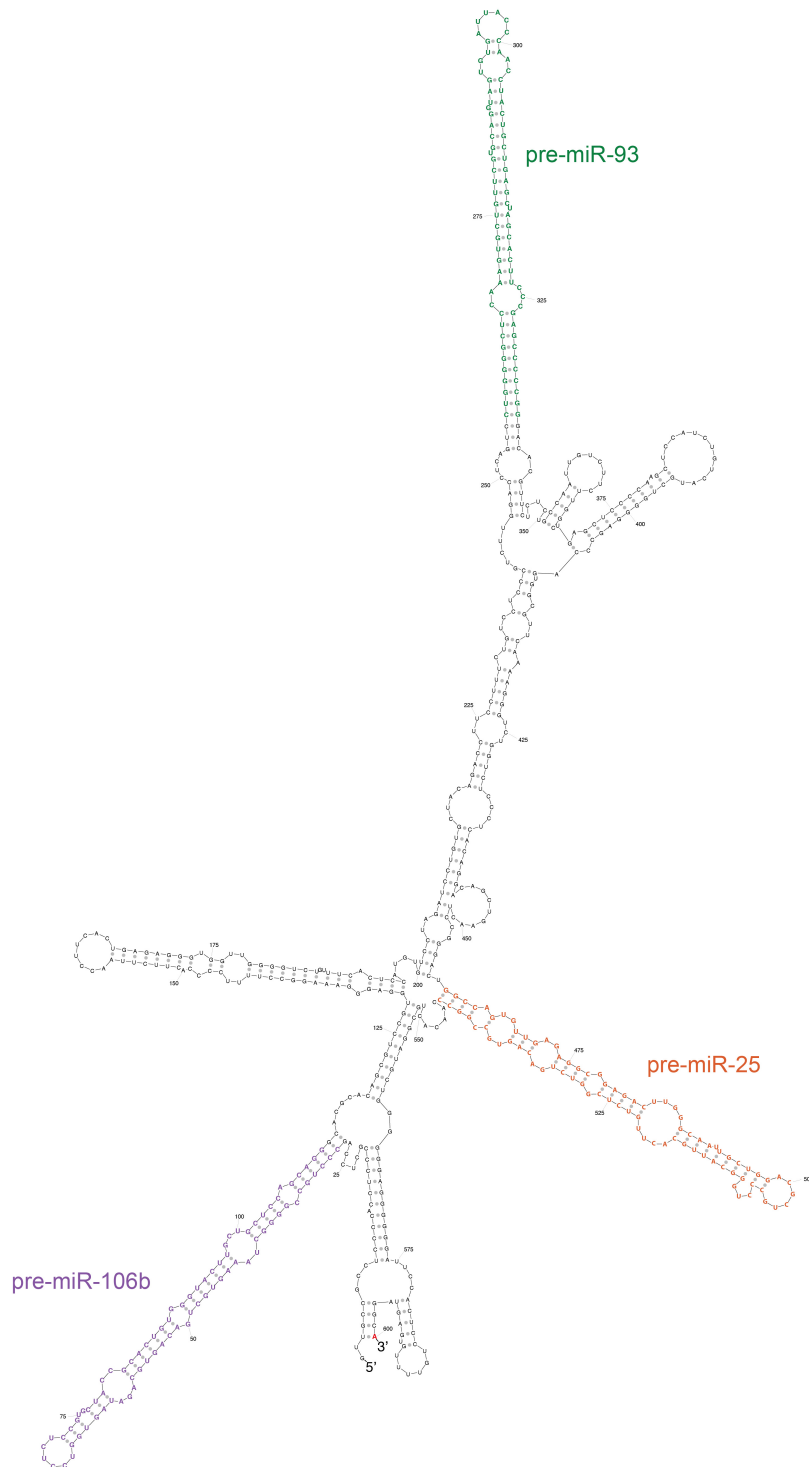


Figure 3.5 : Séquence du polycistron de miARN miR106b-25. La structure secondaire de l'ARN a été prédite grâce au logiciel Sfold⁵⁹, chacun des trois sites pre-miARN sont identifiés en trois couleurs différentes sur la séquence.

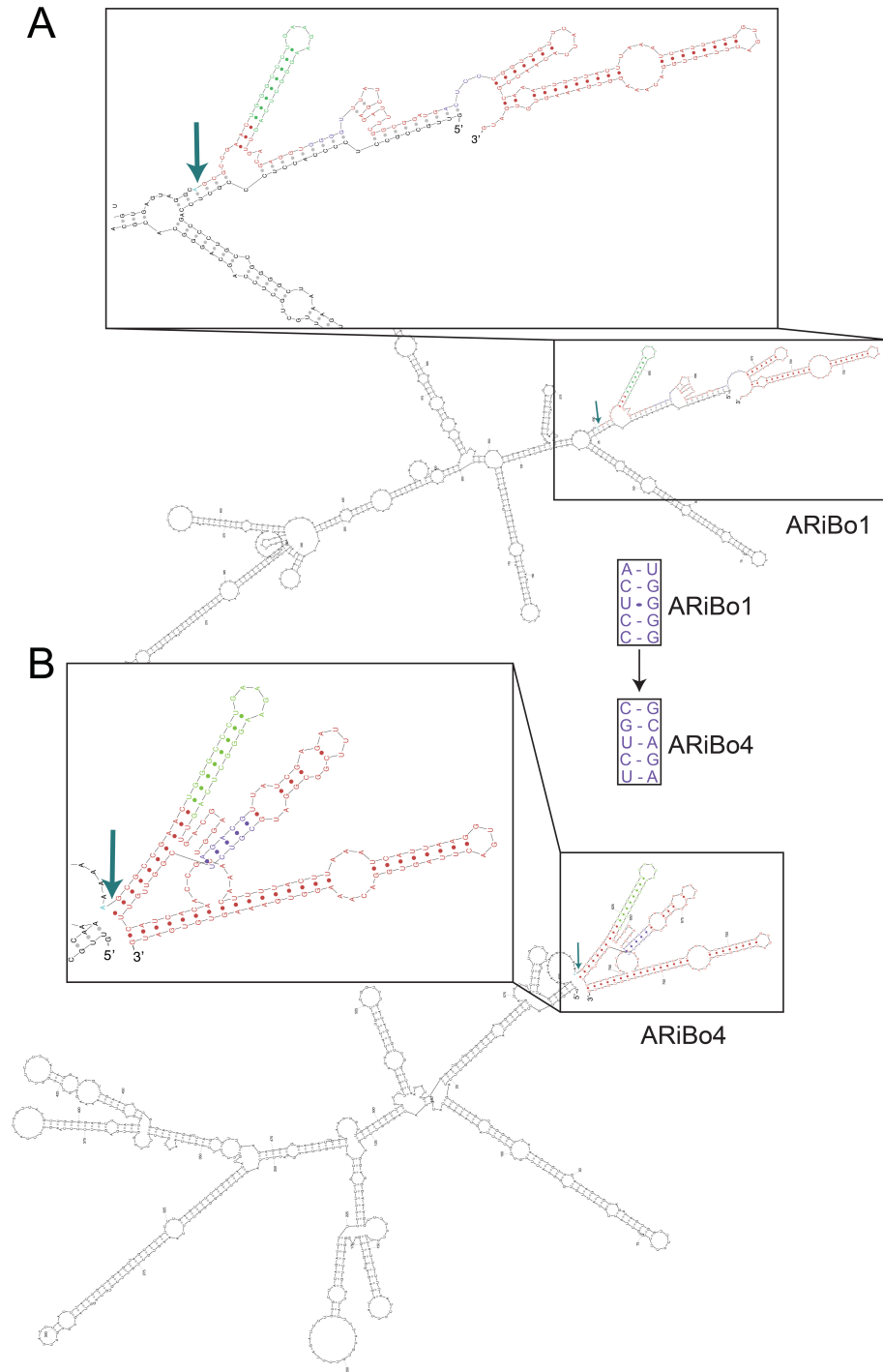


Figure 3.6 : Prédiction Sfold de miR-106b-25-ARiBo1 et miR-106b-25-15A-ARiBo4.

Sont représentés sur la figure : l'étiquette ARiBo (rouge), les mutations (mauve) entre ARiBo1 (A) et ARiBo4 (B), λ BoxB (vert) et le site de clivage de la GlcN6P (flèche bleue). Avec ARiBo1 (A), miR-106b-25 se replie avec l'étiquette alors qu'avec ARiBo4 (B) l'étiquette et miR-106b-25 sont repliés de façon indépendante⁵⁹.

Pour parvenir à purifier le polycistron miR-106b-25, plusieurs ARN de fusion ont été synthétisés et le clivage de l'étiquette par le GlcN6P a été testé directement dans un échantillon de la transcription complétée (Figure 3.7). Le premier ARN de fusion synthétisé est miR-106b-25-ARiBoXII, ARiBoXII étant un autre dérivé de l'étiquette ARiBo originale dont la boucle P2.2 a subi une légère modification. La prédiction *Sfold* de ce premier ARN de fusion semblait donner un bon repliement, par contre, le clivage de cet ARN de fusion avec la GlcN6P ne fonctionnait pas du tout (Figure 3.7A). Le deuxième essai a été fait avec l'étiquette ARiBo4 (Supplementary Figure 2.1). Avec cette étiquette, le clivage s'effectuait, mais n'était pas complet (environ 50%; Figure 3.7B). Pour améliorer le clivage, une adénine supplémentaire a été ajoutée entre miR-106b-25 et l'étiquette ARiBo. L'ajout de cette adénine supplémentaire a permis une amélioration considérable du clivage (environ 80%; Figure 3.7C). Malgré le clivage encore incomplet de ce nouvel ARN de fusion, la purification par affinité a été faite. La purification ARiBo standard à partir de cet ARN de fusion n'a pas donné de bons rendements, car une majorité de l'ARN a été éluée lors du premier lavage (résultats non montrés).

Les résultats encourageants obtenus en ajoutant seulement une adénine supplémentaire laissent présager que l'ajout d'une plus longue séquence connectrice entre l'ARN d'intérêt et l'étiquette ARiBo pourrait permettre d'améliorer le clivage et le rendement de purification. Ainsi, 15 adénines ont été ajoutées entre miR-106b-25 et l'étiquette ARiBo4, ce qui a permis un clivage complet de l'ARN d'intérêt avec la GlcN6P (Figure 3.7D). Les premiers essais de purification avec ce dernier ARN de fusion en utilisant le protocole original de la méthode ARiBo n'étaient pas concluants, donc des modifications ont dû être faites afin de permettre la purification de miR-106b-25 avec un bon rendement. Une première modification a été apportée qui avait déjà améliorée les rendements lors de la purification ARiBo d'ARN de plus de 100 nt au laboratoire⁵³. Il s'agit de l'ajout de sel (NaCl) lors de l'éluion qui peut aider à détacher un ARN qui reste lié à la résine après le clivage de l'étiquette ARiBo. Donc, durant l'éluion, suite au clivage de l'étiquette, du chlorure de sodium est ajouté pour améliorer les

rendements d'élution possiblement en diminuant les liaisons non spécifiques de l'ARN clivé à la protéine de fusion ou la résine.

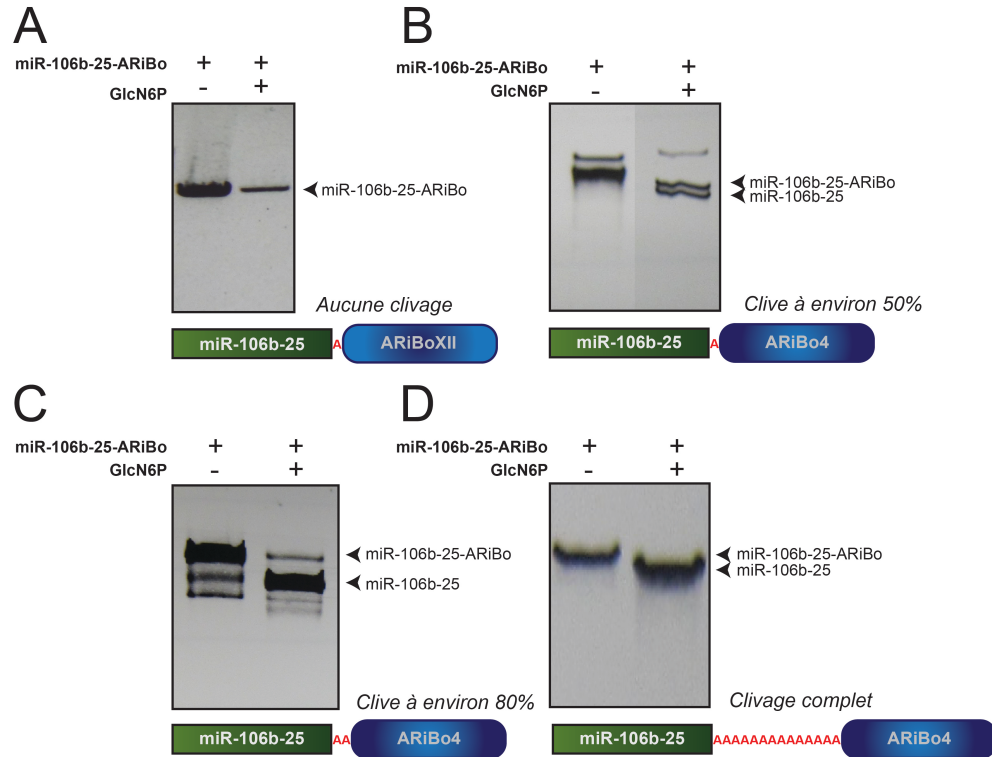


Figure 3.7 : Différentes étiquettes ARiBo utilisées pour la purification du polycistron miR-106b-25. (A-D) Les différentes étiquettes ARiBo sont illustrées ainsi que le résultat du clivage obtenu avec chacune de ces étiquettes.

Une autre modification apportée au protocole est le temps d'incubation. La protéine $\lambda N^+ - L^+ - GST$ a été incubée une nuit entière (environ 16h) au lieu de 15 min et cette incubation s'est faite à 4°C, au lieu de la température de la pièce pour éviter tout risque de dégradation de l'ARN ou de la protéine. Le temps d'incubation du complexe ARN-protéine avec la résine à température ambiante a également été augmenté de 15 min à 1 h. Allonger ainsi ces temps d'incubation a permis d'obtenir de meilleurs résultats de purification, possiblement en favorisant les liaisons spécifiques de l'ARN à la protéine et du complexe ARN-protéine à la résine.

Ainsi, effectuer des modifications au protocole standard ARiBo pour optimiser la purification de miR-106b-25, soit l'utilisation de temps d'incubation plus longs et l'ajout de sel dans l'élution, ont permis la purification du polycistron miR-106b-25 avec 15 A en 3' (Figure 3.8). Pour cet ARN de 614 nt, un rendement d'environ 13% a été obtenu avec une pureté de 98%. De plus, seulement 20% de l'ARN est sorti lors des lavages avant l'élution, ce qui est une nette amélioration par rapport aux premiers essais de purification. La méthode ARiBo a donc pu être adaptée avec succès à un plus large ARN d'environ 600 nt et pourrait possiblement être utilisé pour la purification d'ARN encore plus longs.

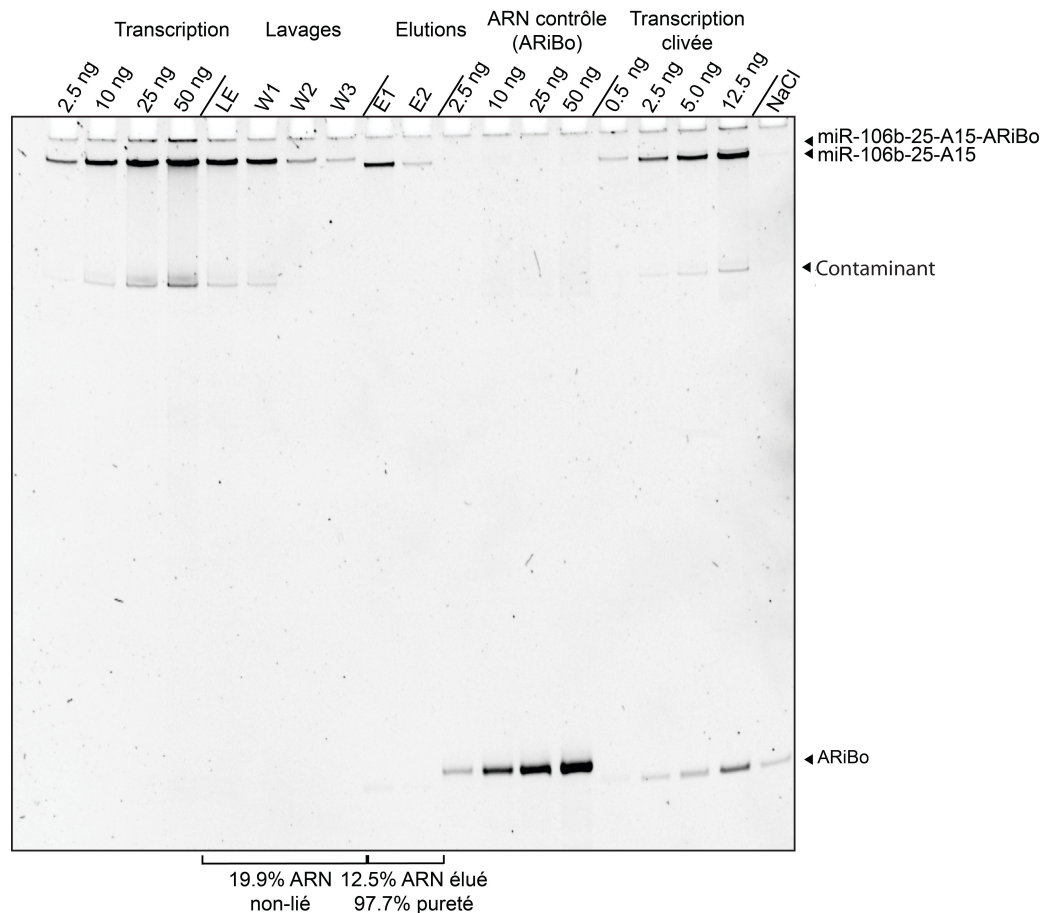


Figure 3.8 : Gel de polyacrylamide de la purification du polycistron miR-106b-25. Purification standard avec chacun des contrôles et lavages (LE, W1, W2, W3) de la purification du polycistron miR-106b-25. L'ARN est élué (E1, E2) à environ 12,5% de rendement et une pureté de 98%.

3.2.2 – Isolation de complexes ribonucléoprotéiques (RNP)

La méthode de purification par affinité a récemment été adaptée dans notre laboratoire pour l'isolation de complexes ribonucléoprotéiques (RNP). Plusieurs méthodes existent pour isoler des complexes RNP, mais ces méthodes utilisent en général une protéine immobilisée afin de repêcher des ARN et autres protéines qui s'y lient. Cependant, seulement quelques méthodes ont été développées pour lier l'ARN à une résine afin d'isoler les protéines qui s'y lient. La méthode du Ribotrap⁶⁰ est basée sur la synthèse d'un ARN transcrit avec des 5-bromo-UTP, ce qui permet à cet ARN de se lier à une résine protéine G couplée à un anticorps liant les 5-bromo-UTP (Figure 3.9A). Le désavantage principal de cette méthode est que l'ARN modifié n'adopte pas nécessairement sa conformation native et par conséquent, les protéines qui s'y lieront ne seront pas nécessairement spécifiques à cet ARN. D'autres méthodes utilisant un ARN pour recruter des protéines existent, comme le StreptoTag qui se lie à une matrice de streptomycine⁶¹ (Figure 3.9B) ou une étiquette formée d'une structure d'ARNt et d'un aptamère d'ARN capable de se lier à une résine Streptavidine⁶² (Figure 3.9C). L'ARNt ajouté entre l'aptamère de Streptavidine et l'ARN d'intérêt n'est pas essentiel, mais permet un meilleur repliement de l'étiquette et rend cette étiquette plus stable. Par rapport à ces méthodes, la méthode ARiBo est basée sur des interactions qui offrent une plus grande affinité entre l'étiquette choisie et la résine utilisée. De plus, un autre avantage est que l'étiquette ARiBo est composée du ribozyme activable *glms*. L'avantage du ribozyme activable est qu'il permet de cliver facilement la séquence d'ARN d'intérêt. Par contre, pour les méthodes avec étiquettes streptomycine et Streptavidine, les ARN d'intérêt seront élués avec leurs étiquettes. Ces étiquettes peuvent lier des protéines non spécifiques à l'ARN d'intérêt qui se retrouveront dans l'éluion.

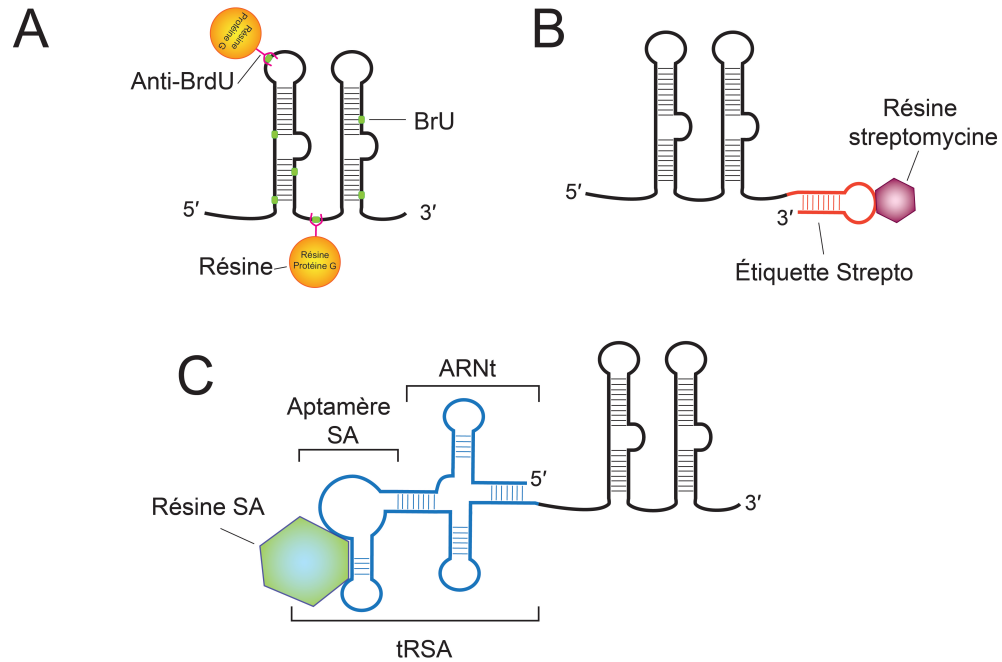


Figure 3.9 : Représentation schématique de différentes méthodes d'immobilisation de l'ARN pour l'isolation de RNP. (A) La méthode du RiboTrap est illustrée⁶⁰, elle utilise des UTP marqués (BrU) reconnue par un anticorps (BrdU) lié à une résine pour immobiliser l'ARN (Adaptée de Beach et Keene, 2008). (B) La méthode du « Strepto tag »⁶³ utilise une étiquette Strepto qui se lie à une résine streptomycine pour immobiliser l'ARN (Adaptée de Windbichler et Schroeder, 2006). (C) La méthode comporte une étiquette tRSA⁶², il s'agit d'une étiquette formée d'une structure d'ARNt et d'un aptamère de streptavidine (SA) qui se lie à une résine contenant la streptavidine (Adaptée de Iioka et al, 2011).

L'isolation de complexes RNP avec la méthode ARiBo se fait comme la purification ARiBo standard (Figure 1.4), mais en ajoutant une étape d'incubation avec un lysat cellulaire juste avant l'étape d'éluion de l'ARN. Ainsi, les protéines du lysat cellulaire pourront se lier spécifiquement à l'ARN et pourront être recueillies dans l'éluion avec l'ARN (Figure 3.10). Pour identifier ces protéines, un immunobuvardage de type Western et/ou de la spectrométrie de masse peuvent être réalisés sur les échantillons d'éluion.

Transcription *in vitro* de l'ARN fusion avec l'étiquette ARiBo



Immobilisation de l'ARN



Saisie des protéines du lysat cellulaire par l'ARN



Autoclivage de l'étiquette ARiBo avec la GlcN6P et élution de l'ARN et de ses protéines associées



Regénération de la matrice

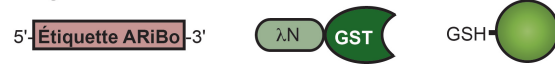


Figure 3.10 : Représentation schématique des étapes d'isolation de complexes ribonucléoprotéiques dérivées de la méthode de purification par affinité utilisant l'étiquette ARiBo. Les grandes étapes de la méthode sont illustrées, de la transcription de l'ARN jusqu'à la régénération de la résine GSH-sepharose.

Pour valider cette méthode impliquant l'ajout d'extrait cellulaire, l'ARN pre-let-7g a été utilisé afin d'isoler la protéine Lin28 (Di Tomasso G. et Legault P., résultats non publiés). L'interaction entre ces deux composés est connue et largement étudiée au laboratoire⁶⁴. Le contrôle négatif utilisé pour ces expériences est l'étiquette ARiBo seule. Dans le cadre d'études avec le lysat cellulaire de la lignée P19⁶⁵ (cellule embryonnaires de souris), il a été observé que plusieurs protéines se retrouvent dans l'élution de la purification de l'ARN de fusion pre-let-7g-ARiBo où seul pre-let-7g est élué, mais non dans l'élution de la purification avec ARiBo seul (mock), où aucun ARN ne devrait normalement être élué. En effet, il a été possible d'identifier Lin28 comme partenaire spécifique de pre-let-7g avec cette méthode. Cette interaction a été révélée par un immunobuvardage de type Western et par spectrométrie de masse. Suite à ces résultats encourageants, le même protocole a été utilisé avec l'ARN pre-miR-153-2 (88 nt; Figure 3.11A). Ce précurseur de miR-153-2 est impliqué dans la régulation de l'α-synuclein, une

protéine dérégulée chez les patients atteints de la maladie de Parkinson⁵⁶. Dans ce cas, la lignée cellulaire utilisée pour l'isolation de protéines est SH-SY5Y (cellules de moelle osseuse humaine). Cette lignée cellulaire est très utilisée pour les études sur la maladie de Parkinson et sera donc plus propice à contenir les protéines régulant pre-miR-153-2 dans les cas de maladie de Parkinson⁶⁶. L'isolation de protéines recrutées par l'ARN pre-miR-153-2 a bien fonctionné. Tout d'abord, j'ai vérifié que la purification de pre-miR-153-2 est efficace avec la méthode standard ARiBo (Figure 3.11B). Ensuite, une purification avec l'ajout d'extrait cellulaire a été faite et les rendements d'ARN obtenus étaient semblables à ceux obtenus avec la méthode standard ARiBo, soit 55% de rendement avec une pureté de 98% (Figure 3.11C). Finalement, les protéines recueillies lors de cette dernière purification et d'une purification contrôle avec ARiBo seul (mock) ont été précipitées et détectées par un gel SDS-PAGE coloré au bleu de coomassie (Figure 3.11D). Les bandes de ce gel, correspondant aux protéines recueillies dans l'élution de pre-miR-153-2 et l'élution contrôle (mock), ont été analysées par spectrométrie de masse (plateforme protéomique de l'IRIC), et les données obtenues ont été analysées grâce au logiciel Scaffold4. Certaines protéines spécifiques à pre-miR-153-2 ont été identifiées grâce aux résultats de spectrométrie de masse. La spécificité d'une des protéines identifiées a été confirmée par immunobuvardage de type Western. Cette protéine semble également se lier au pre-miR7-3 (112 nt) un ARN qui aurait des fonctions similaires à celles de pre-miR153-2⁶⁷ (Figure 3.11). Bien que cette technique requiert encore une certaine optimisation et que les résultats obtenus en spectrométrie de masse pour pre-miR-153-2 soient à reconfirmer par des expériences supplémentaires, la méthode ARiBo semble très prometteuse pour l'isolation de protéines qui lient spécifiquement un ARN et va être utilisée avec différents ARN dans un futur proche.

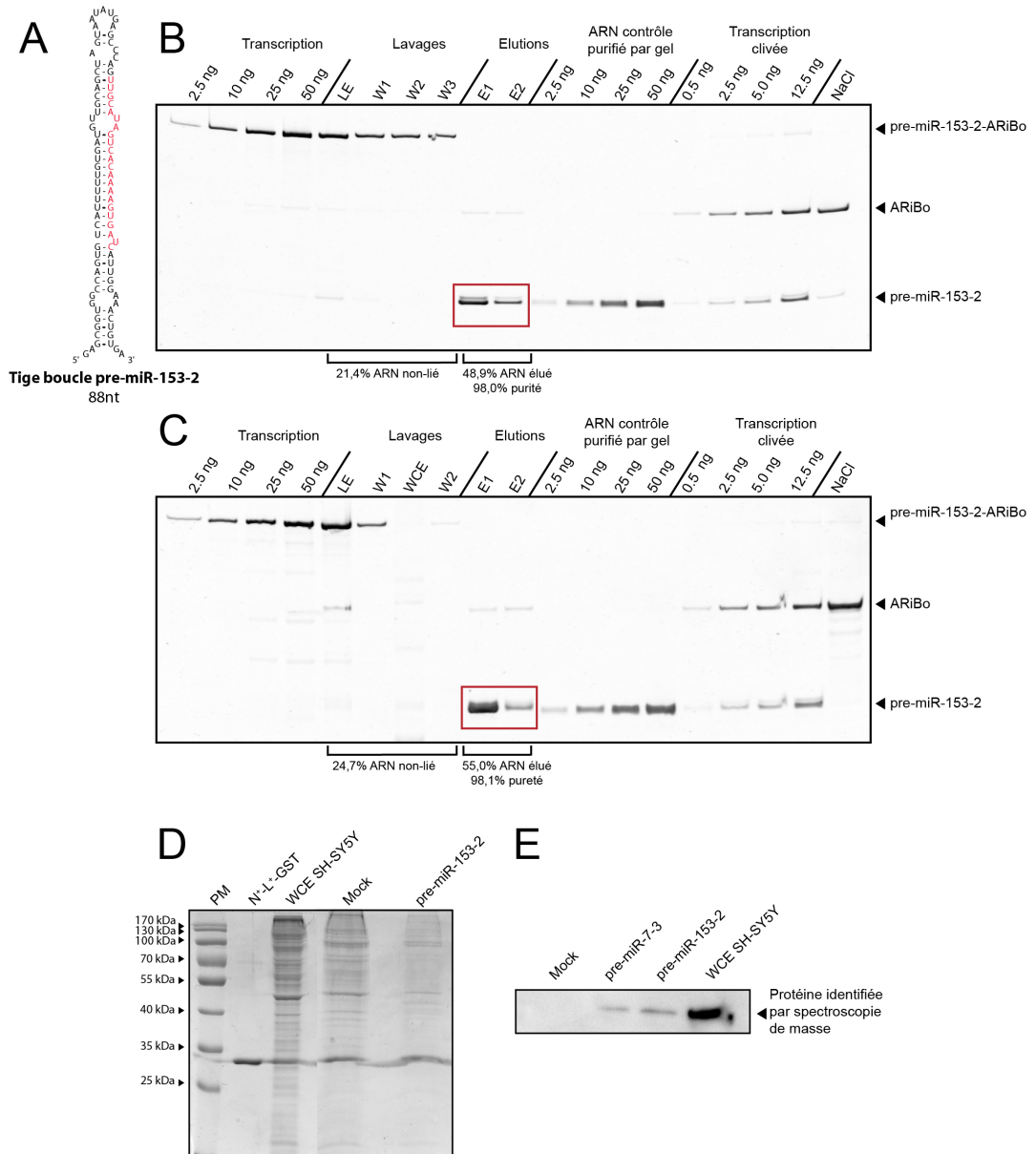


Figure 3.11 : Isolation de complexes ribonucléoprotéiques impliquant pre-miR-153-2.
 (A) Séquence et structure secondaire du pre-miARN. La séquence du miARN mature est identifiée en rouge. (B-C) Gel de polyacrylamide colorés au SYBR Gold. (B) Gel de purification de pre-miR-153-2 avec l'étiquette ARiBo avec tous les contrôles standard. (C) Gel d'ARN de la purification de pre-miR-153-2 avec ajout d'extrait cellulaire (WCE) pour l'isolation de complexes RNP. (D) Gel SDS-PAGE coloré au bleu de coomassie avec les protéines de l'éluion de la purification utilisant l'étiquette ARiBo. Sur le gel, on retrouve le marqueur de poids moléculaire (puits 1), 500 ng de la protéine λN^+L^+ -GST seule (puits 2), 1 μ g d'extrait cellulaire (SH-SY5Y) (puits 3), les protéines de l'éluion du

contrôle mock (puits 4) et les protéines de l'élution de pre-miR-153-2 (puits 5). (E) Un immunobuvardage de type Western montre une protéine identifiée respectivement dans les éluions du mock, de pre-miR-7-3 (un miARN ayant des fonctions communes avec pre-miR153-2), pre-miR-153-2 et l'extrait cellulaire complet (SH-SY5Y) utilisé pour les expériences.

Chapitre 4 – Conclusion

En résumé, la méthode de purification par affinité d'ARN utilisant l'étiquette ARiBo a su répondre à certains problèmes rencontrés avec d'autres méthodes de purification. En comparaison avec la méthode de purification standard par gel de polyacrylamide dénaturant, la méthode ARiBo offre un ARN sous forme native dans un laps de temps beaucoup plus court. De plus, le produit final n'est pas contaminé par des oligomères d'acrylamide. En comparaison avec d'autres méthodes de purification par affinité⁶⁸, la méthode ARiBo procure de meilleurs rendements et une meilleure pureté. Cette méthode de purification de type « batch » permet de purifier plusieurs types d'ARN rapidement en parallèle, procurant un produit pur avec un bon rendement très rapidement, mais surtout sous sa forme native, ce qui n'est pas le cas de la méthode de purification par gel.

Malgré les nombreux avantages que présente la méthode ARiBo originale, deux problèmes persistaient; l'hétérogénéité en 5' des ARN transcrit *in vitro* avec la polymérase à ARN du phage T7 et la purification de longs ARN aux structures secondaires complexes. Pour remédier au problème de l'hétérogénéité trois solutions ont été développées. La première consiste à choisir le début de la séquence en 5' afin qu'il n'y ait pas d'hétérogénéité de séquence. Mon étude systématique de 32 séquences, 16 débutant par GNN et 16 débutant pas ANN, ont permis d'identifier 16 séquences qui présentent 1.5% ou moins d'hétérogénéité lorsque transcrites par les deux polymérases du phage T7. De plus, les séquences GCC, AUG et AGU transcrites avec la polymérase P266L présentent 1.5% d'hétérogénéité ou moins. Pour les ARN ne débutant pas par les séquences de trois nucléotides énumérées plus haut, l'ajout d'une étiquette en 5' de l'ARN d'intérêt est nécessaire. L'utilisation de l'étiquette CRISPR présente plusieurs avantages par rapport à la méthode déjà développée du ribozyme hammerhead²⁸. Il est clivable par l'endonucléase Cse3 à plus de 99% peu importe les nucléotides suivant en 5', est compatible avec la méthode ARiBo et n'a pas besoin d'être adapté d'une séquence d'ARN à l'autre. Mis à part l'élimination de l'hétérogénéité en 5', un autre avantage non négligeable de l'ajout d'une étiquette d'ARN en 5' est le choix plus varié du premier résidu en 5' qui n'a pas à être obligatoirement une purine. Le développement et l'accessibilité de ces deux solutions permettent d'éradiquer

l'hétérogénéité d'un ARN purifié avec la méthode ARiBo.

Ensuite, afin de démontrer la flexibilité de la méthode ARiBo, une large gamme d'ARN, de différentes longueurs et structures, ont été purifiés⁵³. Ainsi, de légères modifications de protocoles ont permis de purifier un polycistron de miARN de 614 nt à la structure secondaire complexe, miR-106b-25. D'abord, le choix de la séquence de l'étiquette ARiBo et l'ajout d'une séquence de 15 adénines entre l'ARN d'intérêt et l'étiquette ARiBo ont permis un bon clivage de la séquence grâce à la GlcN6P. Ensuite, l'augmentation des temps d'incubation et l'ajout de sel lors de l'élution ont permis la purification du polycistron miR-106b-25. Cela démontre donc que la méthode ARiBo, avec de simples modifications au protocole, peut être utilisée pour la purification rapide d'un long ARN de plus de 600 nt ce qui est souvent une difficulté pour les quelques autres méthodes de purification d'ARN par affinité³⁵.

Finalement, la méthode ARiBo a été utilisée afin d'isoler des complexes ARN-protéines en utilisant l'ARN pre-miR-153-2 et un lysat cellulaire provenant des cellules SH-SY5Y. La méthode a été testée avec d'autres ARN au laboratoire et semble donner des résultats très prometteurs.

Ainsi, la méthode ARiBo peut être utilisée pour différentes applications, peu importe la nature ou la taille de l'ARN. Il est donc possible de purifier des ARN homogènes sous leur forme native en très peu de temps. Pour des études structurales qui demandent une très grande pureté et de bons rendements d'ARN, par exemple, cette méthode est idéale. De plus, avec les modifications apportées, elle peut également être utilisée pour isoler des complexes ribonucléoprotéiques. La méthode ARiBo est une des méthodes de purification par affinité les mieux développées et contribuera certainement, dans le futur, à explorer la complexité des systèmes biologiques.

Références

- 1 Risebrough, R. W., Tissieres, A. & Watson, J. D. Messenger-RNA attachment to active ribosomes. *Proc Natl Acad Sci U S A* **48**, 430-436 (1962).
- 2 Gomes, A. Q., Nolasco, S. & Soares, H. Non-coding RNAs: multi-tasking molecules in the cell. *Int J Mol Sci* **14**, 16010-16039, doi:10.3390/ijms140816010 (2013).
- 3 Sanbonmatsu, K. Y. Dynamics of riboswitches: Molecular simulations. *Biochim Biophys Acta*, doi:10.1016/j.bbagr.2014.06.010 (2014).
- 4 Ferre-D'Amare, A. R. The glmS ribozyme: use of a small molecule coenzyme by a gene-regulatory RNA. *Q Rev Biophys* **43**, 423-447, doi:10.1017/s0033583510000144 (2010).
- 5 Caruthers, M. H. A brief review of DNA and RNA chemical synthesis. *Biochem Soc Trans* **39**, 575-580, doi:10.1042/bst0390575 (2011).
- 6 Steitz, T. A. The structural changes of T7 RNA polymerase from transcription initiation to elongation. *Curr Opin Struct Biol* **19**, 683-690, doi:10.1016/j.sbi.2009.09.001 (2009).
- 7 Tang, G. Q. *et al.* Relaxed rotational and scrunching changes in P266L mutant of T7 RNA polymerase reduce short abortive RNAs while delaying transition into elongation. *PLoS One* **9**, e91859, doi:10.1371/journal.pone.0091859 (2014).
- 8 Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* **15**, 8783-8798 (1987).
- 9 Krupp, G. RNA synthesis: strategies for the use of bacteriophage RNA polymerases. *Gene* **72**, 75-89 (1988).
- 10 Zoller, M. J. & Smith, M. Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template. *Methods Enzymol* **154**, 329-350 (1987).

- 11 Bowman, J. C., Azizi, B., Lenz, T. K., Roy, P. & Williams, L. D. Preparation of long templates for RNA in vitro transcription by recursive PCR. *Methods Mol Biol* **941**, 19-41, doi:10.1007/978-1-62703-113-4_3 (2012).
- 12 Chamberlin, M. J. The selectivity of transcription. *Annu Rev Biochem* **43**, 721-775, doi:10.1146/annurev.bi.43.070174.003445 (1974).
- 13 Milligan, J. F. & Uhlenbeck, O. C. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol* **180**, 51-62 (1989).
- 14 Dunn, J. J. & Studier, F. W. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J Mol Biol* **166**, 477-535 (1983).
- 15 Coleman, T. M., Wang, G. & Huang, F. Superior 5' homogeneity of RNA from ATP-initiated transcription under the T7 phi 2.5 promoter. *Nucleic Acids Res* **32**, e14, doi:10.1093/nar/gnh007 (2004).
- 16 Helm, M., Brule, H., Giege, R. & Florentz, C. More mistakes by T7 RNA polymerase at the 5' ends of in vitro-transcribed RNAs. *RNA* **5**, 618-621 (1999).
- 17 Kochetkov, S. N., Rusakova, E. E. & Tunitskaya, V. L. Recent studies of T7 RNA polymerase mechanism. *FEBS Lett* **440**, 264-267 (1998).
- 18 Pleiss, J. A., Derrick, M. L. & Uhlenbeck, O. C. T7 RNA polymerase produces 5' end heterogeneity during in vitro transcription from certain templates. *RNA* **4**, 1313-1317 (1998).
- 19 Guillerez, J., Lopez, P. J., Proux, F., Launay, H. & Dreyfus, M. A mutation in T7 RNA polymerase that facilitates promoter clearance. *Proc Natl Acad Sci U S A* **102**, 5958-5963, doi:10.1073/pnas.0407141102 (2005).
- 20 Ramirez-Tapia, L. E. & Martin, C. T. New insights into the mechanism of initial transcription: the T7 RNA polymerase mutant P266L transitions to elongation at longer RNA lengths than wild type. *J Biol Chem* **287**, 37352-37361, doi:10.1074/jbc.M112.370643 (2012).

- 21 Prody, G. A., Bakos, J. T., Buzayan, J. M., Schneider, I. R. & Bruening, G. Autolytic processing of dimeric plant virus satellite RNA. *Science* **231**, 1577-1580, doi:10.1126/science.231.4745.1577 (1986).
- 22 Sharmeen, L., Kuo, M. Y., Dinter-Gottlieb, G. & Taylor, J. Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J Virol* **62**, 2674-2679 (1988).
- 23 Saville, B. J. & Collins, R. A. A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria. *Cell* **61**, 685-696 (1990).
- 24 Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281-286, doi:10.1038/nature02362 (2004).
- 25 Cochrane, J. C. & Strobel, S. A. Catalytic strategies of self-cleaving ribozymes. *Acc Chem Res* **41**, 1027-1035, doi:10.1021/ar800050c (2008).
- 26 Ferre-D'Amare, A. R. & Scott, W. G. Small self-cleaving ribozymes. *Cold Spring Harb Perspect Biol* **2**, a003574, doi:10.1101/cshperspect.a003574 (2010).
- 27 Scott, W. G., Martick, M. & Chi, Y. I. Structure and function of regulatory RNA elements: ribozymes that regulate gene expression. *Biochim Biophys Acta* **1789**, 634-641, doi:10.1016/j.bbagr.2009.09.006 (2009).
- 28 Ferre-D'Amare, A. R. & Doudna, J. A. Use of cis- and trans-ribozymes to remove 5' and 3' heterogeneities from milligrams of in vitro transcribed RNA. *Nucleic Acids Res* **24**, 977-978 (1996).
- 29 Viladoms, J. & Fedor, M. J. The glmS ribozyme cofactor is a general acid-base catalyst. *J Am Chem Soc* **134**, 19043-19049, doi:10.1021/ja307021f (2012).
- 30 Wyatt, J. R., Chastain, M. & Puglisi, J. D. Synthesis and purification of large amounts of RNA oligonucleotides. *Biotechniques* **11**, 764-769 (1991).

- 31 Booy, E. P., Meng, H. & McKenna, S. A. Native RNA purification by gel filtration chromatography. *Methods Mol Biol* **941**, 69-81, doi:10.1007/978-1-62703-113-4_6 (2012).
- 32 Edwards, A. L., Garst, A. D. & Batey, R. T. Determining structures of RNA aptamers and riboswitches by X-ray crystallography. *Methods Mol Biol* **535**, 135-163, doi:10.1007/978-1-59745-557-2_9 (2009).
- 33 Lukavsky, P. J. & Puglisi, J. D. Large-scale preparation and purification of polyacrylamide-free RNA oligonucleotides. *RNA* **10**, 889-893 (2004).
- 34 Kladwang, W., Hum, J. & Das, R. Ultraviolet shadowing of RNA can cause significant chemical damage in seconds. *Sci Rep* **2**, 517, doi:10.1038/srep00517 (2012).
- 35 Batey, R. T. Advances in methods for native expression and purification of RNA for structural studies. *Curr Opin Struct Biol* **26C**, 1-8, doi:10.1016/j.sbi.2014.01.014 (2014).
- 36 Urh, M., Simpson, D. & Zhao, K. Affinity chromatography: general methods. *Methods Enzymol* **463**, 417-438, doi:10.1016/s0076-6879(09)63026-3 (2009).
- 37 Zhao, X., Li, G. & Liang, S. Several affinity tags commonly used in chromatographic purification. *J Anal Methods Chem* **2013**, 581093, doi:10.1155/2013/581093 (2013).
- 38 Aviv, H. & Leder, P. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc Natl Acad Sci U S A* **69**, 1408-1412 (1972).
- 39 Srisawat, C. & Engelke, D. R. Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA* **7**, 632-641 (2001).
- 40 Srisawat, C., Goldstein, I. J. & Engelke, D. R. Sephadex-binding RNA ligands: rapid affinity purification of RNA from complex RNA mixtures. *Nucleic Acids Res* **29**, E4 (2001).
- 41 Butter, F., Scheibe, M., Morl, M. & Mann, M. Unbiased RNA-protein interaction screen by quantitative proteomics. *Proc Natl Acad Sci U S A* **106**, 10626-10631, doi:10.1073/pnas.0812099106 (2009).

- 42 Batey, R. T. & Kieft, J. S. Improved native affinity purification of RNA. *RNA* **13**, 1384-1389, doi:10.1261/rna.528007 (2007).
- 43 Cheong, H. K., Hwang, E., Lee, C., Choi, B. S. & Cheong, C. Rapid preparation of RNA samples for NMR spectroscopy and X-ray crystallography. *Nucleic Acids Res* **32**, e84, doi:10.1093/nar/gnh081 (2004).
- 44 Jaeger, L. The New World of ribozymes. *Curr Opin Struct Biol* **7**, 324-335 (1997).
- 45 Bevilacqua, P. C. Mechanistic considerations for general acid-base catalysis by RNA: revisiting the mechanism of the hairpin ribozyme. *Biochemistry* **42**, 2259-2265, doi:10.1021/bi027273m (2003).
- 46 Cochrane, J. C., Lipchock, S. V., Smith, K. D. & Strobel, S. A. Structural and chemical basis for glucosamine 6-phosphate binding and activation of the glmS ribozyme. *Biochemistry* **48**, 3239-3246, doi:10.1021/bi802069p (2009).
- 47 Di Tomasso, G., Lampron, P., Dagenais, P., Omichinski, J. G. & Legault, P. The ARiBo tag: a reliable tool for affinity purification of RNAs under native conditions. *Nucleic Acids Res* **39**, e18, doi:10.1093/nar/gkq1084 (2011).
- 48 Di Tomasso, G. *et al.* Affinity purification of RNA using an ARiBo tag. *Methods Mol Biol* **941**, 137-155, doi:10.1007/978-1-62703-113-4_11 (2012).
- 49 Di Tomasso, G., Lampron, P., Omichinski, J. G. & Legault, P. Preparation of lambdaN-GST fusion protein for affinity immobilization of RNA. *Methods Mol Biol* **941**, 123-135, doi:10.1007/978-1-62703-113-4_10 (2012).
- 50 Sorek, R., Kunin, V. & Hugenholtz, P. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**, 181-186, doi:10.1038/nrmicro1793 (2008).
- 51 Gesner, E. M., Schellenberg, M. J., Garside, E. L., George, M. M. & Macmillan, A. M. Recognition and maturation of effector RNAs in a

- CRISPR interference pathway. *Nat Struct Mol Biol* **18**, 688-692, doi:10.1038/nsmb.2042 (2011).
- 52 Sashital, D. G., Jinek, M. & Doudna, J. A. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat Struct Mol Biol* **18**, 680-687, doi:10.1038/nsmb.2043 (2011).
- 53 Di Tomasso, G., Salvail-Lacoste, A., Bouvette, J., Omichinski, J. G. & Legault, P. Affinity purification of in vitro transcribed RNA with homogeneous ends using a 3'-ARiBo tag. *Methods Enzymol* **549c**, 49-84, doi:10.1016/b978-0-12-801122-5.00003-9 (2014).
- 54 Desjardins, G., Bonneau, E., Girard, N., Boisbouvier, J. & Legault, P. NMR structure of the A730 loop of the *Neurospora* VS ribozyme: insights into the formation of the active site. *Nucleic Acids Res* **39**, 4427-4437, doi:10.1093/nar/gkq1244 (2011).
- 55 Schurer, H., Lang, K., Schuster, J. & Morl, M. A universal method to produce in vitro transcripts with homogeneous 3' ends. *Nucleic Acids Res* **30**, e56 (2002).
- 56 Mouradian, M. M. MicroRNAs in Parkinson's disease. *Neurobiol Dis* **46**, 279-284, doi:10.1016/j.nbd.2011.12.046 (2012).
- 57 Winter, J., Jung, S., Keller, S., Gregory, R. I. & Diederichs, S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol* **11**, 228-234, doi:10.1038/ncb0309-228 (2009).
- 58 Chaulk, S. G. *et al.* Role of pri-miRNA tertiary structure in miR-17~92 miRNA biogenesis. *RNA Biol* **8**, 1105-1114, doi:10.4161/rna.8.6.17410 (2011).
- 59 Ding, Y., Chan, C. Y. & Lawrence, C. E. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res* **32**, W135-141, doi:10.1093/nar/gkh449 (2004).
- 60 Beach, D. L. & Keene, J. D. Ribotrap : targeted purification of RNA-specific RNPs from cell lysates through immunoaffinity precipitation to

- identify regulatory proteins and RNAs. *Methods Mol Biol* **419**, 69-91, doi:10.1007/978-1-59745-033-1_5 (2008).
- 61 Bachler, M., Schroeder, R. & von Ahsen, U. StreptoTag: a novel method for the isolation of RNA-binding proteins. *RNA* **5**, 1509-1516 (1999).
- 62 Iioka, H., Loiselle, D., Haystead, T. A. & Macara, I. G. Efficient detection of RNA-protein interactions using tethered RNAs. *Nucleic Acids Res* **39**, e53, doi:10.1093/nar/gkq1316 (2011).
- 63 Windbichler, N. & Schroeder, R. Isolation of specific RNA-binding proteins using the streptomycin-binding RNA aptamer. *Nat Protoc* **1**, 637-640, doi:10.1038/nprot.2006.95 (2006).
- 64 Desjardins, A., Bouvette, J. & Legault, P. Stepwise assembly of multiple Lin28 proteins on the terminal loop of let-7 miRNA precursors. *Nucleic Acids Res* **42**, 4615-4628, doi:10.1093/nar/gkt1391 (2014).
- 65 McBurney, M. W. P19 embryonal carcinoma cells. *Int J Dev Biol* **37**, 135-140 (1993).
- 66 Xie, H. R., Hu, L. S. & Li, G. Y. SH-SY5Y human neuroblastoma cell line: in vitro cell model of dopaminergic neurons in Parkinson's disease. *Chin Med J (Engl)* **123**, 1086-1092 (2010).
- 67 Junn, E. *et al.* Repression of alpha-synuclein expression and toxicity by microRNA-7. *Proc Natl Acad Sci U S A* **106**, 13052-13057, doi:10.1073/pnas.0906277106 (2009).
- 68 Edelmann, F. T., Niedner, A. & Niessing, D. Production of pure and functional RNA for in vitro reconstitution experiments. *Methods*, doi:10.1016/j.ymeth.2013.08.034 (2013).